Time Varying Heteroskedastic Realized GARCH models for tracking measurement error bias in volatility forecasting

Gerlach, Richard and Naimoli, Antonio and Storti, Giuseppe

University of Sydney, University of Salerno, University of Salerno

Time Varying Heteroskedastic Realized GARCH models for tracking measurement error bias in volatility forecasting

Richard Gerlach[1], *Antonio Naimoli*[2] and *Giuseppe Storti*[2]

1. The University of Sydney – Business School

2. University of Salerno – Department of Economics and Statistics

ABSTRACT – This paper proposes generalisations of the Realized GARCH model by Hansen et al. (2012), in three different directions. First, heteroskedasticity in the noise term in the measurement equation is allowed, since this is generally assumed to be time-varying as a function of an estimator of the Integrated Quarticity for intra-daily returns. Second, in order to account for attenuation bias effects, the volatility dynamics are allowed to depend on the accuracy of the realized measure. This is achieved by letting the response coefficient of the lagged realized measure depend on the time-varying variance of the volatility measurement error, thus giving more weight to lagged volatilities when they are more accurately measured. Finally, a further extension is proposed by introducing an additional explanatory variable into the measurement equation, aiming to quantify the bias due to the effect of jumps and measurement errors.

1   Introduction

In the econometric literature it is widely acknowledged that the use of intra-daily information, in the form of realized volatility measures (Hansen and Lunde, 2011), can be beneficial for forecasting financial volatility on a daily scale. This is typically done by taking one of two different approaches.

First, dynamic models are directly fitted to the time series of the realized measure. Examples include the Heterogeneous AutoRegressive (HAR) (Corsi, 2009) and Multiplicative Error Models (MEM) (Engle, 2002; Engle and Gallo, 2006). A drawback of this approach is that the estimate is given by the expected level of the realized measure, rather than by the conditional variance of returns. As will be clarified in the next section, realized measures are designed to consistently estimate the integrated variance that, under general regularity conditions, can be interpreted as an unbiased estimator of the conditional variance of returns, but will not equal this latter desired quantity. In addition, two main sources of bias can also arise, as a consequence of microstructure noise and jumps. In practical applications, an additional major source of discrepancy is due to

1

the fact that realized measures are usually built not taking into account the contribution of overnight volatility, which is a relevant part of the conditional return variance that is of interest to risk managers and other professionals.

The second approach makes use of a volatility model, e.g. GARCH-type models, where conditional variance is driven by one or more realized measures. The main idea here is to replace a noisy volatility proxy, e.g. the squared daily returns as employed in the traditional GARCH, with a more efficient realized measure. Differently from the above-mentioned approach, in this case both low frequency (daily returns) and high frequency (realized measures) information is employed in the model. Examples within this class include: the HEAVY model of Shephard and Sheppard (2010) and the Realized GARCH model of Hansen et al. (2012). These two models are closely related but, nevertheless, they are characterised by some distinctive features. HEAVY models are designed for the generation of multi-step ahead forecasts, guaranteed by the inclusion of a dynamic updating equation for the conditional expectation of the chosen realized measure. On the other hand, Realized GARCH models include a measurement equation allowing, in a fully data driven fashion, deeper insight on the statistical properties of the realized measure, and its relationship with latent volatility, for the empirical problem of interest.

A complication arising with both approaches is that realized measures are noisy estimates of the underlying integrated variance, generating a classical errors-in-variables problem. This typically leads to the rise of what is often called attenuation bias: the realized measure has lower persistence than the latent integrated variance. Although it is evident that correcting for this attenuation bias can potentially lead to improved volatility forecasts, this issue has not received much attention in the literature. Recently, Bollerslev et al. (2016) find, via employing a HAR model, that allowing the volatility persistence to depend on the estimated degree of measurement error allows a marked improvement in the model's predictive performance. In the same vein, Shephard and Xiu (2016) find evidence that the magnitude of the response coefficients associated with different realized volatility measures, in a GARCH-X model, is related to the quality of the measure itself. Finally, Hansen and Huang (2016) observe that the response of the current conditional variance to past unexpected volatility shocks is negatively correlated with the accuracy of the associated realized volatility measure.

In this framework, exploiting the flexibility of the Realized GARCH, this paper develops a novel modelling approach that allows for correcting the attenuation bias effect, in a natural and fully data driven way. To this purpose, the standard Realized GARCH is extended by allowing the variability of the measurement error to be time-varying, as a function of an estimator of the integrated quarticity of intra-daily returns. Consequently,

the volatility dynamics are allowed to depend on the accuracy of the realized measure. Namely, the response coefficient of the lagged realized volatility is proposed to depend on a measure of the latter's accuracy, given by the estimated variance of the volatility measurement error, flexibly designed so that more weight is given to lagged volatilities that are more accurately measured. Finally, the proposed modelling approach is further extended to capture further potential effects, related to jumps and measurement errors, by introducing into the measurement equation an additional component, controlling the amount of bias that is generated by noise and by jumps. This allows the separation of measurement error due to microstructure noise and discretization, from that due to the impact of jumps. A notable feature of the proposed model is that the jump correction occurs only on days in which jumps are most likely to happen, while resorting to the use of more efficient standard measures, such as realized variances and kernels, in jumps-free periods.

The paper is organized as follows: in Section 2 the basic theoretical framework behind the computation of realized measures is reviewed, while Section 3 discusses the Realized GARCH model of Hansen et al. (2012); Section 4 presents a time-varying parameter heteroskedastic Realized GARCH model, that allows and accounts for attenuation bias effects: a jumps-free setting is considered first, then a modification of the proposed model, which aims to explicitly account for the impact of jumps and microstructure noise, is discussed in Section 5; QML estimation of the proposed models is discussed in Section 6, while Sections 7 to 9 are dedicated to the empirical analysis; Section 7 presents the main features of the data used for the analysis; Section 8 focuses on the in-sample performance of the proposed models, compared to the standard Realized GARCH model as benchmark, whereas the out-of-sample forecasting performance is analysed in Section 9; Finally, section 10 concludes.

2   Realized measures: a short review

In recent years, the availability of high-frequency financial market data has enabled researchers to build reliable measures of the latent daily volatility, based on the use of intra-daily returns. In the econometric and financial literature, these are widely known as realized volatility measures. The theoretical background to these measures is given by the dynamic specification of the price process in continuous time. Formally, let the logarithmic price p_t of a financial asset, be determined by the stochastic differential process:

$$dp_t = \mu_t dt + \sigma_t dW_t + dJ_t \qquad 0 \le t \le T \ , \qquad (1)$$

where μ_t, σ_t are the drift and instantaneous volatility processes, respectively, whilst W_t is a standard Brownian motion; σ_t ia assumed to be independent of W_t, while J_t is a finite activity jump process. Under assumption of jump absence ($dJ_t = 0$) and a frictionless market the logarithmic price p_t follows a semi-martingale process.

In that case the Quadratic Variation (QV) of log-returns $r_t = p_t - p_{t-1}$ coincides with the Integrated Variance (IV), given by:

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

In the absence of jumps, microstructure noise and measurement error, Barndorff-Nielsen and Shephard (2002) show that IV is consistently estimated by Realized Volatility (RV):

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

where

$$r_{t,i} = p_{t-1+i\Delta} - p_{t-1+(i-1)\Delta}$$

is the i-th Δ-period intraday return, $M = 1/\Delta$. Although IV and the conditional variance of returns do not coincide, there is a precise relationship between these two quantities: under standard integrability conditions (Andersen et al., 2001) show that:

$$E(IV_t | \mathcal{F}_{t-1}) = var(r_t | \mathcal{F}_{t-1}) \ ,$$

where $\mathcal{F}_{t-1}$ denotes the information set at time $(t-1)$. In other words, the optimal forecast of IV can be interpreted as the conditional variance of returns and the difference between these two quantities is given by a zero mean error.

Barndorff-Nielsen and Shephard (2002) show that RV consistently estimates the true latent volatility, when $\Delta \longrightarrow 0$, but in practice, due to data limitations, the following results hold:

$$RV_t = IV_t + \varepsilon_t \tag{4}$$

and

$$\varepsilon_t \sim N(0, 2\Delta IQ_t) \tag{5}$$

where $IQ_t = \int_{t-1}^{t} \sigma_s^4 ds$ is the Integrated Quarticity (IQ). IQ, in turn, can be consistently estimated as:

$$RQ_t = \frac{M}{3} \sum_{i=1}^{M} r_{t,i}^4 \ . \tag{6}$$

On the other hand, if jumps are present, QV will differ from IV, with the difference given by the accumulated squared jumps. Formally, let

$$dJ_t = k_t dq_t \ ,$$

where $k_t = p_t - p_{t-}$ is the size of the jump in the log-price p_t and q_t is a counting process, with possibly time-varying intensity λ_t, such that:

$$P(dq_t = 1) = \lambda_t dt \ .$$

Then, under the assumptions in Andersen et al. (2007):

$$RV_t \underset{p}{\to} QV = IV + \sum_{t-1 \leq s \leq t} k^2(s) \ .$$

Hence, RV is a consistent estimator of QV, but not of IV. An alternative here is to use jump-robust estimators, such as the *Bipower* and *Tripower Variation* (Barndorff-Nielsen and Shephard, 2004), *minRV* or *medRV* (Andersen et al., 2012), that are consistent for IV even in the presence of jumps. In the empirical applications carried out in this work, among the different proposals arising in the literature, focus here is put on the *medRV* estimator, mainly for theoretical reasons: specifically, Andersen et al. (2012) show that in the jump-free case *"the medRV estimator has better theoretical efficiency properties than the tripower variation measure and displays better finite-sample robustness to both jumps and the occurrence of "zero" returns in the sample"*. In addition, unlike the *Bipower Variation* measure, for the *medRV* estimator an asymptotic limit theory in the presence of jumps is available.

