

Article

Modeling Realized Variance with Realized Quarticity

Hiroyuki Kawakatsu

Business School, Dublin City University, Dublin 9, D09 Dublin, Ireland; hiroyuki.kawakatsu@dcu.ie

Abstract: This paper proposes a model for realized variance that exploits information in realized quarticity. The realized variance and quarticity measures are both highly persistent and highly correlated with each other. The proposed model incorporates information from the observed realized quarticity process via autoregressive conditional variance dynamics. It exploits conditional dependence in higher order (fourth) moments in analogy to the class of GARCH models exploit conditional dependence in second moments.

Keywords: realized variance; realized quarticity; volatility of volatility

1. Introduction

The commonly used measure of financial risk, the conditional variance, of an asset return is not directly observable. The large class of observation-driven GARCH type or parameter-driven stochastic volatility models use asset return data to estimate the conditional variance process. A more recent alternative approach exploits the availability of intraday high-frequency data and treats the realized variance (*RV*) process as observable data. This approach was popularized by the HAR (heterogeneous autoregressive) specification of Corsi [1].

Understanding the time variation in financial risk and the ability to predict its future course is of primary interest for financial practitioners and regulators who need to manage financial risk. In the extensive literature that examines the conditional dynamics of asset return second moments, some have considered utilizing information contained in higher order moments such as skewness and kurtosis [2–4]. This paper contributes to this literature by examining the information in the observed fourth-moment process to predict the realized variance *RV*.

The proposed model exploits two empirical features of the *RV* and the realized quarticity (*RQ*) processes. First, both processes are highly persistent with slowly decaying autocorrelations. The early literature modeled the slow decay in volatility as a long memory process with Hurst exponent $H > 1/2$ [5–7]. Recently, Gatheral et al. [8] have suggested a rough volatility model with $H < 1/2$. As they note, though a rough volatility model is not strictly a long memory process, their autocorrelation is statistically difficult to distinguish from that of a long memory process. For the fourth-moment process Huang et al. [9] find positive but weak dependence in the volatility of volatility measure from options on VIX. Da Fonseca and Zhang [10] document roughness in volatility of volatility with $H < 1/2$.

The second empirical feature the proposed model exploits is the high correlation between *RV* and *RQ* processes. A positive correlation between volatility and volatility of volatility is also documented in Da Fonseca and Zhang [10]. This correlation motivates the use of information in *RQ* to model *RV*. As the variance of variance is the fourth moment and this fourth moment is persistent (first empirical feature), the proposal is to model the conditional variance of the realized variance with GARCH-type dynamics that can capture the persistence. The proposed model is a higher moment extension of the realized GARCH model of Hansen et al. [11].

The common approach to model the persistence in *RV* is to use a restricted distributed lag model of Corsi [1]. Bollerslev et al. [12] extend the HAR specification to account for



Citation: Kawakatsu, H. Modeling Realized Variance with Realized Quarticity. *Stats* **2022**, *5*, 856–880. <https://doi.org/10.3390/stats5030050>

Academic Editor: Magda Sofia Valério Monteiro

Received: 11 August 2022

Accepted: 5 September 2022

Published: 7 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

possible measurement error in RV as an estimator of the integrated variance IV . Their HARQ specification allows the parameters of the HAR model to be time-varying and depend on \sqrt{RQ} . Bucchieri and Corsi [13] consider a HAR model with time-varying parameters in a parameter-driven state-space model. The variance of the observation equation for $\ln RV$ is set to RQ and the state equation parameters are driven by the scores as in Creal et al. [14]. To capture long memory type persistence in RV , the state vector has dimension 22. To economize on the number of parameters to estimate the system is sparse, with score dynamics restricted to follow a diagonal random walk. Ding [15] extends the GARCH model to capture dynamics in the variance of variance.

Section 3 formally describes the proposed model specifications and their properties. Unlike the HAR type models, the dependence in RV is indirectly modeled through dependence in RQ . More specifically, RV is modeled as a function of the (unobservable) conditional variance of RV . The proposed model, therefore, extends Hansen et al. [11] to a higher order moment with a RQ -in-mean type model. The GARCH-in-mean models with asset returns on the left-hand side have a persistence mismatch as returns have little persistence while conditional variances are highly persistent [16]. For the higher-order specification in this paper, the left-hand side RV is persistent as with the right-hand side conditional variance of RV .

Section 3 also discusses the somewhat neglected issue of the use of nonlinear transformations. The three most commonly used transformations in the literature are RV (no transformation), \sqrt{RV} [1], and $\ln RV$ [17]. The first two forms are the variable of interest in financial applications. However, in the majority of the models that are estimated by least squares, the parameters are not restricted to ensure $RV \geq 0$ and may result in negative out-of-sample forecasts. The log transformation ensures non-negative forecasts without parameter restrictions. Another important advantage of the log transformation, as documented in the empirical analysis Section 4, is the removal of the excess kurtosis in RV and RQ . The proposed model with the log transformation can, therefore, be estimated by maximizing the Gaussian likelihood.

The empirical performance of the proposed model is considered in Section 4 using data for the 27 stocks analyzed in Bollerslev et al. [12]. This section first documents the claimed two empirical features of RV and RQ for the sample of 27 stocks. The proposed model is then fitted using a rolling estimation window for each stock. The rolling pseudo-out-of-sample predictions from the proposed model are then compared to those from the baseline HARQ model of Bollerslev et al. [12]. The HARQ model is chosen as the baseline for comparison as it also exploits information from RQ to predict RV .

The rest of the paper is organized as follows. Section 2 provides a summary of related literature, Section 3 formally describes the proposed model specification, Section 4 examines the empirical performance of the proposed model specification, and Section 5 provides some concluding comments.

2. Related Literature

An early parametric model with a long memory for the RV process was suggested in Andersen et al. [18]. They specify a fractionally integrated VAR where the fractional difference parameter d is estimated by the log-periodogram regression of Geweke and Porter-Hudak [19]. The long memory in RV is modeled by a restricted AR(22) specification in Corsi [1]. This long AR specification, known as the HAR (Heterogeneous AutoRegressive) model, has become the standard model to capture long memory like persistence in RV since it can be estimated by simple least squares.

Corsi et al. [17] extends the HAR specification by modeling conditional dependence in the second moment of the residuals. As the outcome variable in a HAR model is the second moment RV , the residual second moment is the variance of variance. This model of conditional fourth-moment dynamics is the higher-order extension of the GARCH class of conditional second-moment dynamic models. Like the GARCH model, the con-

ditional fourth moment is a latent variable that is not directly observable in the model of Corsi et al. [17].

The proposal in this paper is to use the observed fourth-moment variable, the realized quarticity RQ . RQ was used in the HAR extension of Bollerslev et al. [12], which motivates its use as an instrument for potential measurement error in the lagged RV on the right-hand side of the regression. Their HARQ model can be estimated by simple least squares. Buccheri and Corsi [13] deal with the RV measurement error in a state-space model where the observed RV is modeled as the latent signal IV plus a measurement error noise. The unobserved latent signal IV is the state variable that needs to be filtered. To capture long memory type persistence, IV has a HAR type long AR specification resulting in a latent state vector of dimension 22. Their HARK model uses the observed RQ process as the time-varying variance of the observation equation for RV . Although the state-space model can be estimated by maximum likelihood with the use of Kalman filter, the likelihood is nonlinear in the parameters and estimation is computationally more expensive than the least squares based HARQ model of Bollerslev et al. [12].

Bollerslev et al. [12] and Buccheri and Corsi [13] both use the observed fourth moment variable RQ in the HAR class model to capture long memory type persistence in RV . Rather than use long lags of RV to capture persistence, the proposal in this paper is to exploit similar long memory type persistence in RQ to model the RV dynamics. To exploit potential dependence in the residual second moments as in Corsi et al. [17], the proposed model is a higher moment extension of the realized GARCH model of Hansen et al. [11] and Hansen and Huang [20]. Compared to their realized GARCH specification, the proposed model is an RQ -in-mean specification as described below. The likelihood function of the proposed model is nonlinear in the parameters and requires numerical optimization for maximum likelihood estimation.

3. Realized Variance Model

Online Appendix A.1 provides a brief summary of asymptotic distributions of realized variances.

3.1. Model Specification

The proposed model extends the realized GARCH model of Hansen et al. [11] in two ways. First, instead of the asset return (first moment), the conditional mean equation is for the realized variance (second moment) of asset returns. Second, as the realized variance is persistent, the conditional mean equation has a conditional variance of variance term analogous to the GARCH-in-mean specification to capture the persistence in realized variance.

$$y_t = c_0 + c_1 \ln \kappa_t + \sqrt{\kappa_t} \epsilon_t, \quad \epsilon_t \sim iid(0, 1) \quad (1a)$$

$$\ln \kappa_t = \omega + \alpha x_{t-1} + \beta \ln \kappa_{t-1} \quad (1b)$$

$$x_t = \zeta + \phi \ln \kappa_t + \tau(\epsilon_t) + \sigma_u u_t, \quad u_t \sim iid(0, 1) \quad (1c)$$

where ϵ_t and u_t are assumed independent.

The observed outcome variable y_t is some function of the realized variance RV_t . The choice of transformation for y_t is discussed below in Section 3.2. κ_t is the (unobserved) conditional variance of y_t . As κ_t is the variance of variance, the parameter c_1 is the variance-in-mean parameter analogous to the GARCH-in-mean specification [21]. This term is included in the conditional mean Equation (1a) to capture the dependence in y_t .

The conditional dynamics for κ_t is specified in log form in (1b) to ensure κ_t is non-negative without restricting the parameters in (1b). This specification is analogous to that of exponential GARCH of Hansen and Huang [20] and Nelson [22]. Corsi et al. [17] considered a GARCH analogue for κ_t where $x_{t-1} = \kappa_{t-1} \epsilon_{t-1}^2$ in Equation (1b) with κ_t on the left-hand side. For the proposed model x_t is an observed variable analogous to Hansen et al. [11] and Hansen and Huang [20]. Hansen et al. [11] ‘augmented’ the GARCH model by using informa-

tion from realized variance by setting $x_t = RV_t$. For this application, κ_t is the conditional fourth moment and x_t is set to some transformation of the realized quarticity RQ_t .

In Equation (1c) for the observed variable x_t , the term $\tau(\epsilon_t)$ captures potential asymmetric response to negative and positive shocks ϵ_t . As in Hansen et al. [11], the quadratic specification $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$ is used in the empirical application. $\tau(\epsilon_t)$ is then a zero mean process where the parameter τ_1 captures the asymmetric response. (1b) can include additional lags of x_t and $\ln \kappa_t$ but we use only one lag for parsimony.