The *medRV* estimator proposed by Andersen et al. (2012) is:

$$medRV_t = \frac{\pi}{6 - 4\sqrt{3}\pi} \left(\frac{M}{M - 2} \right) \sum_{i=2}^{M-1} med \left(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}| \right)^2 \ . \quad (7)$$

Nevertheless, in the jump-free case, these jump-robust estimators are substantially less efficient than the simple RV estimator: i.e. *Bipower* and *Tripower Variation*, *medRV* and *minRV* are all asymptotically normal, with asymptotic variance proportional (up to different scale factors) to the IQ (Andersen et al., 2012). Further, in presence of jumps, this quantity will be not consistently estimated by RQ; thus, some alternative jump-robust estimator will be needed. Among several proposals in the literature, for the same reasons discussed above, focus here is on the $medRQ$ estimator proposed by Andersen et al. (2012):

$$medRQ_t = \frac{3\pi M}{9\pi + 72 - 52\sqrt{3}} \left(\frac{M}{M - 2} \right) \sum_{i=2}^{M-1} med \left(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}| \right)^4 \ .$$

$$(8)$$

A further issue is how to consistently estimate QV in the presence of market microstructure frictions. In this direction, several estimators are proposed in the literature to mitigate the influence of market microstructure noise, such as the Two Time Scales approach of Zhang et al. (2005), the Realized Kernel of Barndorff-Nielsen et al. (2008) and the pre-averaged RV of Jacod et al. (2009), among others. In this paper, the Realized Kernel (RK), developed by Barndorff-Nielsen et al. (2008), is employed, specified as:

$$RK = \sum_{h=-H}^{H} K\left(\frac{h}{H+1}\right) \zeta_H, \qquad \zeta_H = \sum_{j=|h|+1}^{M} r_{t,i} r_{t,i-|h|} , \qquad (9)$$

where $K(\cdot)$ is a kernel weight function and H a bandwidth parameter[1].

3 Realized GARCH models

The Realized GARCH (RGARCH), introduced by Hansen et al. (2012), extends the class of GARCH models by first replacing squared returns, as the driver of the volatility dynamics, with a more efficient proxy, such as an RV measure. With this change alone, the resulting specification can be seen as a GARCH-X model, where the realized measure is used as an explanatory variable. A second extension is that the Realized GARCH "completes" the GARCH-X, by adding a measurement equation that explicitly models the contemporaneous relationship between the realized measure and the latent conditional variance.

Formally, let $\{r_t\}$ be a time series of stock returns and $\{x_t\}$ be a time series of realized measures of volatility. Focus here is on the logarithmic RGARCH model, defined via:

$$r_t = \mu_t + \sqrt{h_t}\, z_t \qquad (10)$$
$$log(h_t) = \omega + \beta\, log(h_{t-1}) + \gamma\, log(x_{t-1}) \qquad (11)$$
$$log(x_t) = \xi + \varphi\, log(h_t) + \tau(z_t) + u_t \qquad (12)$$

Here $h_t = var(r_t|\mathcal{F}_{t-1})$ is the conditional variance, $\mathcal{F}_{t-1}$ the historical information set at time $t-1$. To simplify the exposition, in the reminder it is assumed that the conditional mean $\mu_t = E(r_t|\mathcal{F}_{t-1}) = 0$. The innovations z_t and u_t are assumed to be mutually independent, with $z_t \overset{iid}{\sim} (0,1)$ and $u_t \overset{iid}{\sim} (0, \sigma_u^2)$.

The function $\tau(z_t)$ can accommodate leverage effects, since it captures the dependence between returns and future volatility. A common choice (see e.g. Hansen et al. (2012)), found to be empirically satisfactory, is:

$$\tau(z_t) = \tau_1\, z_t + \tau_2(z_t^2 - 1) .$$

---

[1]For detail about the optimal choice of the kernel type and the bandwidth selection see Barndorff-Nielsen et al. (2009).

Substituting the measurement equation into the volatility equation, the model implies an AR(1) representation for $log(h_t)$:

$$log(h_t) = (\omega + \xi\gamma) + (\beta + \varphi\gamma)log(h_{t-1}) + \gamma\,w_{t-1}\,, \tag{13}$$

where $w_t = \tau(z_t) + u_t$ and $E(w_t) = 0$. The coefficient $(\beta + \varphi\gamma)$ reflects the persistence in (the logarithm of) volatility, whereas γ represents the impact of both the lagged return and realized measure on future (log-)volatility.

To ensure the volatility process h_t is stationary the required restriction is $\beta + \varphi\gamma < 1$.

Compared to the linear RGARCH, the log-linear specification has two main advantages: first, it is more flexible, since no constraints on the parameters are required in order to ensure positivity of the conditional variance, which holds automatically by construction; and second, the logarithmic transformation substantially reduces, but does not eliminate, the heteroskedasticity of the measurement equation error term. For these reasons, this paper exclusively focuses on the log-linear specification of the Realized GARCH model.

4 Time Varying Coefficient Heteroskedastic Realized GARCH models with dynamic attenuation bias

In this section a generalization of the basic Realized GARCH specification is proposed that accounts and allows for the natural heteroskedasticity of the measurement error u_t, as well as for dynamic attenuation bias.

In a jump-free world any consistent estimator of the IV can be written as the sum of the conditional variance plus a random innovation. Since the variance of this innovation term is function of the IQ, it seems natural to model the variance of the noise u_t in equation (**??**) as function of the RQ. Thus, it is assumed that the measurement noise variance is time-varying, i.e. $u_t \overset{iid}{\sim} (0, \sigma_{u,t}^2)$. In order to model the time-varying variance of the measurement noise, the specification

$$\sigma_{u,t}^2 = exp\left\{\delta_0 + \delta_1 log\left(\sqrt{RQ_t}\right)\right\} \tag{14}$$

is considered, where the exponential formulation guarantees the positivity of the estimated variance, without imposing constraints on the parameters δ_0 and δ_1. The resulting model is denoted the *Heteroskedastic* Realized GARCH (HRGARCH). It is easy to see that the homoskedastic Realized GARCH is nested within this class, i.e. set $\delta_1 = 0$, and that this restriction can be tested by means of a simple Wald statistic.

In order to account for dynamic attenuation effects in the volatility persistence, in the sense of Bollerslev et al. (2016), the basic HRGARCH

specification is further extended, allowing for time-varying persistence in the volatility equation. This is achieved by letting γ, the impact coefficient of the lagged realized measure, depend on the time-varying variance of the measurement noise u_t. In line with Bollerslev et al. (2016), the impact of past realized measures on current volatility is expected to be down-weighted in periods in which the efficiency of the realized measure is low. The resulting model is called the *Time Varying Heteroskedastic* Realized GARCH (TV-HRGARCH).

Focusing on a log-linear specification, the volatility updating equation of the TV-HRGARCH is given by

$$log(h_t) = \omega + \beta \, log(h_{t-1}) + \gamma_t \, log(x_{t-1}) \tag{15}$$

where

$$\gamma_t = \gamma_0 + \gamma_1 \, \sigma_{u,t-1}^2 \tag{16}$$

and $\sigma_{u,t}^2$ follows the specification in (14). Accordingly, as its fixed coefficients counterpart, the TV-HRGARCH can be represented in terms of a time-varying coefficients AR(1) model for $log(h_t)$

$$log(h_t) = (\omega + \xi\gamma_t) + (\beta + \varphi\gamma_t)log(h_{t-1}) + \gamma_t \, w_{t-1}. \tag{17}$$

5 Accounting for bias from jumps and measurement errors

So far focus is on a simplified setting without the possibility of jumps. Consideration is now given to a variant of the proposed modelling approach, featuring a jumps component as an additional variable in the measurement equation, to capture that source of bias. This is achieved by adding the log-ratio between RV and a jump-robust realized measure, as an explanatory variable. In the empirical application, as anticipated in Section 2, the $medRV$ estimator proposed by Andersen et al. (2012) is employed.

Generally, let $C_t = x_t/x_t^J$ be the ratio between a realized measure x_t and a jump-robust realized measure x_t^J. In the limit, this ratio will converge in probability to the ratio between QV and IV. Values of $C_t > 1$ are interpreted as providing evidence of jumps occurring at time t, while the discrepancy between the two measures is expected to disappear in absence of jumps, leading to values of $C_t \approx 1$. Naturally, sampling variability will play a role here and values $C_t < 1$ will be possible, in a small proportion of cases. This is compatible with the fact that the observed C_t is given by the combination of a latent signal $\bar{C}_t \geq 1$ and a measurement error, thus explaining observed values of C_t below the threshold of 1. A simple way to avoid observations below 1 is to truncate the distribution of C_t

at this threshold, setting all the values below the truncation point equal to 1 (see e.g. Andersen et al. (2007)). However, this does not guarantee consistent filtering of the measurement error (the truncation on the left tail is somewhat arbitrary and the right tail would be untouched) with the potential drawback of introducing an additional source of bias into the analysis. Therefore, taking into account the limited empirical incidence of values of $C_t < 1$, it is decided to work with uncensored values of C_t.

After adding the bias correction variable C_t, the proposed modified measurement equation is:

$$log(x_t) = \xi + \varphi log(h_t) + \eta log(C_t) + \tau(z_t) + u_t^* \tag{18}$$

or equivalently

$$log(x_t^*) = \xi + \varphi log(h_t) + \tau(z_t) + u_t^* , \tag{19}$$

where $log(x_t^*) = log(x_t/C_t^\eta)$.

The specification in (19) implies that when $x_t = x_t^J$, x_t corresponds to x_t^*, so no bias correction is applied. In this way, in the jumps free case, the dynamics of the predicted volatility are still driven by the standard RV estimator, or some noise-robust variant such as the *Realized Kernel* (RK), that in this situation are known to be much more efficient than jump-robust estimators. On the other hand, assuming $\eta > 0$ (as systematically confirmed by our empirical results), when x_t is greater than x_t^J, in the spirit of Barndorff-Nielsen and Shephard (2004) and Andersen et al. (2007), it follows that there is evidence of an upward bias, due to the presence of jumps, meaning that x_t should be reduced, in order to be consistent for the latent IV. This result is achieved, thanks to the correction variable C_t, which, in this case, takes values higher than one and consequently makes $x_t^* < x_t$.

Considering the chosen RGARCH and the AR(1) representation for the log-conditional variance, it follows that:

$$log(h_t) = (\omega + \xi\gamma) + (\beta + \varphi\gamma)log(h_{t-1}) + \gamma w_{t-1}^* , \tag{20}$$

where

$$w_t^* = \tau(z_t) + u_t^*$$

and

$$u_t^* = log(x_t^*) - \xi - \varphi log(h_t) - \tau(z_t). \tag{21}$$

By substituting equation (21) in (20), the log-conditional variance can be alternatively written as

$$log(h_t) = \omega + \beta\, log(h_{t-1}) + \gamma\, log(x_{t-1}) - \gamma\eta\, log(C_{t-1}) \tag{22}$$

9

or equivalently

$$log(h_t) = \omega + \beta\, log(h_{t-1}) + \gamma\, log(x^*_{t-1}). \tag{23}$$

In this modified framework, it then turns out that the log-conditional variance $log(h_t)$ is driven not only by past values of the realized measure but also, with opposite sign, by past values of the associated bias. The additional parameter η allows to adjust the contribution of C_{t-1}. From a different point of view, equation (19) suggests that the volatility updating equation can be rewritten in a form similar to that of the standard RGARCH model, with the substantial difference that the volatility changes are driven instead by the bias-corrected measure x_{t-1}/C^{η}_{t-1}; the amount of correction is determined by the estimated scaling parameter η.

This specification, of course, extends to the HRGARCH and TV-HRGARCH models, with the additional modification that, in order to account for the presence of jumps, the RQ estimator in the specification of $\sigma^2_{u,t}$ must be replaced by a jump-robust estimator, as will be more extensively discussed in the empirical section. In the remainder, to distinguish models incorporating the bias correction variable C_t in the measurement equation, these models will be denoted by addition of the superscript " * ", namely: RGARCH*, HRGARCH* and TV-HRGARCH*.

6 Quasi Maximum Likelihood Estimation

The model parameters can be estimated by standard Quasi Maximum Likelihood (QML) techniques. Let Y_t indicate any additional explanatory variable eventually included in the measurement equation. Following Hansen et al. (2012), the quasi log-likelihood function, conditionally on past information $\mathcal{F}_{t-1}$ and Y_t, is given by

$$\mathcal{L}(r, x; \boldsymbol{\theta}) = \sum_{t=1}^{T} log\, f(r_t, x_t | \mathcal{F}_{t-1}, Y_t)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_h, \boldsymbol{\theta}_x, \boldsymbol{\theta}_\sigma)'$ with $\boldsymbol{\theta}_h$, $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_\sigma$ respectively being the vectors of parameters appearing in the volatility equation ($\boldsymbol{\theta}_h$), in the level of the measurement equation ($\boldsymbol{\theta}_x$) and in the noise variance specification ($\boldsymbol{\theta}_\sigma$).

An attractive feature of the Realized GARCH structure is that the conditional density $f(r_t, x_t | \mathcal{F}_{t-1}, Y_t)$ can be easily decomposed as

$$f(r_t, x_t | \mathcal{F}_{t-1}) = f(r_t | \mathcal{F}_{t-1}) f(x_t | r_t; \mathcal{F}_{t-1}, Y_t).$$

Assuming a Gaussian specification for z_t and u_t, such as $z_t \overset{iid}{\sim} N(0,1)$

and $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$, the quasi log-likelihood function is:

$$\mathcal{L}(r, x; \theta) = \underbrace{-\frac{1}{2} \sum_{t=1}^{T} log(2\pi) + log(h_t) + \frac{r_t^2}{h_t}}_{\ell(r)} + \underbrace{-\frac{1}{2} \sum_{t=1}^{T} log(2\pi) + log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}}_{\ell(x|r)} \ . \quad (24)$$

Since the standard GARCH models do not include an equation for x_t, the overall maximized log-likelihood values are not comparable to that returned from the estimation of standard GARCH-type models; the former will always be larger. Nevertheless, the partial log-likelihood value of the returns component, $\ell(r) = \sum_{t=1}^{T} log f(r_t|\mathcal{F}_{t-1})$, can be still meaningfully compared to the maximized log-likelihood value achieved for a standard GARCH type model.

7   The Data

To assess the performance of the proposed models, an empirical application to four stocks traded on the Xetra Market in the German Stock Exchange is conducted.   This section presents the salient features of the data analysed. The following assets are considered: Allianz (ALV), a financial services company dealing mainly with insurance and asset management; Bayerische Motoren Werke (BMW), a company engaged in vehicle and engine manufacturing; Metro Group (MEO), a cash and carry group and RWE (RWE), a company providing electric utilities.

The original dataset included tick-by-tick data on transactions (trades only) in the period 02/01/2002 to 27/12/2012. The raw data are cleaned, using the procedure described in Brownlees and Gallo (2006), then converted to an equally spaced series of five-minute log-returns, which are aggregated on a daily basis to: compute a time series of 2791 daily open-to-close log-returns; two different realized volatility measures: RV and RK; as well as the jump-robust *medRV* estimator. Only continuous trading transactions during the regular market hours 9:00 am - 5:30 pm are considered.

Table 1 reports some descriptive statistics for daily log-returns (r_t), RV, RK and $medRV$; as well as for the bias correction variables related to RV_t and RK_t, denoted by C_t^{RV} and C_t^{RK}, respectively. For ease of presentation, the values associated with RV, RK are multiplied by 100.

The daily returns range have standard deviation typically around 0.020 and are slightly skewed, negatively so for ALV, BMW and MEO, but positively for RWE. Furthermore, the high kurtosis values indicate much heavier tails than the normal distribution, as expected. As expected, all

Table 1: Summary statistics

		Min.	1Qu.	Med.	Mean	3Qu.	Max.	S.dev.	Skew.	Kurt.
r_t	ALV	-0.147	-0.010	0.000	-0.001	0.008	0.135	0.021	-0.066	8.402
	BMW	-0.135	-0.010	0.000	0.000	0.010	0.153	0.020	-0.039	7.497
	MEO	-0.150	-0.010	-0.001	-0.001	0.009	0.122	0.019	-0.377	8.900
	RWE	-0.108	-0.009	0.000	-0.001	0.008	0.097	0.016	0.065	7.415
$RV_t \times 100$	ALV	0.002	0.011	0.021	0.050	0.047	1.732	0.089	6.999	93.681
	BMW	0.004	0.016	0.028	0.045	0.051	0.842	0.057	5.254	49.634
	MEO	0.004	0.016	0.026	0.045	0.048	1.047	0.060	5.030	48.277
	RWE	0.003	0.012	0.020	0.034	0.035	1.011	0.046	6.635	92.977
$RK_t \times 100$	ALV	0.002	0.011	0.021	0.049	0.047	1.730	0.089	7.012	94.289
	BMW	0.004	0.016	0.028	0.045	0.050	0.842	0.056	5.289	50.367
	MEO	0.004	0.015	0.025	0.044	0.047	1.041	0.059	5.077	49.355
	RWE	0.003	0.012	0.019	0.033	0.035	1.009	0.045	6.703	96.421
$medRV_t \times 100$	ALV	0.002	0.010	0.019	0.045	0.043	1.606	0.083	7.241	99.614
	BMW	0.003	0.014	0.025	0.041	0.046	0.773	0.052	4.999	44.906
	MEO	0.002	0.013	0.023	0.040	0.042	1.032	0.053	5.395	61.048
	RWE	0.002	0.011	0.018	0.031	0.032	0.876	0.042	5.991	75.030
$C_t^{RV} = \frac{RV_t}{medRV_t}$	ALV	0.760	1.000	1.100	1.134	1.219	2.655	0.195	1.536	7.596
	BMW	0.738	0.996	1.093	1.126	1.213	3.131	0.191	1.867	11.612
	MEO	0.744	1.015	1.120	1.162	1.263	3.704	0.223	2.355	17.469
	RWE	0.742	1.003	1.092	1.127	1.213	2.751	0.187	1.635	8.842
$C_t^{RK} = \frac{RK_t}{medRV_t}$	ALV	0.747	0.989	1.090	1.123	1.208	2.654	0.193	1.567	7.835
	BMW	0.736	0.984	1.080	1.114	1.196	3.126	0.190	1.889	11.761
	MEO	0.741	1.005	1.105	1.148	1.243	3.670	0.221	2.399	17.998
	RWE	0.742	0.988	1.076	1.111	1.195	2.747	0.185	1.684	9.111

Summary statistics of daily log-returns r_t, daily Realized Variance RV_t* (*: ×100), daily Realized Kernel RK_t* (*: ×100), daily $medRV_t*$ (*: ×100), bias correction variable C_t^{RV} for RV_t and bias correction variable C_t^{RK} for RK_t . Sample period: January 2002 – December 2012. Min.: Minimum; 1Qu.: First Quartile; Med.: Median; Mean; 3Qu.: Third Quartile; Max.: Maximum; S.dev.: Standard deviation; Skew.: Skewness; Kurt.: Kurtosis.

RV and RK series present very strong positive skew; $medRV$ has smaller standard deviations than RV, RK, which may be expected as a jump-robust estimator.