Model (Section 3.1) implies that $\ln \kappa_t$ follows an AR(1) process and x_t an ARMA(1,1) process [11]

$$\begin{aligned}\ln \kappa_t &= \omega + \alpha \zeta + (\beta + \alpha \phi) \ln \kappa_{t-1} + \alpha e_{t-1} \\ x_t &= \omega \phi + \zeta(1 - \beta) + (\beta + \alpha \phi)x_{t-1} + e_t - \beta e_{t-1}\end{aligned}\quad (2)$$

where $e_t \equiv \tau(\epsilon_t) + \sigma_u u_t$ is a zero mean *iid* process. The stationarity condition for $\ln \kappa_t$ and x_t is $\beta + \alpha \phi < 1$ with unconditional means

$$E[\ln \kappa_t] = \frac{\omega + \alpha \zeta}{1 - \beta - \alpha \phi}, \quad E[x_t] = \frac{\omega \phi + \zeta(1 - \beta)}{1 - \beta - \alpha \phi} \quad (3)$$

$\rho \equiv \beta + \alpha \phi$ is a measure of persistence of the model implied series $\ln \kappa_t$ and x_t as their autocorrelations decay with powers of ρ .

3.2. Variable Transformations

The proposed model (Section 3.1) can be considered a class of models depending on the choice of transformed variables y_t and x_t . The literature that models realized variance RV_t have used alternative transformations of RV_t for y_t . The HAR model of Corsi [1] uses $y_t = \sqrt{RV_t}$, Corsi et al. [17] consider $y_t = \sqrt{RV_t}$, $y_t = \ln RV_t$ and Bollerslev et al. [12] consider $y_t = RV_t$. Strictly speaking the non-negativity of $y_t = RV_t$, $y_t = \sqrt{RV_t}$ requires restrictions on the mean Equation (1a) parameters to ensure out of sample forecasts remain non-negative. Such restrictions, however, do not appear to be imposed for least squares based estimators in Corsi et al. [17] and Bollerslev et al. [12]. As a consequence out-of-sample predictions may produce negative values of $y_t = RV_t$ or $y_t = \sqrt{RV_t}$.

For this reason, the empirical application below uses the log transformation $y_t = \ln RV_t$ which ensures non-negative RV_t predictions of RV_t without restrictions on the mean equation parameters (1a). A problem with the use of log transformation is that the variable of interest in financial applications is often the variance itself RV_t or the volatility $\sqrt{RV_t}$, not the log transformation. If we use the log transformation in (Section 3.1), the predicted values need to go through a nonlinear transformation to obtain predictions of the variable of interest. The nonlinear transformation results in approximate inference for the variable of interest even under the rather strong assumption of a correctly specified distribution for the transformed variable y_t .

The variable x_t is an observed measure of the conditional variance of y_t . As y_t is a second moment variable, it is natural to use a realized fourth moment or quarticity RQ_t for x_t . For transformations $y_t = \sqrt{RV_t}$ or $y_t = \ln RV_t$, the asymptotic variance expression in online Appendix A.1 have IV_t in the denominator. This may cause problems for the values of RV_t close to zero, the so-called inlier problem. As the model (Section 3.1) restricts x_t to follow an ARMA(1,1) process, an alternative approach is to choose a transformation of RQ_t in the sample that has similar dependence as an ARMA(1,1) process. A further consideration is that the model implied process for x_t does not ensure positive x_t without restrictions on the parameters. The log transformation $x_t = \ln RQ_t$ ensures positive RQ_t without imposing additional restrictions on the parameters.

For these reasons, the empirical application below uses the log transformation $x_t = \ln RQ_t$. Alternative transformations of the conditional variance have been considered for the (G)ARCH-in-mean term. The RQ -in-mean term with coefficient c_1 in (1a) can also use κ_t itself or transformations such as $\sqrt{\kappa_t}$. However, (1a) uses the log transformation as the

dynamics of κ_t in logs in (1b) ensures non-negative κ_t without restricting the parameters. Furthermore, the log transformation of κ_t matches the log transformation of RV_t for y_t and of RQ_t for x_t .

3.3. Maximum Likelihood Estimation

The parameters of the model can be estimated by maximum likelihood assuming the error terms have a Gaussian distribution. While one can interpret the Gaussian assumption as a quasi-likelihood, for finite sample performance, it is desirable that the transformations discussed in the previous section are chosen so that the distribution of ϵ_t and u_t are not ‘too’ different from the Gaussian. The levels of RV_t and RQ_t are well known to have very high kurtosis and an important reason to prefer the log transformation for y_t and x_t is to remove the excess kurtosis. Corsi et al. [17] and Barndorff-Nielsen and Shephard [23] find that the finite sample distribution of $\ln RV_t$ is closer to the Gaussian than RV_t .

For model (Section 3.1), the contribution to the Gaussian log-likelihood from the t -th observation is

$$\begin{aligned}\ell_t(y_t, x_t) &= \ell_t(\epsilon_t) - \frac{1}{2} \ln \kappa_t + \ell_t(u_t) - \frac{1}{2} \ln \sigma_u^2 \\ &= -\ln(2\pi) - \frac{1}{2}(\epsilon_t^2 + \ln \kappa_t + u_t^2 + \ln \sigma_u^2)\end{aligned}$$

For the quadratic leverage function $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$ used in Hansen et al. [11], the parameter vector is $\theta = (\theta_1, \theta_2, \theta_3)$, $\theta_1 = (c_0, c_1)$, $\theta_2 = (\omega, \alpha, \beta)$, $\theta_3 = (\xi, \phi, \tau_1, \tau_2, \sigma_u)$. The parameter restrictions are $\sigma_u > 0$ and for stationarity of κ_t and x_t , $\beta + \alpha\phi < 1$.

To start the recursion in (1b), we need to specify presample values of x_t, κ_t . If stationarity of x_t, κ_t are imposed, we can set these to the unconditional means $x_0 = E[x_t]$, $\kappa_0 = E[\kappa_t]$ given in (3). A simple alternative used in the empirical application below is to use the sample variance of y_t .

For statistical inference, the ‘sandwich’ QML (quasi-maximum likelihood) parameter covariance matrix can be used by evaluating the first and second derivatives of the contributions to the Gaussian log-likelihood. Analytical expressions to recursively evaluate the first derivatives are given in the online Appendix A.2. The second derivatives can be evaluated by numerically differentiating the analytical first derivatives.

3.4. Model Evaluation by Pseudo Out-of-Sample Forecasting

A standard approach to evaluating model performance is to compare the accuracy of pseudo-out-of-sample forecasts. For alternative models with the same outcome variable y_t and provided the actual outcome y_t is observable, the comparison can be made by specifying a loss, or scoring, function.

There are two additional issues to consider in this application where the outcome of interest is (some function of) the daily realized variance RV_t . First, if the statistic of interest is the integrated variance IV_t , the outcome y_t based on RV_t is likely measured with error. Patton [24] suggested scoring functions robust to additive error in the outcome variable.

The second issue is the comparison of models where the outcome variable of interest is some transformation of the modeled variable y_t . For example, in finance applications rather than $y_t = \ln RV_t$ we are often more interested in $\sqrt{RV_t} = \exp(y_t/2)$. To compare models with outcomes $y_{1t} = RV_t$, $y_{2t} = \sqrt{RV_t}$, $y_{3t} = \ln RV_t$, we need to specify a common outcome variable of interest. In this case setting RV_t as the variable of interest is the least problematic as obtaining forecasts of $\sqrt{RV_t}$ from RV_t under additive error is mathematically intractable. Furthermore RV_t is the observed data variable constructed from intraday data.

A commonly used approach to obtain forecasts for RV_t is to assume Gaussianity. For $y_t = \ln RV_t$ and $\epsilon_t \sim N(0, 1)$, the h -step forecast of y is Gaussian with

$$\begin{aligned}E_t[y_{t+h}] &\sim N(\hat{y}_{t+h|t}, \kappa_t) \\ \hat{y}_{t+h|t} &= c_0 + c_1 \ln \kappa_t\end{aligned}$$

and the forecasts for the transformed variables are

$$E_t[RV_{t+h}] = E_t[\exp(y_{t+h})] = \exp(\hat{y}_{t+h|t} + \frac{1}{2}\kappa_t)$$

$$E_t[\sqrt{RV}_{t+h}] = E_t[\exp(y_{t+h}/2)] = \exp(\frac{1}{2}\hat{y}_{t+h|t} + \frac{1}{8}\kappa_t)$$

An alternative approach that does not assume Gaussianity is the ‘smearing’ estimate of Duan [25] and Wooldridge ([26], 6-4). Let $\hat{e}_t = y_t - \hat{y}_t$, $t = 1, \dots, T$ denote the in-sample residuals. The smearing forecasts are

$$\widehat{RV}_{T+h} = \exp(\hat{y}_{T+h}) \left(\frac{1}{T} \sum_{t=1}^T \exp(\hat{e}_t) \right)$$

$$\widehat{\sqrt{RV}}_{T+h} = \exp(\frac{1}{2}\hat{y}_{T+h}) \left(\frac{1}{T} \sum_{t=1}^T \exp(\frac{1}{2}\hat{e}_t) \right)$$

4. Empirical Application

This section uses the 5-minute return data from Bollerslev et al. [12] to evaluate the performance of the proposed specification (Section 3.1) for the realized variance. The data, made publicly available by the authors, is a daily sample of 27 Dow Jones constituent stocks from 22 April 1997 to 31 December 2013 (4200 trading days). A list of the 27 stocks and their ticker symbols are provided in online Appendix A.3.

4.1. Preliminary Analysis

Before we fit the model to the data, we examine the log transformations chosen for the observed variables $y_t = \ln RV_t$ and $x_t = \ln RQ_t$. As explained above, the log transformations ensure non-negative RV_t and RQ_t without restricting the model parameters. Another important reason for using the log transformation for y_t is to approximate the Gaussian distribution assumed for the likelihood function closely. Table A2 in the online Appendix A.3 shows the sample third (skewness) and fourth (kurtosis) moments of RV_t , \sqrt{RV}_t , and $\ln RV_t$ for the 27 stocks in the sample.