The bias correction variables C_t^{RV}, C_t^{RK} have mean slightly above one and positive skewness. Their minimums are ≈ 0.75 and maximums $\in [2.66, 3.70]$. This preliminary analysis suggests that the impact of jumps is more important and prevalent compared to the measurement error bias. These aspects are also confirmed by the distributional information on C_t presented in Table 2. Only approximately 5% of observations have C_t below 0.90, suggesting that the incidence of the measurement error in the observed C_t series is rather limited, compared to that of jumps.

Figure 1 displays the daily returns for the four analysed stocks. These reveal three periods of high volatility common to the four stocks: the

Table 2: C_t distribution for RV and RK

	Distribution of C_t^{RV}				Distribution of C_t^{RK}			
	ALV	BMW	MEO	RWE	ALV	BMW	MEO	RWE
0%	0.760	0.738	0.744	0.742	0.747	0.736	0.741	0.742
5%	0.894	0.893	0.900	0.895	0.887	0.883	0.892	0.883
10%	0.930	0.930	0.941	0.931	0.923	0.919	0.929	0.919
25%	1.000	0.996	1.015	1.003	0.989	0.984	1.005	0.988
50%	1.100	1.093	1.120	1.092	1.090	1.080	1.105	1.076
75%	1.219	1.213	1.263	1.213	1.208	1.196	1.243	1.195
90%	1.375	1.354	1.428	1.359	1.362	1.342	1.410	1.343
95%	1.492	1.475	1.565	1.477	1.472	1.463	1.542	1.453
100%	2.655	3.131	3.704	2.751	2.654	3.126	3.670	2.747

Figure 1: Time series of daily log-returns
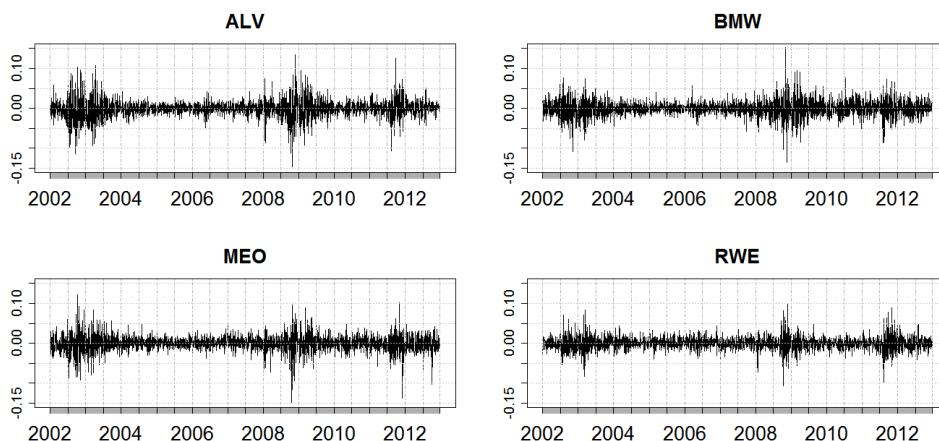
Daily log-returns for the stocks ALV (top-left), BMW (top-right), MEO (bottom-left) and RWE (bottom-right) for the sample period $02/01/2002 - 27/12/2012$.

first relates to the dot com bubble in 2002; the second is the financial crisis starting in mid 2007 and peaking in 2008; the crisis in Europe then progressed from the banking system to a sovereign debt crisis, with the highest turmoil level in the late 2011, the 3rd period. These are clearly evident in Figure 2, time plots of the daily 5-minute RV series.

Finally, Figure 3 shows the bias correction variables C_t over time. This fluctuates approximately around a base level ≈ 1, with an evident positive skewness due to the the upward peaks (jumps) while downward variations due to measurement noise appear to be much less pronounced and negligible.

13

Figure 2: Daily Realized Volatility
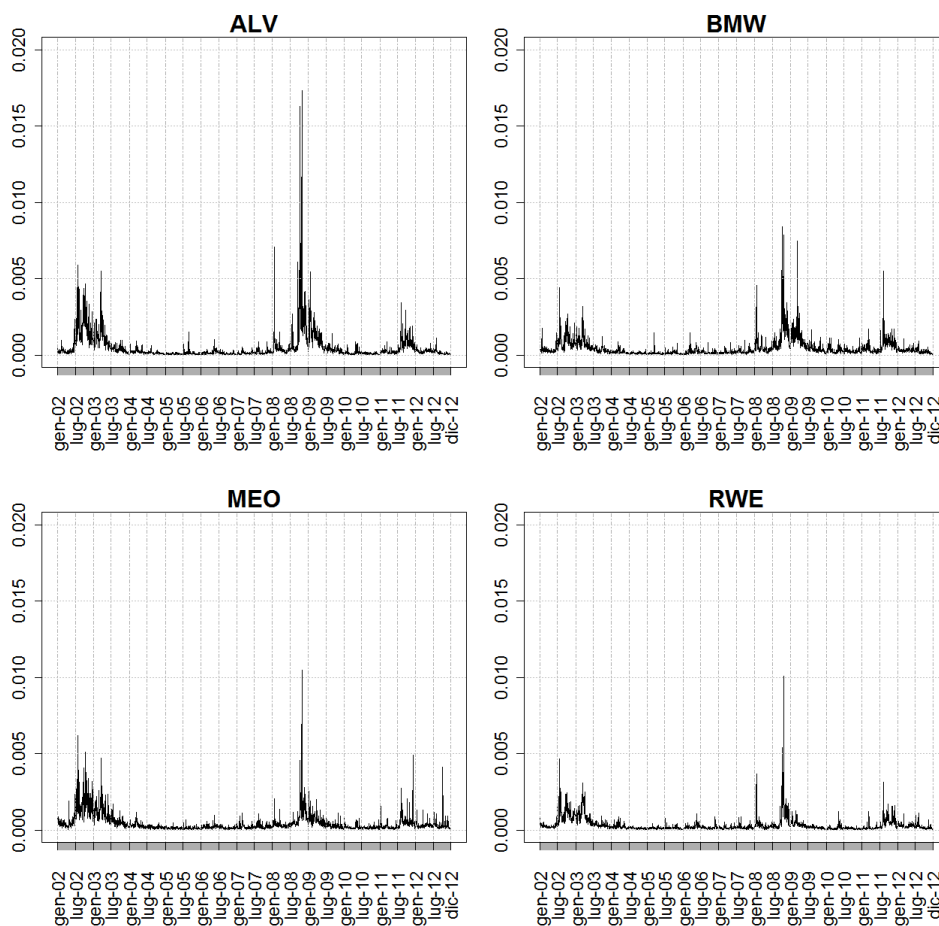
Daily Realized Volatility computed using a sampling frequency of 5 minutes. ALV (top-left), BMW (top-right), MEO (bottom-left) and RWE (bottom-right). Sample period $02/01/2002 - 27/12/2012$.

8 In-sample estimation

This section discusses the in-sample performance of the proposed models. The full data is employed here and focus is on the log-linear specification. For ease of exposition, sections 8.1 and 8.2 present the estimation results obtained for the jump free models (RGARCH, HRGARCH and TV-HRGARCH) and for the modified models in which the impact of jumps and of noise due to sampling variability are considered in the measurement equation (RGARCH*, HRGARCH* and TV-HRGARCH*).

Figure 3: Time series of daily bias correction variable C_t^{RV}
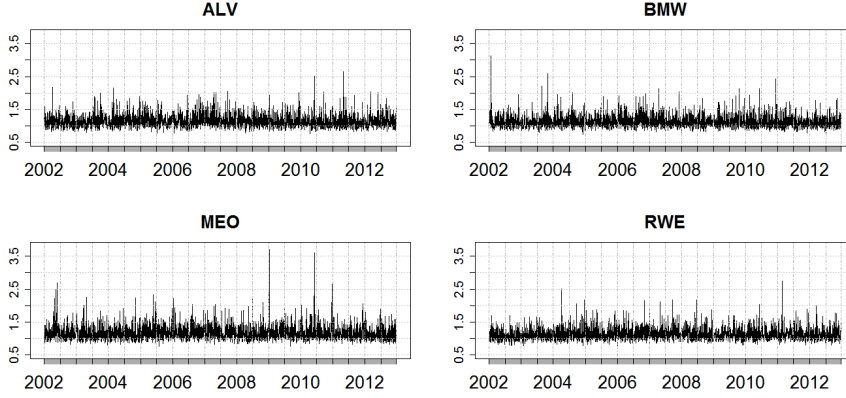
Daily bias correction variable $C_t = RV_t/medRV_t$ for the stocks ALV (top-left), BMW (top-right), MEO (bottom-left) and RWE (bottom-right) for the sample period 02/01/2002 – 27/12/2012.

8.1 Estimation results for RGARCH, HRGARCH and TV-HRGARCH

The estimation results for the HRGARCH and TV-HRGARCH models, and for comparison also the standard RGARCH model, are in Table 3. The top panel reports parameter estimates and robust standard errors, while the second panel shows corresponding values of the log-likelihood $\mathcal{L}(r, x)$ and partial log-likelihood $\ell(r)$, together with Bayesian Information Criterion (BIC), for the four analysed stocks.

The parameter ω is in most cases not significant, except for the TV-HRGARCH model, where the value of ω is considerably greater than other models, since it is influenced by the dynamics of the time-varying coefficient γ_t. The parameter β is always slightly higher for HRGARCH than RGARCH and TV-HRGARCH, whereas φ takes values closer to one both for the RGARCH and for RGARCH models which account for heteroskedasticity in the variance of the noise u_t, but also for the attenuation bias effects. These results are in line with the findings in Hansen et al. (2012), since $\varphi \simeq 1$ suggests the log-transformed realized measure x_t, is roughly proportional to the log-conditional variance. The parameters of the leverage function $\tau(z)$ are always significant, with τ_1 negative and τ_2 positive, as expected.

The parameter δ_1 is always positive and significant, at the 0.05 level. This means that, as expected, RQ_t positively affects the dynamics of the variance of the error term u_t in the measurement equation. This also implies that $\sigma_{u,t}^2$ tends to take on higher values in periods of turmoil and lower values when volatility tends to stay low. The coefficient γ, which summarizes the impact of the realized measure on future volatility,

15

Table 3: In-Sample Estimation Results for RGARCH, HRGARCH and TV-HRGARCH using 5-minutes Realized Volatility

	ALV			BMW			MEO			RWE		
	RGARCH	HRGARCH	TV-HRGARCH	RGARCH	HRGARCH	TV-HRGARCH	RGARCH	HRGARCH	TV-HRGARCH	RGARCH	HRGARCH	TV-HRGARCH
ω	-0.132 ⋆ (0.257)	0.011 ⋆ (0.094)	1.999 (0.386)	-0.127 ⋆ (0.111)	0.036 ⋆ (0.082)	1.611 (0.323)	-0.025 ⋆ (0.105)	-0.075 ⋆ (0.109)	1.435 (0.300)	-0.180 (0.085)	-0.368 (0.132)	0.940 (0.325)
γ	0.401 (0.037)	0.381 (0.031)	-	0.286 (0.027)	0.280 (0.024)	-	0.257 (0.026)	0.240 (0.025)	-	0.338 (0.028)	0.293 (0.027)	-
γ_0	-	-	0.490 (0.051)	-	-	0.410 (0.041)	-	-	0.339 (0.039)	-	-	0.389 (0.037)
γ_1	-	-	1.059 (0.238)	-	-	0.616 (0.151)	-	-	0.698 (0.201)	-	-	0.616 (0.204)
β	0.583 (0.033)	0.620 (0.028)	0.565 (0.035)	0.701 (0.023)	0.727 (0.023)	0.692 (0.026)	0.741 (0.025)	0.752 (0.024)	0.712 (0.029)	0.646 (0.026)	0.670 (0.024)	0.634 (0.028)
ξ	-0.094 ⋆ (0.610)	-0.478 (0.219)	-0.468 (0.211)	0.113 ⋆ (0.353)	-0.589 (0.254)	-0.574 ⋆ (0.375)	-0.284 ⋆ (0.375)	-0.164 ⋆ (0.421)	-0.185 ⋆ (0.285)	0.088 ⋆ (0.230)	0.608 ⋆ (0.474)	0.644 ⋆ (0.355)
φ	0.989 (0.073)	0.948 (0.027)	0.949 (0.026)	1.004 (0.043)	0.924 (0.031)	0.926 (0.046)	0.961 (0.045)	0.981 (0.050)	0.977 (0.034)	0.993 (0.027)	1.058 (0.056)	1.061 (0.042)
τ_1	-0.069 (0.008)	-0.069 (0.008)	-0.069 (0.007)	-0.023 (0.008)	-0.032 (0.008)	-0.032 (0.007)	-0.018 (0.009)	-0.024 (0.008)	-0.024 (0.008)	-0.038 (0.008)	-0.040 (0.007)	-0.040 (0.007)
τ_2	0.108 (0.006)	0.111 (0.006)	0.111 (0.006)	0.082 (0.006)	0.086 (0.005)	0.085 (0.005)	0.082 (0.007)	0.090 (0.006)	0.090 (0.006)	0.082 (0.005)	0.083 (0.005)	0.082 (0.005)
σ_u^2	0.185 (0.006)	-	-	0.171 (0.006)	-	-	0.189 (0.006)	-	-	0.164 (0.006)	-	-
δ_0	-	-0.403 ⋆ (0.258)	-0.393 (0.195)	-	0.322 ⋆ (0.320)	0.148 ⋆ (0.237)	-	-0.042 ⋆ (0.313)	-0.224 ⋆ (0.262)	-	0.283 ⋆ (0.351)	-0.133 ⋆ (0.283)
δ_1	-	0.163 (0.032)	0.166 (0.024)	-	0.270 (0.041)	0.251 (0.030)	-	0.209 (0.039)	0.189 (0.033)	-	0.261 (0.043)	0.212 (0.034)
$\ell(r)$	7605.335	7606.093	7608.968	7466.828	7468.052	7466.639	7460.021	7460.750	7466.254	7986.446	7988.292	7990.691
$\mathcal{L}(r,x)$	6001.795	6030.413	6061.639	5970.296	6024.041	6050.956	5825.634	5858.409	5887.817	6567.031	6614.276	6632.093
BIC	-11940.116	-11989.418	-12043.935	-11877.120	-11976.674	-12022.571	-11587.795	-11645.411	-11696.293	-13070.588	-13157.144	-13184.844

In-sample parameter estimates for the full sample period 02 January 2002 - 27 December 2012 using 5-minutes Realized Volatility. ⋆: parameter not significant at 5%. $\ell(r)$: partial log-likelihood. $\mathcal{L}(r,x)$: log-likelihood. BIC: Bayesian Information Criterion. Robust standard errors are reported in small font under the parameter values.

ranges from ≈ 0.25 to 0.40 for both RGARCH and HRGARCH. For the TV-HRGARCH this effect is explained, in an adaptive fashion, by the time varying γ_t, which is a function of the past noise variance $\sigma^2_{u,t-1}$. Since γ_1 is always positive and $log(x_t)$ is negative, when the lagged variance of the error term of the realized measure $\sigma^2_{u,t-1}$ is high, the impact of the lagged log-transformed realized measure $log(x_{t-1})$ on $log(h_t)$ will be negative and lower than what would have been implied by the same value of $log(x_{t-1})$ in correspondence of a lower value of $\sigma^2_{u,t-1}$. Said differently, the impact of x_{t-1} on h_t will be down-scaled towards zero when $\sigma^2_{u,t-1}$ increases. Equivalently, variations in h_t ($\nabla h_t = h_t - h_{t-1}$) will be negatively correlated with the values of γ_t and $\sigma^2_{u,t-1}$. These results are in line with the recent findings of Bollerslev et al. (2016).

In terms of fit to the data, from the second panel of Table 3 it emerges that the TV-HRGARCH features the lowest value of the BIC for the four examined stocks. Looking at $\mathcal{L}(r, x)$, the TV-HRGARCH model improves over the standard RGARCH[2] in all series, as expected; the improvement is similar, but less pronounced, from the HRGARCH to the RGARCH. Simple likelihood ratio tests reveal that both HRGARCH and TV-HRGARCH give rise to significant improvements over the benchmark RGARCH model in all series. Further, the partial returns log-likelihood $\ell(r)$ component also shows all positive differences over RGARCH, though these are smaller in scale, where the TV_HRGARCH prevails in three cases out of 4 (for BMW the HRGARCH is preferred).

As a robustness check, all the models are re-estimated using the 5-minute Realized Kernel as volatility proxy. The results and conclusions reported above are highly similar to those obtained using the 5-minute RV, and so not reported to save space.

Figure 4 compares the constant variance σ^2_u estimated by RGARCH with the time-varying variance $\sigma^2_{u,t}$ given by the HRGARCH model estimated using the 5 minute Realized Volatility[3]. For the four analysed stocks the trend of $\sigma^2_{u,t}$ follows the dynamics of the realized measure, being higher in turbulent periods and lower in calmer periods, while the constant variance σ^2_u estimated within the RGARCH (red line in the plots) is approximately equal to the average level of the time-varying variance of the measurement noise.

Figure 5 displays the time plot of the γ_t coefficient for the four

---

[2]Differently from Hansen et al. (2012) positive values are obtained for the log-likelihood. This is mainly due to the fact that they use percentage log-returns, which approximately fall in the range (-30, 30). It follows that the conditional variances are often above 1, returning positive log-variances that multiplied by -1 in the log-likelihood, explaining the comparatively large negative log-likelihoods that they typically get.

[3]We do not report results for models using the RK as volatility proxy, since these are virtually identical to those reported here for the 5 minute RV.

Figure 4: Constant versus time-varying variance of the noise u_t of the HRGARCH fitted using the 5 minute RV
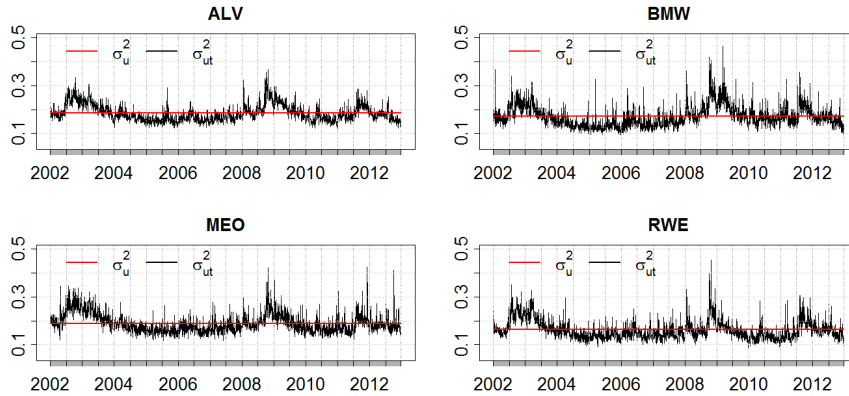
The Figure shows the constant variance σ_u^2 (red-line) estimated with RGARCH together with the time-varying variance $\sigma_{u,t}^2$ (black-line) estimated with HRGARCH. Both models have been fitted taking the 5-minutes RV as volatility proxy. Sample period 02 January 2002 - 27 December 2012.