Table A2 shows that all three transformations are positively skewed. The untransformed RV has the largest positive skewness followed by \sqrt{RV} , which have skewness all above one. The log transformation $\ln RV$ has the smallest skewness all below one, but they are all significantly different (at size 0.05) from the Gaussian value of zero. The untransformed RV also has the largest kurtosis, all well above the Gaussian value of three. \sqrt{RV} has somewhat smaller kurtosis, but the smallest value, 7.42 (for INTC), is still well above three. The log transformation $\ln RV$ has kurtosis much closer to Gaussian with the largest value at 4.15 (for CVX).

As mentioned in the introduction, the proposed specification (Section 3.1) is based on two empirical features of RV and RQ : their common persistence and their correlation. Figure A1 in the online Appendix A.3 shows the sample autocorrelations of $\ln RV$ and $\ln RQ$ for the 27 stocks in the sample. For all stocks, $\ln RV$ is somewhat more persistent than $\ln RQ$. The autocorrelations for $\ln RV$ slowly decay from about 0.8 while those for $\ln RQ$ slowly decay from about 0.6. Both autocorrelations die out slowly with the lag, a feature of long memory series. The two series are highly correlated with each other with all pairwise sample correlations above 0.95.

The model (Section 3.1) implies that the observed realized process x_t should follow the restricted ARMA(1,1) process (2). The slowly decaying autocorrelations for $x_t = \ln RQ_t$ in Figure A1 is not inconsistent with an ARMA(1,1) process with a large AR(1) coefficient. As a further check, Table A5 in the online Appendix A.3 reports estimates of an unrestricted ARMA(1,1) model fitted to x_t for the 27 stocks in the sample. Both the AR and MA coefficients are statistically significant, with a large positive AR(1) coefficient and a large negative MA(1) coefficient. The estimated AR(1) coefficient ranges from 0.976 (XOM) to 0.992 (MCD) and the MA(1) coefficient from -0.709 (XOM) to -0.844 (NKE). Although the

portmanteau test for residual correlation up to lag 5 rejects the white noise residual null (at size 0.05), the residual serial correlations are small in size. The first order residual serial correlation ranges from 0.103 (NKE) to 0.036 (CVX).

4.2. In-Sample Estimates

To evaluate the performance of the model out of the sample, pseudo-out-of-sample rolling forecasts were obtained. As parameter estimation by maximum likelihood is computationally expensive compared to least squares, a rolling forecast window of one (calendar) month was moved forward each month starting from January 2006. The first set of parameter estimates were obtained for the estimation sample from the beginning of the data sample April 1997 to December 2005 (104 months). Using these parameter estimates, h -step forecasts for trading days in the month of January 2006 are obtained. These forecasts use the same parameter estimates but the conditioning information set, i.e., the lagged variables on the right-hand side are updated as we move forward within the forecast window.

The next set of estimates were obtained for the sample from May 1997 to January 2006 and forecasts for trading days in the month of February 2006 were obtained. This resulted in 96 sets of parameter estimates with forecasts from the beginning of January 2006 to the end of December 2013. This forecast sample included the financial crisis period 2008–2009 when there was a spike in RV .

Numerical maximization of the Gaussian likelihood can be sensitive to the choice of starting values. To guard against getting stuck in local maxima, a few alternative random starting values are tried for each estimation window. To start the recursion, the presample values for x_t and $\ln \kappa_t$ were set to the estimation sample variance of y_t . (The alternative of setting these presample values to the model unconditional means (3) often resulted in the nonconvergence of the numerical optimizer and was sensitive to the choice of starting parameter values).

There are a large number of estimated parameters (10 parameters for each of the 96 estimation windows for each of the 27 stocks). Figure 1 shows a summary of the estimated parameters for stepsizes $h = 1, 5, 22$ days. Following Bollerslev et al. [12], for the h -step forecast the outcome variable on the left-hand side of (1a) is

$$y_t^{(h)} \equiv \begin{cases} \frac{1}{h} \sum_{j=1}^h y_{t+j-1}, & y_t = RV_t, \sqrt{RV_t} \\ \ln\left(\frac{1}{h} \sum_{j=1}^h Y_{t+j-1}\right), & y_t = \ln(RV_t), Y_t = RV_t, \sqrt{RV_t} \end{cases} \quad (4)$$

The h -step ahead variable for the log transformation is defined so that the forecasts from $y_t^{(h)}$ can be compared to those from the other transformations. For $h = 1$, taking the log of RV or \sqrt{RV} just results in a different scaling. For $h > 1$, the log of RV averages and the log of \sqrt{RV} averages are considered since one cannot be recovered from the other just by rescaling. Strictly speaking, the model parameters change with the forecast stepsize h and should be written as a function of h .

Each panel in Figure 1 corresponds to a parameter and the shaded area is the interquartile range (from the 0.25 to 0.75 quantile) across the 27 stocks. The thick solid line is the median estimate across the 27 stocks. The intercept c_0 and the RQ -in-mean parameter c_1 of Equation (1a) are both positive. A positive RQ -in-mean parameter c_1 is to be expected given the similarity of the $\ln RV$ and $\ln RQ$ dynamics documented above. c_1 declines with stepsize h and for $h > 1$, c_1 is smaller for the log average of \sqrt{RV} than for the log average of RV .

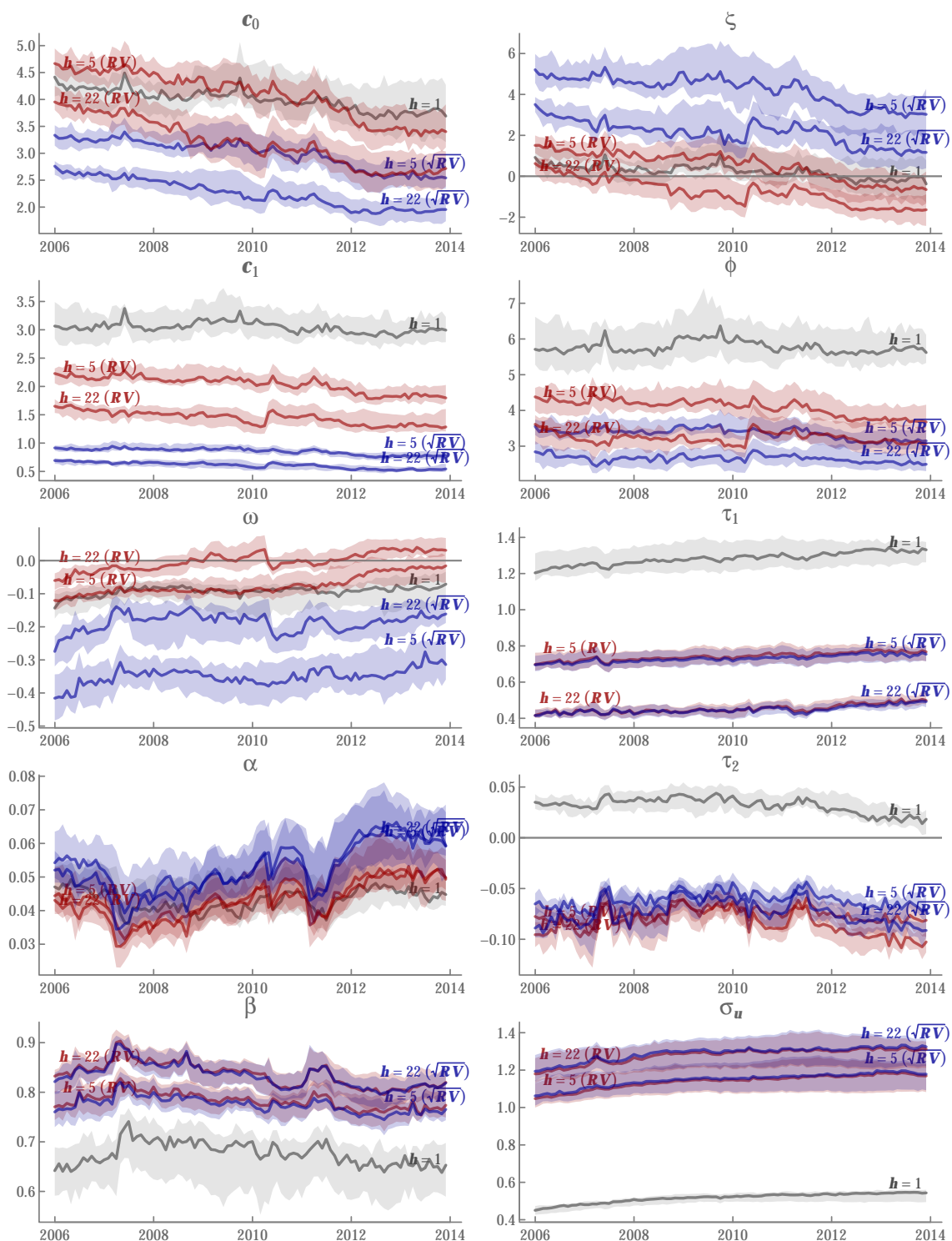


Figure 1. Rolling parameter estimates for log-in-mean specification (Section 3.1) for stepsizes $h = 1, 5, 22$. For $h > 1$, two models are estimated, one using the log of the average of RV (red) and one with the log of the average of \sqrt{RV} (blue). The shaded area is the interquartile range across the 27 stocks and the solid line is the median. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days). The rolling estimation window is 8.67 years (104 months) followed by a one-month prediction window shifted each month over the sample for 96 estimation windows.

The parameters α, β, ϕ that determine the persistence $\rho \equiv \beta + \alpha\phi$ of the $\ln \kappa_t$ and x_t processes are all positive and imply ρ close to but below the stationarity boundary of one (Figure 2). As a function of stepsize h , α does not change much for the log average of RV

but somewhat increases for the log average of \sqrt{RV} . β increases with h from about 0.7 for $h = 1$ to above 0.8 for $h = 22$ and ϕ decreases with h from about 6 for $h = 1$ to less than 4 for $h = 22$. Figure 2 shows the rolling estimates of the persistence parameter ρ increase with the stepsize h . This is to be expected as the h -step outcome variable $y_t^{(h)}$ gets smoother and hence more persistent with h .