Figure 5: Time-varying coefficient γ_t given by the TV-HRGARCH model
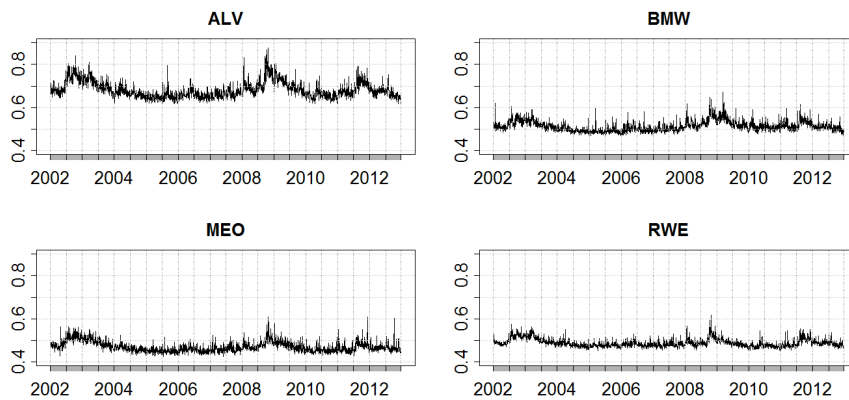
The Figure shows the time-varying coefficient $\gamma_t = \gamma_0 + \gamma_1 \sigma_{u,t-1}^2$ for the sample period 02 January 2002 - 27 December 2012.

considered stocks. It is evident that when the variance of the measurement error is high, γ_t is also high, leading to a less substantial increase of h_t compared to days in which, ceteris paribus, $\sigma_{u,t}^2$ is low and the realized measure provides a stronger more reliable signal. Further, the value of γ_t tends to be higher than the value of the time invariant γ estimated within the RGARCH and HRGARCH models.

8.2 Estimation results for RGARCH*, HRGARCH* and TV-HRGARCH*

The estimation results for RGARCH*, HRGARCH* and TV-HRGARCH* are reported in Table 4. Since jumps are accounted for in these models, in equation (14) RQ_t is replaced with the jump-robust estimator $medRQ_t$.

Results for ω are again not significant in most cases. The estimates of φ, τ_1, τ_2 are also apparently unaffected by using x_t^*. The coefficient η, related to the bias correction variable C_t, is always positive and statistically significant and its values ranges from 0.25 to 0.4 for RGARCH* and HRGARCH* and, slightly higher, from 0.3 to 0.47 for the TV-HRGARCH*. Given that $0 < \eta < 1$, the impact of the rescaled realized measure $log(x_t/C_t^\eta)$ is determined along the same lines as for $log(x_t)$. Therefore, the amount of smoothing is not arbitrarily chosen, but data driven through the estimated parameter η. A quite interesting aspect is that for RGARCH* the variance $\sigma_{u^*}^2$ of the measurement equation error u_t^* is reduced over that for the RGARCH model for each series, providing evidence of an improved goodness of fit in the modified measurement equation and efficiency of the bias corrected realized measure x^*t. Therefore, correcting the upward and downward bias through the variable C_t in the measurement equation reduces the variability and thus improves the efficiency of the realized estimator.

For HRGARCH* and TV-HRGARCH* the parameter δ_1 is always statistically significant and positive, giving empirical confirmation to the intuition that the variance of the measurement error is time-varying and in accordance with the asymptotic theory suggesting that this is positively related to the IQ.

The impact of the past realized measure on future volatility is increased, with the exception of the RGARCH* for the asset MEO, by the introduction of the bias correction variable C_t, as the coefficient γ always takes on higher values; this confirms the idea that accounting for jumps further reduces the attenuation bias effect on γ. Also, for the TV-HRGARCH* model the coefficient γ_1 is positive (even if lower than the ones showed for the TV-HRGARCH), confirming that more weight is given to the realized measure when it is more accurately measured. Thus, this class of models provide stronger persistence when the measurement error is relatively low.

The second panel of Table 4 shows that, even in this framework,

Table 4: In-Sample Estimation Results for RGARCH*, HRGARCH* and TV-HRGARCH* using 5-minute Realized Volatility

	ALV			BMW			MEO			RWE		
	RGARCH*	HRGARCH*	TV-HRGARCH*	RGARCH*	HRGARCH*	TV-HRGARCH*	RGARCH*	HRGARCH*	TV-HRGARCH*	RGARCH*	HRGARCH*	TV-HRGARCH*
ω	-0.026 ⋆ (0.098)	0.012 ⋆ (0.106)	1.537 (0.448)	0.104 ⋆ (0.072)	0.041 ⋆ (0.101)	1.066 (0.295)	-0.123 ⋆ (0.097)	-0.065 ⋆ (0.103)	1.226 (0.313)	-0.289 (0.129)	-0.375 (0.125)	0.294 (0.129)
γ	0.406 (0.032)	0.394 (0.032)	-	0.342 (0.028)	0.287 (0.026)	-	0.249 (0.024)	0.248 (0.024)	-	0.362 (0.032)	0.296 (0.026)	-
γ_0	-	-	0.449 (0.053)	-	-	0.349 (0.037)	-	-	0.314 (0.040)	-	-	0.339 (0.031)
γ_1	-	-	0.884 (0.232)	-	-	0.468 (0.139)	-	-	0.633 (0.193)	-	-	0.308 (0.101)
β	0.588 (0.031)	0.605 (0.030)	0.579 (0.033)	0.673 (0.027)	0.720 (0.023)	0.707 (0.025)	0.735 (0.024)	0.744 (0.024)	0.721 (0.027)	0.610 (0.030)	0.665 (0.024)	0.654 (0.026)
ξ	-0.324 ⋆ (0.212)	-0.472 ⋆ (0.242)	-0.474 (0.202)	-0.701 (0.196)	-0.574 ⋆ (0.317)	-0.521 ⋆ (0.383)	0.103 ⋆ (0.357)	-0.200 ⋆ (0.379)	-0.218 ⋆ (0.293)	0.247 ⋆ (0.354)	0.633 ⋆ (0.439)	0.662 ⋆ (0.339)
φ	0.967 (0.026)	0.953 (0.030)	0.953 (0.025)	0.909 (0.024)	0.928 (0.039)	0.935 (0.047)	1.014 (0.043)	0.981 (0.045)	0.980 (0.034)	1.014 (0.041)	1.064 (0.051)	1.067 (0.040)
τ_1	-0.068 (0.008)	-0.069 (0.007)	-0.069 (0.007)	-0.030 (0.008)	-0.031 (0.008)	-0.031 (0.008)	-0.016 ⋆ (0.009)	-0.020 (0.008)	-0.021 (0.008)	-0.041 (0.008)	-0.040 (0.007)	-0.041 (0.007)
τ_2	0.108 (0.006)	0.110 (0.006)	0.110 (0.006)	0.080 (0.005)	0.084 (0.005)	0.083 (0.005)	0.081 (0.007)	0.088 (0.006)	0.088 (0.006)	0.079 (0.005)	0.081 (0.005)	0.081 (0.005)
η	0.399 (0.050)	0.402 (0.049)	0.465 (0.051)	0.278 (0.054)	0.243 (0.052)	0.309 (0.053)	0.360 (0.051)	0.363 (0.049)	0.430 (0.048)	0.279 (0.050)	0.272 (0.049)	0.309 (0.051)
σ_{u*}^2	0.180 (0.006)	-	-	0.169 (0.006)	-	-	0.185 (0.006)	-	-	0.160 (0.005)	-	-
δ_0	-	-0.551 (0.255)	-0.511 (0.256)	-	-0.115 ⋆ (0.336)	-0.135 ⋆ (0.298)	-	-0.267 ⋆ (0.295)	-0.322 ⋆ (0.299)	-	0.081 ⋆ (0.333)	-0.151 ⋆ (0.142)
δ_1	-	0.145 (0.031)	0.151 (0.031)	-	0.212 (0.042)	0.211 (0.038)	-	0.180 (0.037)	0.175 (0.037)	-	0.235 (0.040)	0.207 (0.042)
$\ell(r)$	7606.606	7606.523	7610.151	7467.651	7468.990	7467.012	7461.572	7461.340	7467.222	7989.563	7988.720	7988.765
$\mathcal{L}(r,x)$	6039.755	6062.608	6081.077	5989.603	6022.243	6033.203	5857.461	5881.498	5900.556	6586.725	6622.973	6627.373
BIC	-12008.103	-12045.874	-12074.879	-11907.799	-11965.144	-11979.131	-11643.514	-11683.655	-11713.837	-13102.043	-13166.604	-13167.470

In-sample parameter estimates for the full sample period 02 January 2002 - 27 December 2012 using 5-minutes Realized Volatility. ⋆: parameter not significant at 5%. $\ell(r)$: partial log-likelihood. $\mathcal{L}(r,x)$: log-likelihood. BIC: Bayesian Information Criterion. Robust standard errors are reported in small font under the parameter values.

the model with time-varying persistence provides the lowest BIC values. Focusing on $\mathcal{L}(r, x)$, the TV-HRGARCH* is still the dominant model since it maximizes the log-likelihood. The partial returns log-likelihood component $\ell(r)$ for ALV and MEO the TV-HRGARCH* proves more powerful and for BMW this holds for the HRGARCH*, while for RWE the RGARCH* prevails.

Estimation results are also obtained using the Realized Kernel as volatility proxy that, as in the previous case, are virtually identical to those based on 5-minutes Realized Volatility, and so not reported here.

Table 5: BIC in-sample comparison

		ALV	BMW	MEO	RWE
RV	RGARCH	-11940.116	-11877.120	-11587.795	-13070.588
	HRGARCH	-11989.418	-11976.674	-11645.411	-13157.144
	TV-HRGARCH	-12043.935	**-12022.571**	-11696.293	**-13184.844**
	RGARCH*	-12008.103	-11907.799	-11643.514	-13102.043
	HRGARCH*	-12045.874	-11965.144	-11683.655	-13166.604
	TV-HRGARCH*	**-12074.879**	-11979.131	**-11713.837**	-13167.470
RK	RGARCH	-11910.095	-11853.344	-11521.500	-13041.133
	HRGARCH	-11963.895	-11947.007	-11597.809	-13128.268
	TV-HRGARCH	-12013.475	**-11986.876**	-11637.261	**-13148.254**
	RGARCH*	-11992.565	-11890.365	-11612.761	-13085.465
	HRGARCH*	-12031.170	-11942.638	-11653.035	-13145.653
	TV-HRGARCH*	**-12060.500**	-11956.648	**-11682.216**	-13145.546

BIC values for the analysed models using the 5-min RV (first panel) and the 5-min RK (second panel). Best models are reported in **bold**.