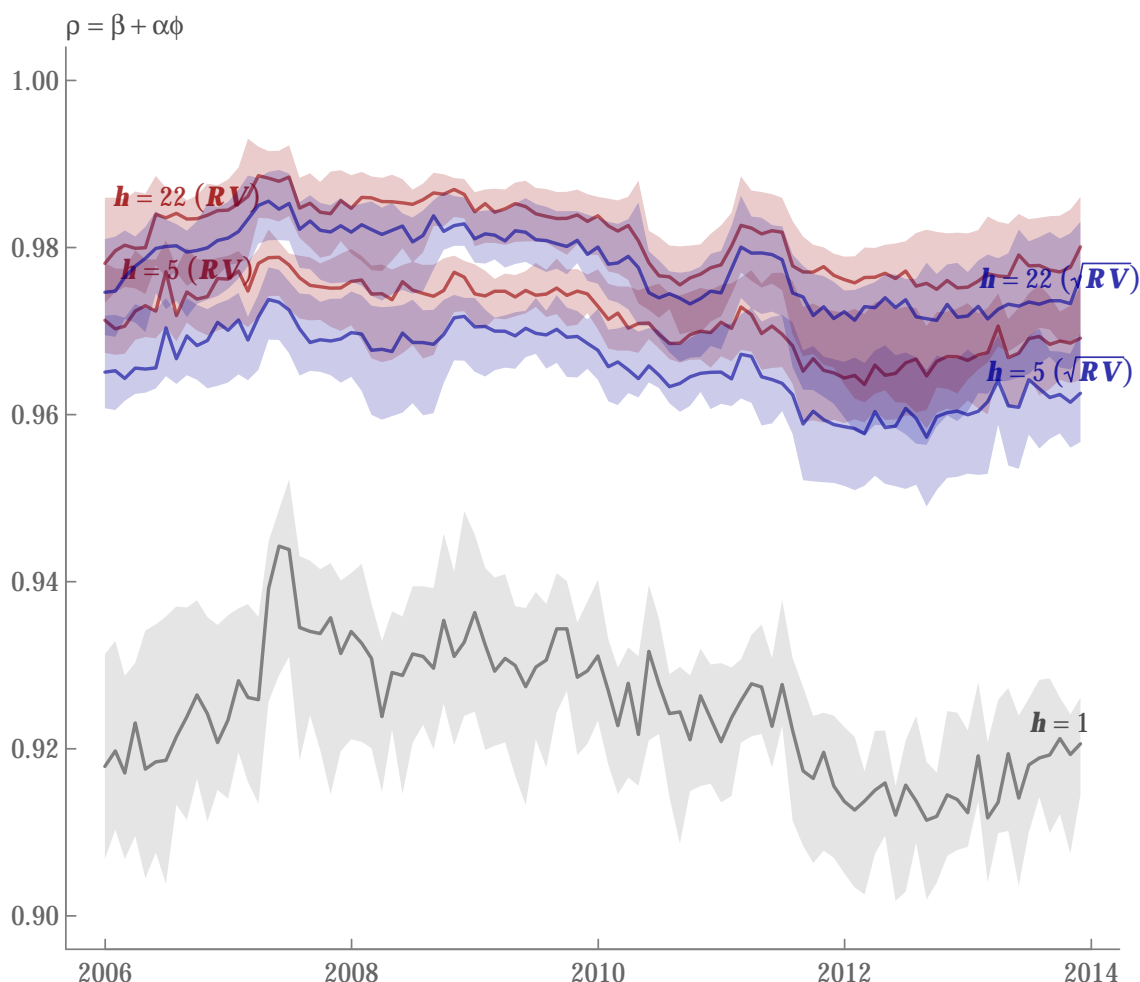


Figure 2. Rolling estimates of the implied persistence parameter ρ for different stepsize h . For $h > 1$, two models are estimated: one using the log of the average of RV (red) and one with the log of the average of \sqrt{RV} (blue). The shaded area is the interquartile range across the 27 stocks and the solid line is the median. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days). The rolling estimation window is 8.67 years (104 months) followed by a one-month prediction window shifted each month over the sample for 96 windows.

The asymmetric response parameter τ_1 is positive and decreases with stepsize h from about 1.3 for $h = 1$ to below 0.5 for $h = 22$. The coefficient on the quadratic term τ_2 switches sign from positive for $h = 1$ to negative for $h = 5, 22$. The volatility parameter σ_u for the $x_t = \ln RQ_t$ Equation (1c) increases with stepsize h .

4.3. Pseudo Out-of-Sample Forecasts

This section evaluates the performance of the proposed model in terms of (pseudo) out-of-sample rolling forecasts described above in Section 4.2. If the model captures the time series dependence of the outcome variable, the forecast errors should not be serially correlated. A formal such test needs to account for sampling error in generating the forecasts based on the estimated parameters. Figures A2–A4 in the online Appendix A.4

provide an informal check by plotting the autocorrelation functions of the rolling forecast errors. In addition to forecasts error from the proposed model (Section 3.1), these figures also compare the forecast error correlations from the HAR model of Corsi [1] with outcome $y = \sqrt{RV}$ and the HARQ model of Bollerslev et al. [12] with outcome $y = RV$.

For stepsize $h = 1$, the forecast error serial correlation is small in magnitude, but most of them fall outside the asymptotic interval for a white noise series. (A portmanteau test for serial correlation up to lag 5 all reject the null hypothesis of a white noise at the conventional size of 0.05. However, these are asymptotic p -values that ignore parameter estimation sampling error.) For stepsizes $h > 1$, the forecast errors show a large positive correlation up to lag h for all models. This dependence is due to the overlapping sample used in generating the multistep outcome variable (4).

To evaluate the relative forecast performance of the proposed model, its forecast accuracy is compared against a baseline model. The baseline model is the HARQ model of Bollerslev et al. [12], which extends the HAR model of Corsi et al. [17] with an additional interaction term involving \sqrt{RQ} . The model proposed in this paper also uses the realized quarticity RQ but without the lagged RV terms of HAR(Q). Following Bollerslev et al. [12], the outcome variable of the baseline HARQ model is the untransformed $y = RV$. The model parameters are estimated by least squares without restricting them to ensure the predicted values are non-negative. Bollerslev et al. ([12], footnote 17) apply the ‘insanity’ filter and replace predicted values that are outside the in-sample range with the in-sample mean value. The baseline predictions in this paper do not apply this somewhat ad hoc insanity filter.

An alternative natural baseline that ensures non-negative predicted RV values is the HARQ model with $y = \ln(RV)$ and an interaction term with $\ln(RQ)$. This log version of the HARQ specification, however, does not perform as well as the untransformed specification of Bollerslev et al. [12]. Table A3 in the online Appendix A.3 reports estimated coefficients on the interaction term for the full sample for stepsizes $h = 1, 5, 22$. For the level (untransformed) specification, these coefficients are all negative with t -ratios above two in absolute value (with two exceptions for $h > 1$) as reported in Bollerslev et al. [12]. For the log specification, many coefficients are positive and insignificant. For $h > 1$ all t -ratios (except one) are below two in absolute value. Bollerslev et al. [12] motivate the use of \sqrt{RQ} to correct for possible measurement error in RV . The negative coefficient of the interaction term may be correcting for this error mainly in the tails of the high excess kurtosis of RV . This may explain the insignificant coefficient in the interaction for the log specification as the excess kurtosis largely disappears for $\ln(RV)$. The issue of appropriate HARQ specification is not the focus of this study and is left for further research.

To evaluate the relative accuracy of predictions from alternative models, we need to specify a loss or scoring function. For this study, two members from the Bregman family of consistent scoring functions that are robust to additive noise in RV [24] are used.

$$S_{ms}(RV, F) = (RV - F)^2$$

$$S_{ql}(RV, F) = \frac{RV}{F} - \ln\left(\frac{RV}{F}\right) - 1$$

where RV is the actual realized value and F its forecast. S_{ms} is the mean squared and S_{ql} is the QLIKE scoring function. These two scoring functions were also used in Bollerslev et al. [12]. S_{ms} is defined for all values of F while S_{ql} is undefined for $F < 0$. As the baseline HARQ predictions are not filtered, the small cases of non-positive predictions are removed when evaluating S_{ql} .

The difference in scores between the baseline and comparison model is tested with the t -ratio of equal forecast accuracy of Diebold and Mariano [27]. As mentioned above, the forecast errors are correlated for $h > 1$. To account for this correlation, the denominator of the t -ratio, the standard error of the difference in average scores, is computed using the Bartlett kernel with bandwidth set to the forecast stepsize h . Table 1 reports these t -ratios

for the full forecast sample. A positive τ value indicates a better forecast (smaller score) from the comparison model than the baseline HARQ.

Table 1. t -ratios of Diebold and Mariano [27] test of equal predictive accuracy of h -step RV rolling forecasts. The baseline forecast from the HARQ model (no transformation) is compared against the forecast from the exponential transform of the log forecast from model (Section 3.1) assuming log-normality. A positive τ value indicates a better forecast from the comparison model against the baseline HARQ forecasts. τ_{MS} for equality of mean squared error loss and τ_{QL} for QLIKE loss of Patton [24]. $\widehat{RV} < 0$ is the number of negative forecasts (in the forecast sample) produced by the HARQ model. The bottom four rows are the fraction out of 27 stocks that satisfy the inequalities; for $\widehat{RV} < 0$, it is the fraction of 27 stocks that produced at least one negative forecast. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one-month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