To clarify the results, Table 5 reports the comparison of the BIC values for the estimated models according to RV (first panel) and RK (second panel): the models with optimum BIC are exactly the same when using each realized measure. Interestingly, for ALV and MEO, which have the highest number of jumps, the BIC is minimized by the TV-HRGARCH*, whereas for BMW and RWE the TV-HRGARCH model is optimal. This scenario is repeated when using the log-likelihood and RV (see Table 6); highly similar results are obtained, but not reported, using the RK instead. The partial returns likelihood, $\ell(r)$, reported in small font under each $\mathcal{L}(r, x)$ value has the best results achieved by models allowing for bias effects in the measurement equation, in particular HRGARCH* for BMW and TV-HRGARCH* for ALV and MEO, while for RWE the TV-HRGARCH

prevails, using both the realized measures considered in our analysis.

Summarizing: the introduction of heteroskedasticity and time-varying persistence, as well as the bias correction for jumps and measurement errors, has positive effects on the estimated volatility, since the log-likelihood and the partial log-likelihood of the returns component tend to be maximized by this class of models. Consequently, the proposed models show notable improvements in volatility fitting over standard RGARCH, in sample.

Table 6: Log-likelihood and partial log-likelihood using 5-minutes RV

	ALV	BMW	MEO	RWE
RGARCH	6001.795	5970.296	5825.634	6567.031
	7605.335	7466.828	7460.021	7986.446
HRGARCH	6030.413	6024.041	5858.409	6614.276
	7606.093	7468.052	7460.750	7988.292
TV-HRGARCH	6061.639	**6050.956**	5887.817	**6632.093**
	7608.968	7466.639	7466.254	7990.691
RGARCH*	6039.755	5989.603	5857.461	6586.725
	7606.606	7467.651	7461.572	7989.563
HRGARCH*	6062.608	6022.243	5881.498	6622.973
	7606.523	7468.990	7461.340	7988.720
TV-HRGARCH*	**6081.077**	6033.203	**5900.556**	6627.373
	7610.151	7467.012	7467.222	7988.765

In **bold** models maximizing the log-likelihood $\mathcal{L}(r, x)$. In blue models maximizing the partial log-likelihood $\ell(r)$, reported in small font under the corresponding $\mathcal{L}(r, x)$ value.

9   Out-of-sample Analysis

In this section the out-of-sample predictive ability of the models, estimated in sections 8.1 and 8.2, is assessed via a rolling window forecasting exercise, using a window of 1500 days. Furthermore, the framework is extended to models that allow only for "significant" jumps. The out-of-sample period starts 02 January 2008 and includes 1270 daily observations, covering the credit crisis and the turbulent period from November 2011 to the beginning of 2012.

In order to assess the forecasting performance of the proposed models the predictive log-likelihood and the QLIKE loss function are employed. Further, the Model Confidence Set (MCS) of Hansen et al. (2011) is used to evaluate the comparative predictive ability of all the models, considering confidence levels 75% and 90%; the T_{max} statistic, based on a block-bootstrap procedure of 5000 re-samples, is employed to test the hypothesis of equal predictive ability, where the optimal block length has been chosen through the method described in Patton et al. (2009).

9.1 Testing for significant jumps

On the basis of theoretical results in Barndorff-Nielsen and Shephard (2006) the following test statistic can be used to identify "significant" jumps in a given price series:

$$Z_{TPQ_t} = \frac{RV_t - BPV_t}{\sqrt{(\theta - 2)\frac{1}{M}TPQ_t}} \ , \tag{25}$$

where BPV and TPQ are the realized bipower variation and realized tripower quarticity, respectively, while $\theta = (\pi/2)^2 + \pi - 3 \simeq 2.61$, M is the sampling frequency.

However, the simulation-based evidence reported in Huang and Tauchen (2005) suggests that the Z_{TPQ_t} statistic defined in (25) tends to over-reject the null hypothesis of no jumps for large critical values. These findings suggest the use of the ratio jump statistic:

$$Z^R_{TPQ_t} = \frac{(RV_t - BPV_t)/RV_t}{\sqrt{(\theta - 2)\frac{1}{M}\frac{TPQ_t}{BPV_t^2}}} \tag{26}$$

where $Z^R_{TPQ_t}$ is very closely approximated by a standard normal distribution as $M \to \infty$.

In our empirical application BPV and TPQ estimators are replaced by $medRV$ and $medRQ$, respectively and consequently $\theta = 2.96$ (Andersen et al., 2012), namely

$$Z_{J_t} = \frac{(RV_t - medRV_t)/RV_t}{\sqrt{0.96\frac{1}{M}\frac{medRQ_t}{medRV_t^2}}} \tag{27}$$

In this context, in order to consider the further possible scenario in which only "significant" jumps are modelled, the measurement equation is specified as

$$log(x_t) = \xi + \varphi \, log(h_t) + \eta \, log(C_t) \, I(Z_{J_t}) + \tau(z_t) + \tilde{u}_t \tag{28}$$

where $I\left(Z_{J_t}\right)$ is an indicator function equal to 1 if the null hypothesis of no jumps is rejected and 0 otherwise. Therefore in jumps-free periods the measurement equation x_t corresponds to that of the standard RGARCH. The models within this class are denoted as RGARCH-J, HRGARCH-J and TV-HRGARCH-J.

9.2 Forecasting comparison

Before summarizing the out-of-sample evidence provided by the predictive log-likelihood and the QLIKE loss function, it is interesting to look at Table 7, illustrating the summary statistics of the C_t ratio, only for days in which jumps are significant, based on the Z_{J_t} statistic in (27) at the 1% significance level. Since models including the bias correction variable C_t in the measurement equation allow for both small and large jump variations, focus is on a cut-off of 0.99 to test the presence of jumps in order to consider only days characterised by highly significant jumps[4].

---

[4]The proportion of jumps is a function of the significance level α that is employed. Although the use of $\alpha = 0.05$ identifies more significant jumps, the results provided by the out-of-sample analysis are practically the same.

Table 7: Summary statistics of Ct for significant jumps at 1% level

		ALV	BMW	MEO	RWE
C_t^{RV}	Prop.	0.054	0.041	0.069	0.051
	Min.	1.358	1.352	1.371	1.362
	1Qu.	1.482	1.507	1.475	1.494
	Med	1.547	1.596	1.547	1.547
	Mean	1.616	1.635	1.656	1.595
	3Qu.	1.646	1.710	1.712	1.686
	Max.	2.655	2.426	3.704	2.180
	S.dev	0.232	0.205	0.371	0.166
		ALV	BMW	MEO	RWE
C_t^{RK}	Prop.	0.047	0.036	0.064	0.043
	Min.	1.355	1.389	1.392	1.357
	1Qu.	1.478	1.505	1.473	1.493
	Med	1.549	1.594	1.551	1.549
	Mean	1.620	1.651	1.661	1.602
	3Qu.	1.657	1.716	1.720	1.697
	Max.	2.654	2.424	3.670	2.161
	S.dev	0.243	0.207	0.377	0.169

Summary statistics for significant jumps determined by the bias correction variable C_t using 5-min RV (first panel) and 5-min RK (second panel) at the significance level of $\alpha = 0.01$ according to the test statistic Z_{J_t} specified in (27). Prop.: proportion of days with significant jumps.

The first row of the top and bottom panel in Table 7 reveals that MEO and ALV are the assets showing the highest proportion of forecast sample days with significant jumps, whereas BMW is affected by relatively fewer jumps. However, BMW shows an average value of C_t higher than that for ALV and RWE, where the latter exhibits the lowest mean, and standard deviation, meaning that the impact of jumps is less marked than other stocks (both using RV and RK in estimation), even if it presents a proportion of significant jumps very close to that for ALV.

Table 8: Predictive log-likelihood and predictive partial log-likelihood using 5-minute RV.

	ALV	BMW	MEO	RWE
RGARCH	2568.420 3334.337	2430.166 3131.385	2444.677 3199.957	2887.809 3562.323
HRGARCH	2583.630 3334.156	2451.359 3132.354	2488.939 3200.557	2925.647 3561.528
TV-HRGARCH	2602.690 3336.313	**2466.983** 3130.976	2505.271 3206.451	2932.639 3565.076
RGARCH*	2583.520 3334.529	2439.665 3131.750	2468.083 3200.824	2907.078 3562.610
HRGARCH*	2598.877 3334.595	2453.199 3132.383	2502.194 3201.749	**2934.486** 3561.946
TV-HRGARCH*	**2607.071** 3336.773	2457.562 3131.430	**2517.105** 3207.527	2934.071 3562.597
RGARCH-J	2573.770 3333.654	2433.761 3131.507	2454.004 3200.087	2894.117 3562.042
HRGARCH-J	2587.530 3333.760	2447.852 3132.045	2488.987 3200.776	2924.222 3561.484
TV-HRGARCH-J	2604.238 3334.679	2452.133 3131.239	2495.016 3204.609	2921.216 3561.791

The table reports the predictive log-likelihood and the predictive partial log-likelihood in smaller font underneath. In **bold** the preferred model according to predictive log-likelihood. In blue the best model in terms of predictive partial log-likelihood of returns. Out-of-sample period 02 January 2008 – 27 December 2012.