	$h = 1$			$h = 5$			$h = 22$		
	τ_{MS}	τ_{QL}	$\widehat{RV} < 0$	τ_{MS}	τ_{QL}	$\widehat{RV} < 0$	τ_{MS}	τ_{QL}	$\widehat{RV} < 0$
AXP	0.94	−2.65	1	0.92	0.96	1	−0.14	1.75	1
BA	−2.12	0.55	0	−1.44	0.44	0	−0.31	0.43	0
CAT	1.10	−3.33	2	1.04	0.58	1	1.05	1.25	0
CSCO	−2.48	−6.23	0	−1.69	−2.46	0	−1.80	−0.92	0
CVX	0.89	−1.40	1	0.96	0.56	1	1.16	1.28	1
DD	−2.49	−0.71	0	0.04	0.32	0	−0.99	2.16	1
DIS	−1.47	−2.47	0	−0.79	−2.09	0	−1.47	−0.73	0
GE	1.05	0.86	2	0.66	1.28	2	−0.61	2.57	3
HD	−1.96	−2.62	0	−1.24	−1.96	0	−0.01	0.89	0
IBM	−0.28	−2.17	0	−0.02	−2.43	0	−0.95	0.21	0
INTC	−1.60	−6.43	1	0.49	−5.41	1	−0.20	−1.08	1
JNJ	−1.64	−3.24	1	−1.53	−1.70	1	0.89	0.13	2
JPM	−1.65	4.18	1	−1.64	0.12	1	−1.65	2.42	1
KO	−1.49	−4.84	1	−0.10	−0.53	1	0.39	2.39	1
MCD	0.01	1.65	0	−0.49	6.11	0	0.41	5.44	0
MMM	1.03	−5.68	2	1.20	−0.91	2	1.11	0.59	1
MRK	0.98	−1.35	1	1.01	1.57	1	1.03	4.62	1
MSFT	−1.96	−2.62	0	−1.44	−0.07	0	−1.37	0.34	0
NKE	−2.53	−5.08	1	−1.38	−2.04	1	−1.24	0.85	1
PFE	−2.14	−3.35	0	−1.68	−0.82	0	−1.14	0.55	0
PG	−0.51	−0.41	0	−0.87	−1.27	0	1.71	2.22	0
TRV	1.15	−4.57	2	1.01	−2.84	4	1.14	1.90	4
UNH	−0.94	−4.75	1	0.32	−2.72	1	0.85	−0.43	0
UTX	−0.57	−0.38	0	0.95	−3.51	0	−0.21	−0.53	0
VZ	−1.31	−3.06	0	−0.73	−3.78	0	−0.12	0.26	0
WMT	0.54	−1.16	2	0.81	−1.52	2	0.67	1.45	2
XOM	0.59	−0.94	1	1.01	−2.12	1	1.11	−0.20	0
$\tau < 0$	[0.63]	[0.85]	[0.56]	[0.52]	[0.67]	[0.56]	[0.56]	[0.22]	[0.48]
$\tau < -2$	[0.19]	[0.59]		[0.00]	[0.37]		[0.00]	[0.00]	
$\tau > 0$	[0.37]	[0.15]		[0.48]	[0.33]		[0.44]	[0.78]	
$\tau > 2$	[0.00]	[0.04]		[0.00]	[0.04]		[0.00]	[0.26]	

Table 1 shows that in about half of the 27 stocks, a negative RV forecast is produced for the baseline HARQ model. As mentioned above, τ_{QL} is computed for the forecast sample, excluding these negative predictions. The sign switch between τ_{MS} and τ_{QL} may be due to the difference in the forecast evaluation sample for the two score functions. With this caveat, the test result is somewhat sensitive to the choice of scoring function and forecast stepsize h . For $h = 1$, both τ_{MS} and τ_{QL} indicate lower average score, i.e., better forecast accuracy, for the baseline HARQ model than the proposed model (Section 3.1). The forecast accuracy of the proposed model relative to the baseline generally improves with the forecast stepsize

h. For $h = 22$, τ_{MS} is inconclusive in the sense that all their values are less than two in absolute value. None of the τ_{QL} values are less than -2 , but about a quarter of the stocks have values above $+2$ indicating more accurate forecasts from model (Section 3.1) than from the baseline.

The tests in Table 1 are based on the full forecast sample. The test results could be specific to the choice of the forecast sample and a particular sample could be cherry-picked to obtain certain results. As a guard against such potential cherry picking, Figures A5–A7 in the online Appendix A.5 show running *t*-ratios of the Diebold and Mariano [27] equal forecast accuracy tests for all possible end-of-forecast samples. These are the values of *t*-ratios when the test is applied to the forecast sample from the beginning of the forecast sample (2007-01-03) up to each date of an expanding evaluation sample. (The three figures use alternative (un)transformations to obtain forecasts for *RV* from $\ln RV$.)

One common feature of the running *t*-ratio results is the change in performance shortly after the financial crisis period 2008–2009. As expected the test performance up to the financial crisis period is quite noisy, but the majority of the *t*-ratios for both *MS* and *QL* are positive for $h > 1$, indicating better forecast accuracy from model (Section 3.1) compared to the baseline. The problem with the HARQ model in levels producing negative forecasts mostly occurs during the financial crisis period. Table A4 in the online Appendix A.3 lists all dates with negative predictions from the HARQ model, the majority of which occur during 2008. The *t*-ratios remain quite stable after the financial crisis period indicating robustness to the choice of forecast sample post-financial crisis. When judging the significance of these running *t*-ratios, one should be aware of the multiple comparisons problem and that the usual critical values are likely to be too small.

5. Concluding Remarks

This paper proposed a model for *RV* dynamics that exploits the information in the higher order moment of *RQ*. In contrast to the HAR(Q) models that exploit dependence in the *RV* variable itself, the idea is to exploit dependence in the *RQ* series. The empirical analysis using 27 stocks suggests that the proposed model may perform better than the HAR(Q) type specifications for multistep predictions and during periods of market turmoil such as the financial crisis of 2008.

As the empirical analysis was based on large cap (blue chip) US stocks, it remains to be seen how the proposed model performs for other markets or asset classes. These include small-cap stocks, which are known to be volatile and fat-tailed, or cryptocurrencies. Performance during other market turmoil periods, such as the recent pandemic, or for commodities during the recent conflict in Ukraine could also be analyzed.

For policymakers, the performance of the proposed model for risk management purposes could be of interest. The fundamental review of the trading book for the latest Basel regulation has introduced the use of expected shortfall for market risk capital requirements. The forecast performance analysis in this paper could be extended to evaluation of the accuracy of risk measures such as expected shortfalls. Additional tools or models for managing tail risks for regulatory purposes could prove beneficial for large financial institutions.

A somewhat puzzling result is the better accuracy for multistep forecasts from the proposed model compared to the baseline HARQ model. The HARQ specification uses RV_{t-22} while the proposed specification only uses one lag (though, in principle, additional lags could be considered). Whether the implied ARMA(1,1) process for $x_t = \ln RQ$ can better capture long-term dependence than the AR(22) term could be further investigated.

Several extensions of the proposed model could be considered. Bollerslev et al. [12] consider the use of alternative estimators of *RV* such as realized kernel of Barndorff-Nielsen et al. [28] and of *RQ* such as *QP* and *TP* mentioned in Appendix A.1. Alternatively, option-based fourth-moment measures from VVIX as used in Huang et al. [9] could also be considered. Rather than select one of these alternative measures, one can try to incorporate information from all of these alternative measures as in Hansen and Huang [20].

The challenge in further generalizing the model is the increase in the number of parameters to estimate. With ten parameters to estimate, the model is already somewhat over-parameterized as a model of a single outcome variable. To reduce the number of parameters to estimate, one can either impose a priori ‘reasonable’ restrictions or sparsity consistent with the data. Alternatively, a penalty term can be added to the likelihood function for regularization [29,30].

Funding: This research received no external funding

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Appendix A.1. Summary of Asymptotic Distributions of Realized Variances

Assume the log asset price p_t follows a semimartingale process

$$dp_t = \mu(t)dt + \sigma(t)dW(t)$$

with drift $\mu(t)$ and volatility $\sigma(t)$. $W(t)$ is the standard Wiener process. The realized variance over day t is $RV_t = \sum_{j=1}^m r_{t,j}^2$ where $r_{t,j}$ is the log return over the j -th subinterval in day t and m is the number of subintervals over one trading day. In the diffusion limit $m \rightarrow \infty$ [31,32],

$$RV_t \rightarrow MN(IV_t, \frac{2}{m} IQ_t)$$

where $IV_t = \int_{t-1}^t \sigma^2(s)ds$ is the integrated variance and $IQ_t = \int_{t-1}^t \sigma^4(s)ds$ is the integrated quarticity. The (approximate) asymptotic distribution of nonlinear transformations of RV_t can be obtained by the delta method [17,23]

$$\begin{aligned}\sqrt{RV_t} &\rightarrow MN(\sqrt{IV_t}, \frac{1}{2m} \frac{IQ_t}{IV_t}) \\ \ln RV_t &\rightarrow MN(\ln IV_t, \frac{2}{m} \frac{IQ_t}{IV_t^2})\end{aligned}$$

To estimate the (asymptotic) variance of RV_t , we need an estimate of integrated quarticity IQ_t . Commonly used consistent estimators (under no microstructure noise) are

$$\begin{aligned}RQ_t &= \frac{m}{3} \sum_{j=1}^m r_{t,j}^4 \\ QP_t &= \frac{m\pi^2}{4} \sum_{j=1}^m |r_{t,j}| |r_{t,j-1}| |r_{t,j-2}| |r_{t,j-3}| \\ TP_t &= \frac{m\Gamma(1/2)^3}{4\Gamma(7/6)^3} \sum_{j=1}^m |r_{t,j}|^{4/3} |r_{t,j-1}|^{4/3} |r_{t,j-2}|^{4/3}\end{aligned}$$

RQ_t is the realized quarticity, QP_t is the realized quad-power quarticity [33,34], and TP_t is the realized tri-power quarticity [35].

Appendix A.2. Derivatives of Log-Likelihood

For (Section 3.1) the contribution to the Gaussian log-likelihood from the t -th observation is

$$\begin{aligned}\ell_t(y_t, x_t) &= \ell_t(\epsilon_t) - \frac{1}{2} \ln \kappa_t + \ell_t(u_t) - \frac{1}{2} \ln \sigma_u^2 \\ &= -\ln(2\pi) - \frac{1}{2}(\epsilon_t^2 + \ln \kappa_t + u_t^2 + \ln \sigma_u^2)\end{aligned}$$

Contribution to score is

$$\frac{\partial \ell_t}{\partial \theta} = -\epsilon_t \frac{\partial \epsilon_t}{\partial \theta} - u_t \frac{\partial u_t}{\partial \theta} - \frac{1}{\sigma_u} \frac{\partial \sigma_u}{\partial \theta} - \frac{1}{2} \frac{\partial \ln \kappa_t}{\partial \theta}$$

where

$$\begin{aligned} \frac{\partial \ln \kappa_t}{\partial \theta_1} &= 0, \quad \frac{\partial \ln \kappa_t}{\partial \theta_2} = \begin{pmatrix} 1 \\ x_{t-1} \\ \ln \kappa_{t-1} \end{pmatrix} + \beta \frac{\partial \ln \kappa_{t-1}}{\partial \theta_2}, \quad \frac{\partial \ln \kappa_t}{\partial \theta_3} = \beta \frac{\partial \ln \kappa_{t-1}}{\partial \theta_3} \\ \frac{\partial \epsilon_t}{\partial \theta_1} &= -\frac{1}{\sqrt{\kappa_t}} \begin{pmatrix} 1 \\ \ln \kappa_t \end{pmatrix} - \left(\frac{\epsilon_t}{2} + \frac{c_1}{\sqrt{\kappa_t}} \right) \frac{\partial \ln \kappa_t}{\partial \theta_1}, \quad \frac{\partial \epsilon_t}{\partial \theta_j} = -\left(\frac{\epsilon_t}{2} + \frac{c_1}{\sqrt{\kappa_t}} \right) \frac{\partial \ln \kappa_t}{\partial \theta_j}, \quad j = 2, 3 \\ \frac{\partial u_t}{\partial \theta_j} &= -\frac{\phi}{\sigma_u} \frac{\partial \ln \kappa_t}{\partial \theta_j} - \frac{1}{\sigma_u} (\tau_1 + 2\tau_2 \epsilon_t) \frac{\partial \epsilon_t}{\partial \theta_j}, \quad j = 1, 2, \\ \frac{\partial u_t}{\partial \theta_3} &= -\frac{1}{\sigma_u} \begin{pmatrix} 1 \\ \ln \kappa_t \\ \epsilon_t \\ \epsilon_t^2 - 1 \\ u_t \end{pmatrix} - \frac{\phi}{\sigma_u} \frac{\partial \ln \kappa_t}{\partial \theta_3} - \frac{1}{\sigma_u} (\tau_1 + 2\tau_2 \epsilon_t) \frac{\partial \epsilon_t}{\partial \theta_3} \end{aligned}$$

For $t = 1$, if $x_0, \ln \kappa_0$ are fixed, e.g., set to estimation sample variance of y_t ,

$$\frac{\partial \ln \kappa_1}{\partial \theta_2} = \begin{pmatrix} 1 \\ x_0 \\ \ln \kappa_0 \end{pmatrix}, \quad \frac{\partial \ln \kappa_1}{\partial \theta_3} = 0$$

and if $x_0, \ln \kappa_0$ are set to their unconditional means

$$\frac{\partial \ln \kappa_1}{\partial \theta_2} = \frac{1}{1 - \beta - \alpha \phi} \begin{pmatrix} 1 \\ x_0 \\ \ln \kappa_0 \end{pmatrix}, \quad \frac{\partial \ln \kappa_1}{\partial \theta_3} = \frac{\alpha}{1 - \beta - \alpha \phi} \begin{pmatrix} 1 \\ \ln \kappa_0 \\ 0_3 \end{pmatrix}$$

Appendix A.3. Ticker Symbols

Twenty-seven constituents from the Dow Jones (Table 2, [12]).

Table A1. Ticker symbols from Dow Jones constituents. Reproduced from (Table 2, [12]).

Symbol	Exchange	Company
AXP	NYSE	American Express Company
BA	NYSE	Boeing Company
CAT	NYSE	Caterpillar Inc.
CSCO	NASDAQ	Cisco Systems, Inc.
CVX	NYSE	Chevron Corporation
DD	NYSE	DuPont de Nemours, Inc.
DIS	NYSE	Walt Disney Company
GE	NYSE	General Electric Company
HD	NYSE	Home Depot
IBM	NYSE	International Business Machine Corporation
INTC	NASDAQ	Intel Corporation
JNJ	NYSE	Johnson & Johnson
JPM	NYSE	JPMorgan Chase & Co.
KO	NYSE	Coca-Cola Company
MCD	NYSE	McDonald's Corporation
MMM	NYSE	3M Company
MRK	NYSE	Merck & Co., Inc.
MSFT	NASDAQ	Microsoft Corporation

Table A1. *Cont.*

Symbol	Exchange	Company
NKE	NYSE	Nike, Inc.
PFE	NYSE	Pfizer Inc.
PG	NYSE	Procter & Gamble Company
TRV	NYSE	Travelers Companies, Inc.
UNH	NYSE	UnitedHealth Group Incorporated
UTX	NYSE	United Technologies Corporation
VZ	NYSE	Verizon Communications Inc.
WMT	NYSE	Walmart Inc.
XOM	NYSE	ExxonMobil Corporation

Table A2. Sample third (*Skew*) and fourth (*Kurt*) moments of transformations of realized variance *RV*. *p*-values in square brackets are for *Kurt* = 3 for the log transformation. All other *p*-values for *Skew* = 0 and *Kurt* = 3 are less than 0.05 and are not reported to avoid cluttering the table. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days).

	<i>RV</i>		\sqrt{RV}		<i>ln RV</i>		
	<i>Skew</i>	<i>Kurt</i>	<i>Skew</i>	<i>Kurt</i>	<i>Skew</i>	<i>Kurt</i>	
AXP	11.59	262.02	2.74	17.68	0.34	2.95	[0.53]
BA	6.57	76.69	2.21	12.35	0.34	3.19	[0.01]
CAT	7.42	110.65	2.33	13.47	0.39	3.32	[0.00]
CSCO	4.63	36.34	1.95	8.61	0.43	2.90	[0.17]
CVX	15.71	400.32	3.87	35.70	0.49	4.15	[0.00]
DD	5.82	64.82	1.90	10.01	0.22	2.88	[0.12]
DIS	7.95	130.39	2.13	12.20	0.37	2.78	[0.00]
GE	9.79	153.91	3.26	21.52	0.53	3.52	[0.00]
HD	7.83	121.54	2.32	13.24	0.44	3.07	[0.39]
IBM	6.63	78.68	2.22	11.72	0.45	2.91	[0.25]
INTC	4.13	31.95	1.75	7.42	0.41	2.82	[0.02]
JNJ	8.23	123.38	2.25	13.72	0.27	2.82	[0.02]
JPM	9.63	150.45	2.99	18.91	0.44	3.20	[0.01]
KO	6.55	83.30	2.07	11.00	0.34	2.90	[0.18]
MCD	12.52	283.70	2.46	18.66	0.11	2.86	[0.06]
MMM	13.84	348.47	2.95	23.13	0.45	3.38	[0.00]
MRK	21.02	774.03	3.70	36.26	0.51	3.89	[0.00]
MSFT	4.69	40.03	1.82	8.40	0.37	2.82	[0.02]
NKE	5.28	55.88	1.82	8.77	0.35	2.71	[0.00]
PFE	5.39	55.00	1.97	9.89	0.38	3.01	[0.91]
PG	9.63	171.20	2.55	16.20	0.46	3.01	[0.94]
TRV	15.45	401.55	3.35	26.92	0.54	3.16	[0.03]
UNH	8.32	126.42	2.78	16.56	0.61	3.55	[0.00]
UTX	7.65	106.79	2.40	13.99	0.43	3.21	[0.00]
VZ	7.49	114.79	2.23	12.43	0.41	3.00	[0.98]
WMT	8.83	176.11	1.99	11.48	0.33	2.59	[0.00]
XOM	13.52	322.34	3.26	26.45	0.39	3.63	[0.00]

Table A3. Estimates of interaction term in HARQ models. ‘level’ is HARQ with $y = RV$, $q = \sqrt{RV}$ and ‘log’ is with $y = \ln(RV)$, $q = \ln(RQ)$. κ is the sample kurtosis of y where $\kappa = 3$ for a Gaussian. τ_h are the HAR t -ratios for the interaction term $q_{t-1}y_{t-1}$ for forecast step size h days. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days).

	Level				log			
	κ	τ_1	τ_5	τ_{22}	κ	τ_1	τ_5	τ_{22}
AXP	262.02	−11.84	−6.52	−4.04	2.95	1.23	1.12	−0.39
BA	76.69	−4.65	−3.67	−3.33	3.19	0.48	−0.69	−0.76
CAT	110.65	−6.40	−2.85	−2.70	3.32	0.69	0.25	0.31
CSCO	36.34	−4.15	−3.54	−4.33	2.90	2.65	1.71	0.61

Table A3. Cont.

	Level				log			
	κ	τ_1	τ_5	τ_{22}	κ	τ_1	τ_5	τ_{22}
CVX	400.32	−5.57	−1.81	−2.12	4.15	1.76	0.91	0.96
DD	64.82	−5.78	−3.01	−3.54	2.88	0.97	0.26	−0.02
DIS	130.39	−3.90	−5.44	−4.60	2.78	−0.08	−0.33	−0.74
GE	153.91	−5.61	−4.79	−2.42	3.52	2.29	0.77	−0.27
HD	121.54	−6.96	−6.33	−4.83	3.07	0.05	−0.22	−0.34
IBM	78.68	−2.68	−5.52	−5.31	2.91	3.42	0.68	0.58
INTC	31.95	−8.25	−3.40	−3.97	2.82	1.73	0.91	1.08
JNJ	123.38	−4.02	−5.29	−4.05	2.82	2.34	0.54	0.37
JPM	150.45	−5.92	−7.57	−5.83	3.20	2.47	1.26	0.11
KO	83.30	−9.07	−6.21	−1.32	2.90	0.16	0.45	0.24
MCD	283.70	−3.78	−3.10	−3.35	2.86	−2.05	−1.78	−2.16
MMM	348.47	−9.81	−6.43	−5.93	3.38	1.33	0.83	1.31
MRK	774.03	−4.70	−5.44	−5.44	3.89	−0.17	−0.42	−0.73
MSFT	40.03	−4.28	−5.22	−2.21	2.82	2.68	1.49	0.70
NKE	55.88	−5.78	−4.20	−4.11	2.71	−0.53	−0.80	−0.11
PFE	55.00	−6.08	−5.86	−5.53	3.01	−2.11	−1.20	−1.13
PG	171.20	−4.68	−5.68	−4.26	3.01	0.76	0.09	0.31
TRV	401.55	−4.15	−3.98	−3.38	3.16	0.98	0.84	0.54
UNH	126.42	−3.38	−3.32	−2.21	3.55	0.37	0.43	1.23
UTX	106.79	−3.06	−4.61	−3.57	3.21	1.73	0.66	1.25
VZ	114.79	−4.63	−4.75	−3.99	3.00	−0.79	−0.19	0.73
WMT	176.11	−4.98	−7.64	−7.70	2.59	1.69	0.92	−0.27
XOM	322.34	−6.14	−2.92	−2.88	3.63	2.07	0.86	0.64

Table A4. Dates when HARQ predictions for RV are negative. The one month rolling forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