The first criterion for assessing predictive accuracy is, as in Hansen et al. (2012), the predictive log-likelihood, given for time $t + 1$ by:

$$\hat{\ell}(r,x)_{t+1} = -\frac{1}{2}\left[log(2\pi) + log(\hat{h}_{t+1}) + \frac{r_{t+1}^2}{\hat{h}_{t+1}}\right] - \frac{1}{2}\left[log(2\pi) + log(\hat{\sigma}_u^2) + \frac{u_{t+1}^2}{\hat{\sigma}_u^2}\right]. \quad (29)$$

Subsequently, the aggregated predictive log-likelihood is computed by summing the density estimates for each day in the forecast period.

Table 8 shows the values of the predictive log-likelihood, and of the predictive partial log-likelihood (reported in smaller font under each

26

predictive log-likelihood value), corresponding to all models employed. The models that allow for upward and downward bias, i.e. those including the "full" bias correction variable C_t in the measurement equation, maximize both the predictive log-likelihood and partial predictive log-likelihood in 3 cases out of 4. In particular for ALV and MEO, the stocks with the highest proportion of jumps, the specification that allows for heteroskedasticity and time-varying persistence, TV-HRGARCH*, is favoured by both measures; for RWE the HRGARCH* is favoured by predictive likelihood, but TV-HRGARCH is favoured by partial predictive likelihood. Finally, for BMW the HRGARCH* is favoured by partial predictive likelihood, while the TV-HRGARCH is favoured by predictive likelihood. Though not reported, results when employing the RK instead of RV, favour exactly the same models in all cases. In all cases one of the models proposed in this paper is favoured.

Table 9: MCS p-values of predictive log-likelihood using 5-min RV as volatility proxy. For each stock in red models $\in$ 75% MCS and in blue models $\in$ 90% MCS. The p-values refer to the denied predictive log-likelihoods.

	MCS p-values			
	ALV	BMW	MEO	RWE
RGARCH	0.0088	0.0258	0.0006	0.0098
HRGARCH	0.0182	0.0934	0.0030	0.2584
TV-HRGARCH	0.7772	1.0000	0.1118	0.9412
RGARCH*	0.2371	0.0792	0.0014	0.0714
HRGARCH*	0.3718	0.0970	0.1118	1.0000
TV-HRGARCH*	1.0000	0.1708	1.0000	0.9412
RGARCH-J	0.0088	0.0418	0.0010	0.0202
HRGARCH-J	0.0088	0.0522	0.0022	0.0830
TV-HRGARCH-J	0.7772	0.0934	0.0070	0.0268

In order to evaluate if the differences in the maximized log-likelihood of the considered models are statistically relevant, the Model Confidence Set approach of Hansen et al. (2011) is employed. Table 9 shows the MCS p-values associated to the predictive log-likelihood (multiplied by -1), where the 5-min RV is used. Interestingly, for MEO (which has the largest proportion of jumps) the only model coming into the MCS at the more restrictive 0.75 confidence level is the TV-HRGARCH*, while

the HRGARCH* and the TV-HRGARCH enter at 0.90 level. On the other hand, for BMW (which has the lowest proportion of jumps) the model that allows for both heteroskedasticity and time-varying persistence uniquely falls in the 75% MCS, with its bias corrected version, the TV-HRGARCH*, entering into the 90% MCS. For the asset ALV all the time-varying specifications enter into the *set of superior models* at the 0.75 level, as well as the HRGARCH*, while the RGARCH* falls into the MCS at the 0.90 level. Finally, for RWE four models are included into the 75% MCS: the HRGARCH and TV-HRGARCH, together with the corresponding counterparts corrected for jumps and measurement errors. The results, not reported to save space, are highly similar when employing the RK insytead of the RV in all models. Therefore, the standard RGARCH never enters into the set of superior models (as in Table 9), nor do the models capturing only significant jumps.

As a further criterion for assessing and comparing the forecasting accuracy of the fitted models, the QLIKE loss function is employed. This choice is motivated by the consideration that, compared to other robust alternatives, this loss function is more powerful in rejecting poorly performing predictors (Liu et al., 2015). The QLIKE loss has been computed according to the formula:

$$QLIKE = \frac{1}{T} \sum_{t=1}^{T} log(\hat{h}_t) + \frac{x_t}{\hat{h}_t} \tag{30}$$

where the 5-minute RV and RK are again chosen as volatility proxy[5]. The significance of differences in the QLIKE values across different models is again tested by means of the Model Confidence Set (MCS) approach of Hansen et al. (2011).

---

[5]Results are highly similar also when employing the 5-min $medRV$ as volatility proxy

Table 10: Average values of QLIKE loss using 5-min RV as volatility proxy (top) and MCS p-values (bottom). For each stock in **bold** minimum loss, in red model $\in$ 75% MCS and in blue model $\in$ 90% MCS

QLIKE				
	ALV	BMW	MEO	RWE
RGARCH	-6.9639	-6.6661	-6.9402	-7.2539
HRGARCH	-6.9634	-6.6650	-6.9408	-7.2533
TV-HRGARCH	-6.9667	**-6.6687**	**-6.9472**	**-7.2580**
RGARCH*	-6.9645	-6.6673	-6.9419	-7.2543
HRGARCH*	-6.9641	-6.6661	-6.9428	-7.2540
TV-HRGARCH*	**-6.9668**	-6.6673	-6.9465	-7.2548
RGARCH-J	-6.9644	-6.6660	-6.9408	-7.2538
HRGARCH-J	-6.9641	-6.6653	-6.9413	-7.2533
TV-HRGARCH-J	-6.9658	-6.6656	-6.9434	-7.2534

MCS p-values				
	ALV	BMW	MEO	RWE
RGARCH	0.0634	0.0856	0.0158	0.0478
HRGARCH	0.0688	0.0406	0.0288	0.0462
TV-HRGARCH	0.9004	1.0000	1.0000	1.0000
RGARCH*	0.1546	0.3380	0.0612	0.0478
HRGARCH*	0.1254	0.1574	0.2654	0.0478
TV-HRGARCH*	1.0000	0.3380	0.7216	0.0478
RGARCH-J	0.1254	0.0738	0.0290	0.0478
HRGARCH-J	0.1254	0.0534	0.0556	0.0462
TV-HRGARCH-J	0.4498	0.0678	0.0882	0.0462

Observing the average values of the QLIKE loss function for each considered model, shown in top panel of Table 10, the lowest value is obtained by the TV-HRGARCH model, except for ALV where it is the TV-HRGARCH*. Moreover, in order to assess the significance of differences across different models, the bottom panel of Table 10 displays the MCS p-values, representing the probability that a particular model belongs to the MCS at the considered confidence level of 0.75 and 0.90. The MCS

p-values show that the TV-HRGARCH is always included in the MCS, at both confidence levels; for RWE, which has the lowest mean and maximum value of C_t for significant jumps (Table 7), it is the only model in the MCS at the 0.75 level. The standard RGARCH never comes into the set of superior models, nor does the HRGARCH, for all examined assets. Within the class of models which allow for upward and downward bias effects in the measurement equation, the RGARCH* is included into the MCS at the level 0.75 for BMW, and at the level 0.90 for ALV, whereas the HRGARCH* falls into the 75% MCS only for MEO and into the 90% MCS for ALV and BMW. The TV-HRGARCH* is always included in the set of superior models at the more restrictive 0.75 level, excepting for RWE. On the other hand, the models considering significant jumps come into the set of superior models only for the stock ALV and in particular the RGARCH-J and the HRGARCH-J enter into the 90% MCS, while the TV-HRGARCH-J is included in the 75% MCS.

As a robustness check, the above results are repeated using instead the 5-minute RK as volatility proxy: the results are highly similar and qualitatively the same conclusions are reached; thus these results are not reported here.

10   Conclusion

A new class of realized volatility models, accounting for heteroskedasticity in measurement error and temporal variation in the volatility persistence is proposed. The latter novelty explicitly reduces the magnitude of the attenuation bias. Further, the presence of jumps and measurement errors arising from the sampling variability lead to extending the proposed modelling approach, through the use of a bias correction variable controlling the amount of upward (jumps) and downward (measurement errors) bias, in a data-driven fashion.

The empirical analysis illustrates that the proposed modelling approach outperforms the standard Realized GARCH, both for in-sample fit and in out-of-sample forecasting of volatility and return distributions. The models allowing for both heteroskedasticity and time-varying persistence show a higher predictive capability, in maximizing the predictive log-likelihood and in minimizing the QLIKE loss function, and always entering into the Model Confidence Set at the different confidence levels considered.

The time-varying model regulating the bias, generated by jumps and measurement errors, better reproduces the dynamic properties of volatility of the stocks characterised by a high proportion of jumps, while models that account only for significant jumps show a relatively lower predictive performance than the competitors. Furthermore, the use of different

realized volatility measures confirms the robustness of our findings.

The models developed in this paper can easily be used in other financial applications, such as forecasting of Value at Risk and Expected Shortfall, where the latter is currently receiving a lot of attention in the literature.

References

Andersen, T. G., T. Bollerslev, and F. X. Diebold (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., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association 96*(453), 42–55.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal 12*(3).

Barndorff-Nielsen, O. E. and N. Shephard (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.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics 4*(1), 1–30.

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

Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis 51*(4), 2232–2245.

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

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

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

Hansen, P. and A. Lunde (2011). Forecasting volatility using high frequency data. *The Oxford Handbook of Economic Forecasting, Oxford: Blackwell*, 525–556.

Hansen, P. R. and Z. Huang (2016). Exponential garch modeling with realized measures of volatility. *Journal of Business & Economic Statistics 34*(2), 269–287.

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

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

Huang, X. and G. Tauchen (2005). The relative contribution of jumps to total price variance. *Journal of financial econometrics 3*(4), 456–499.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications 119*(7), 2249–2276.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Patton, A., D. N. Politis, and H. White (2009). Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews 28*(4), 372–375.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics 25*(2), 197–231.

Shephard, N. and D. Xiu (2016). Econometric analysis of multivariate realised qml: efficient positive semi-definite estimators of the covariation of equity prices.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.