	$h = 1$	$h = 5$	$h = 22$
AXP	13 October 2008	13 October 2008	30 September 2008
CAT	8 October 2008	13 October 2008	
	13 October 2008		
CVX	13 October 2008	13 October 2008	16 July 2008
DD			25 July 2007
GE	17 September 2008	17 September 2008	17 September 2008
	22 September 2008	22 September 2008	19 September 2008
			22 September 2008
INTC	13 October 2008	13 October 2008	13 October 2008
JNJ	7 May 2010	7 May 2010	13 October 2008
			7 May 2010
JPM	31 December 2013	13 October 2008	13 October 2008
KO	22 September 2008	22 September 2008	22 September 2008
MMM	13 October 2008	13 October 2008	7 May 2010
	7 May 2010	7 May 2010	
MRK	28 January 2008	28 January 2008	28 January 2008
NKE	7 May 2010	7 May 2010	7 May 2010
TRV	19 September 2008	23 April 2007	23 April 2007
	22 September 2008	19 September 2008	19 September 2008
		22 September 2008	22 September 2008
		13 October 2008	26 September 2008
UNH	22 September 2008	22 September 2008	
WMT	9 October 2008	9 October 2008	9 October 2008
	13 October 2008	13 October 2008	13 October 2008
XOM	13 October 2008	13 October 2008	

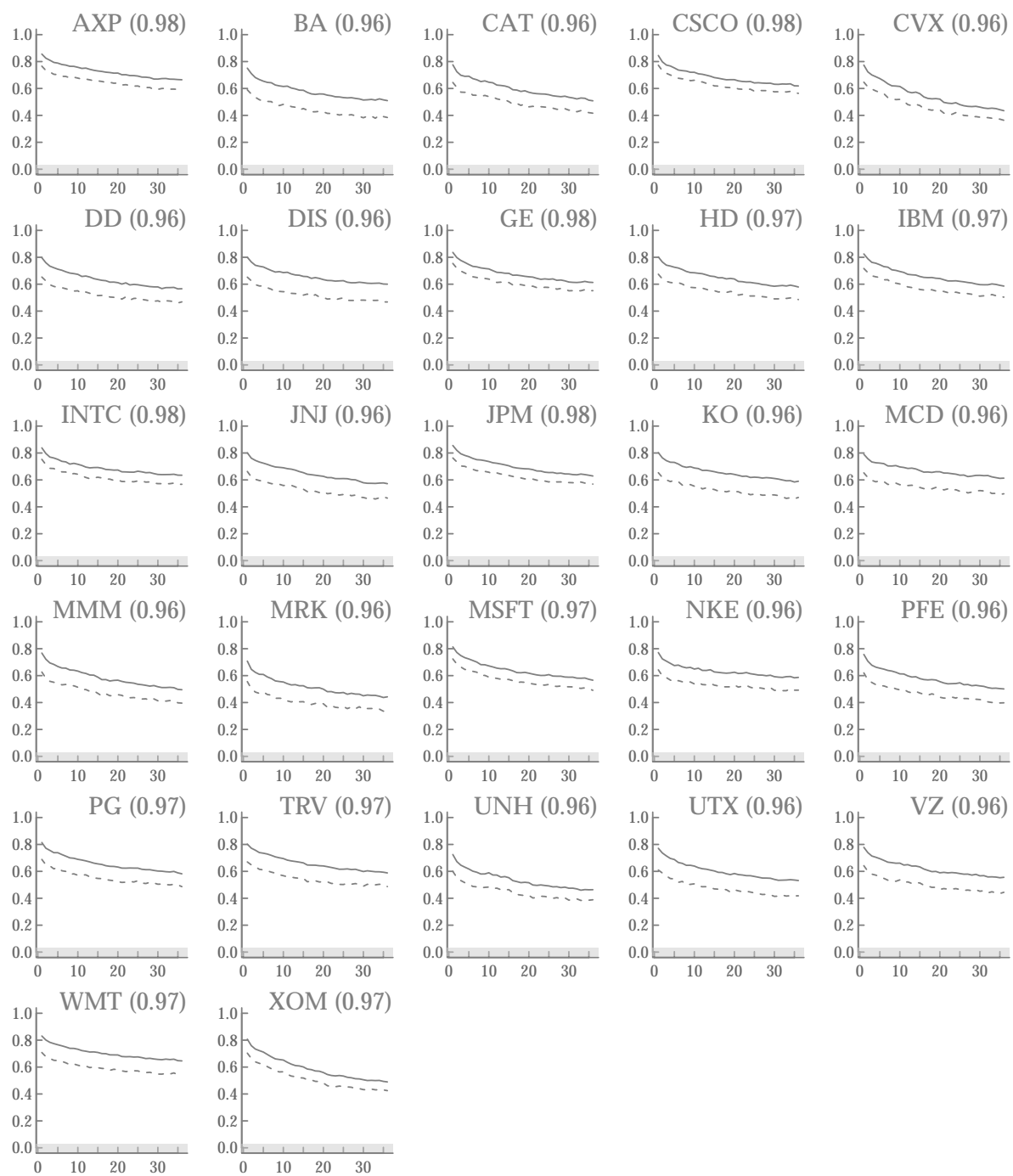


Figure A1. Sample autocorrelations of log realized variance (solid line) and log realized quarticity (dashed line). Numbers in parentheses next to the ticker symbols are the sample correlation between the two log realized series. The shaded area is the two standard error band for a white noise process. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days).

Table A5. ARMA(1,1) estimates for daily log realized quarticity. The ARMA(1,1) model is parametrized as $x_t = \phi x_{t-1} + e_t + \theta e_{t-1}$ where $x_t = \ln(RQ_t) - \mu$, μ is the mean parameter. t are the t -ratios of the estimated parameters, ρ_1 is the first order autocorrelation of the residuals \hat{e}_t , p_5 is the p -value from the Ljung–Box portmanteau test of residual correlation up to lag 5. The daily sample is from 22 April 1997 to 31 December 2013 (4200 trading days).

Ticker	ϕ	(t_ϕ)	θ	(t_θ)	ρ_1	$[p_5]$
AXP	0.989	(379.8)	−0.769	(−56.3)	0.079	[0.000]
BA	0.982	(264.4)	−0.798	(−59.1)	0.055	[0.000]
CAT	0.985	(304.1)	−0.796	(−61.5)	0.074	[0.000]
CSCO	0.983	(296.6)	−0.717	(−46.2)	0.084	[0.000]
CVX	0.977	(241.8)	−0.742	(−55.0)	0.036	[0.035]
DD	0.985	(305.3)	−0.793	(−58.4)	0.063	[0.000]
DIS	0.985	(302.6)	−0.792	(−57.6)	0.060	[0.000]
GE	0.981	(278.2)	−0.713	(−45.8)	0.067	[0.000]
HD	0.988	(349.7)	−0.802	(−64.7)	0.068	[0.000]
IBM	0.983	(292.6)	−0.743	(−53.9)	0.052	[0.000]
INTC	0.984	(299.0)	−0.735	(−47.5)	0.084	[0.000]
JNJ	0.986	(326.6)	−0.799	(−62.8)	0.075	[0.000]
JPM	0.985	(320.0)	−0.736	(−49.9)	0.072	[0.000]
KO	0.987	(334.2)	−0.805	(−63.9)	0.064	[0.000]
MCD	0.992	(422.7)	−0.842	(−72.0)	0.083	[0.000]
MMM	0.982	(275.0)	−0.788	(−59.2)	0.071	[0.000]
MRK	0.980	(241.5)	−0.806	(−57.3)	0.075	[0.000]
MSFT	0.980	(266.5)	−0.724	(−48.5)	0.063	[0.000]
NKE	0.991	(390.8)	−0.844	(−64.2)	0.103	[0.000]
PFE	0.982	(267.6)	−0.793	(−56.2)	0.084	[0.000]
PG	0.984	(302.9)	−0.771	(−56.4)	0.063	[0.000]
TRV	0.986	(322.2)	−0.786	(−61.8)	0.039	[0.009]
UNH	0.981	(253.0)	−0.794	(−54.8)	0.078	[0.000]
UTX	0.982	(272.9)	−0.791	(−57.6)	0.041	[0.007]
VZ	0.983	(282.4)	−0.788	(−56.0)	0.074	[0.000]
WMT	0.989	(373.0)	−0.798	(−61.1)	0.068	[0.000]
XOM	0.976	(239.6)	−0.709	(−48.6)	0.051	[0.000]

Appendix A.4. Forecast Error Diagnostics

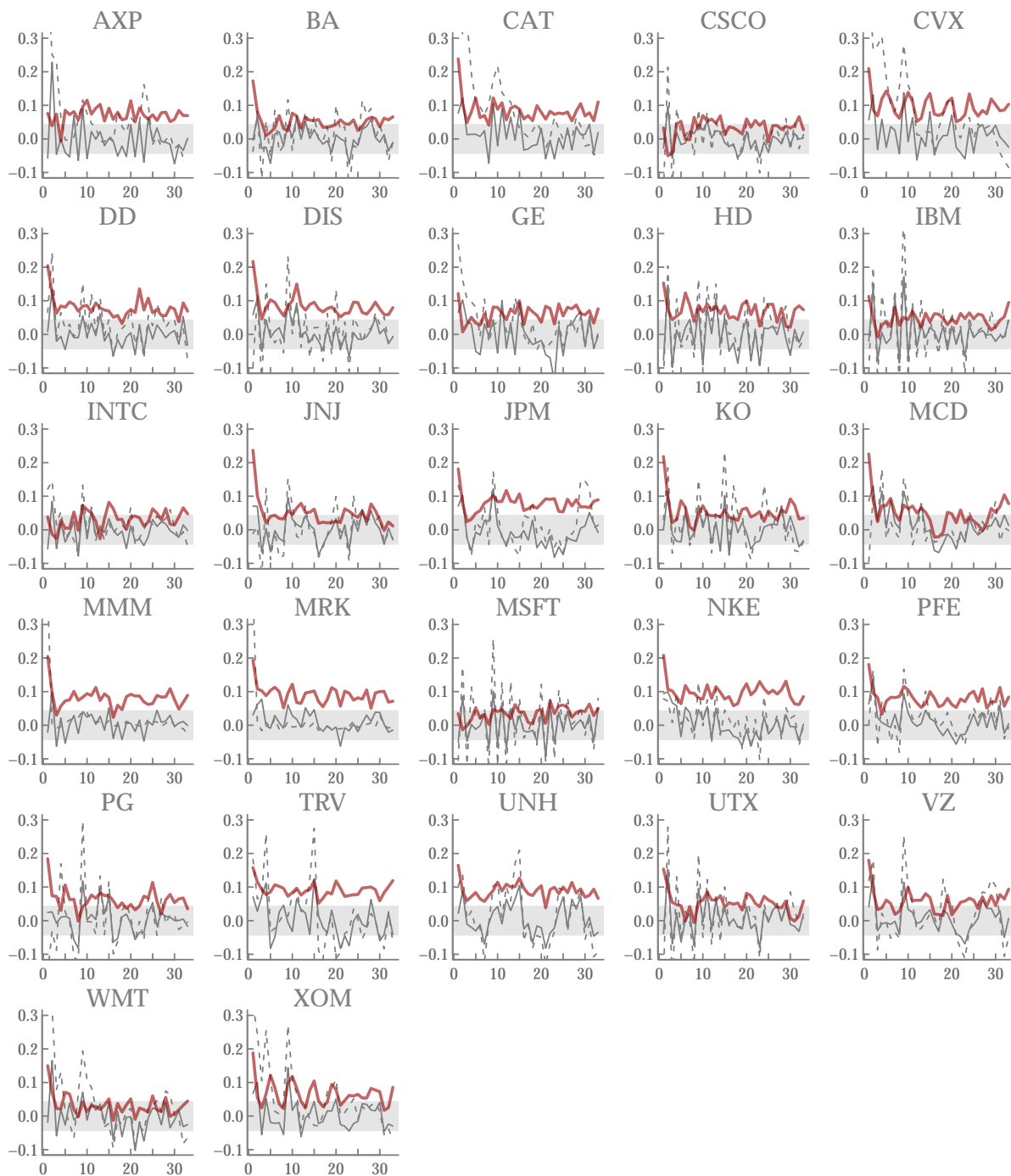


Figure A2. Autocorrelations of forecast errors $e_t = y_t^{(h)} - \hat{y}_t^{(h)}$ for stepsize $h = 1$ day. The autocorrelations for HAR with $y = \sqrt{RV}$ (gray), for HARQ with $y = RV$ (dashed gray), and for RQ-in-mean with $y = \ln(RV)$ (red line) should all be white noise if e_t is *iid*. The gray shaded area is the two standard error band for a white noise series. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

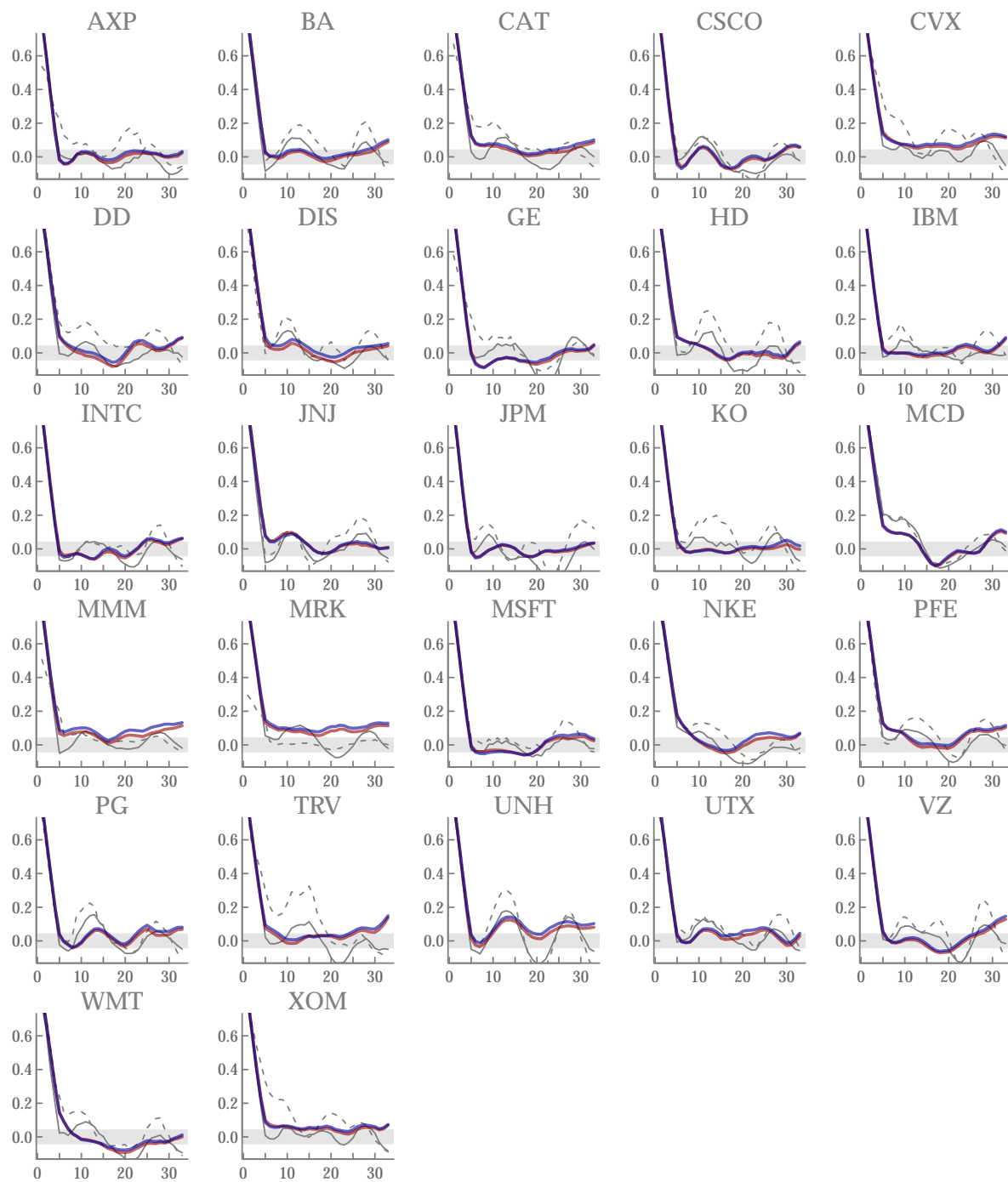


Figure A3. Autocorrelations of forecast errors $e_t = y_t^h - \hat{y}_t^{(h)}$ for stepsize $h = 5$ days. The autocorrelations for HAR with $y = \sqrt{RV}$ (gray), for HARQ with $y = RV$ (dashed gray), and for RQ-in-mean with log of average RV (red) and with log of average \sqrt{RV} (blue) should all be white noise if e_t is *iid*. The gray shaded area is the two standard error band for a white noise series. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

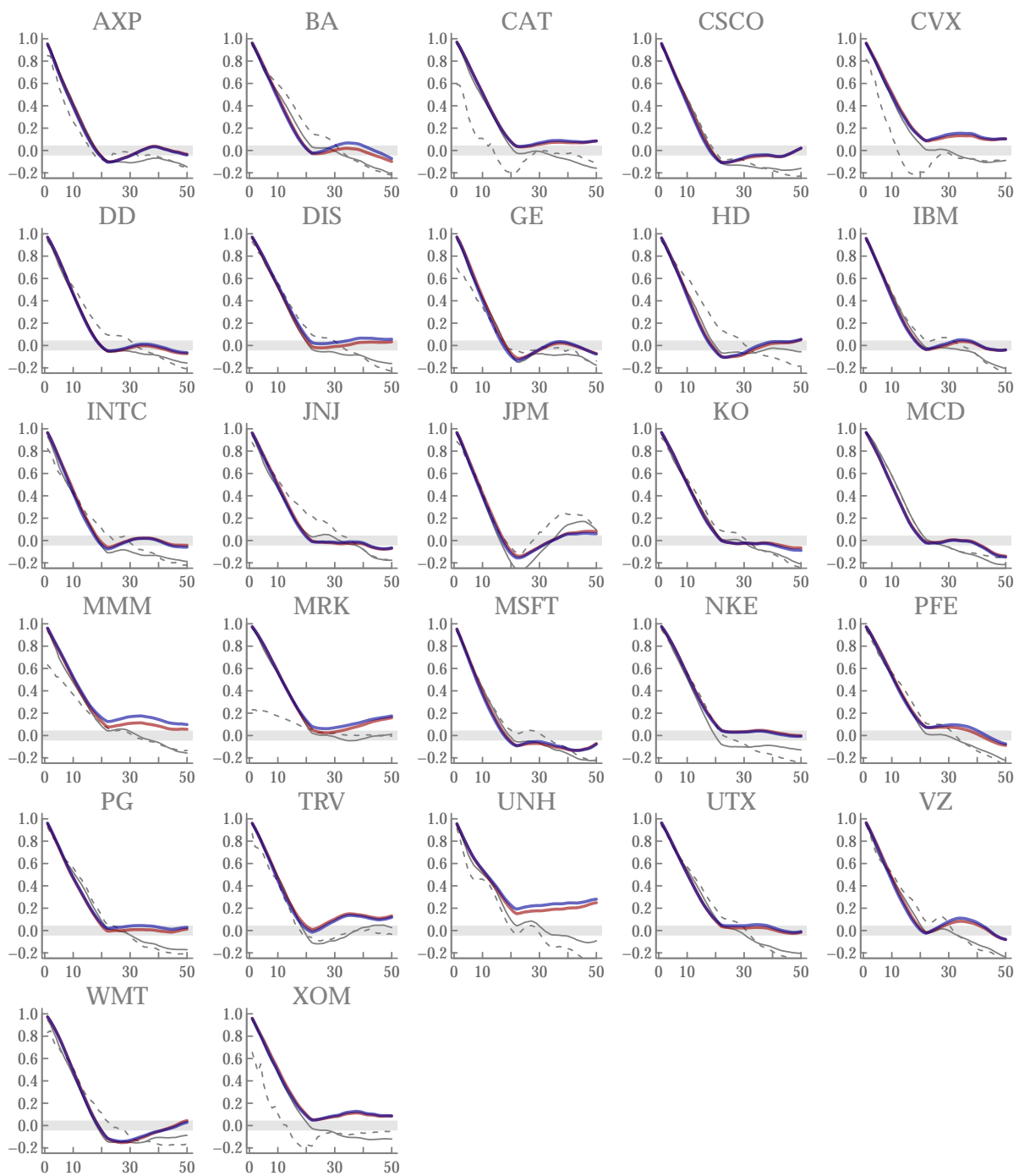


Figure A4. Autocorrelations of forecast errors $e_t = y_t^h - \hat{y}_t^{(h)}$ for stepsize $h = 22$ days. The autocorrelations for HAR with $y = \sqrt{RV}$ (gray), for HARQ with $y = RV$ (dashed gray), and for RQ-in-mean with log of average RV (red) and with log of average \sqrt{RV} (blue) should all be white noise if e_t is *iid*. The gray shaded area is the two standard error band for a white noise series. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

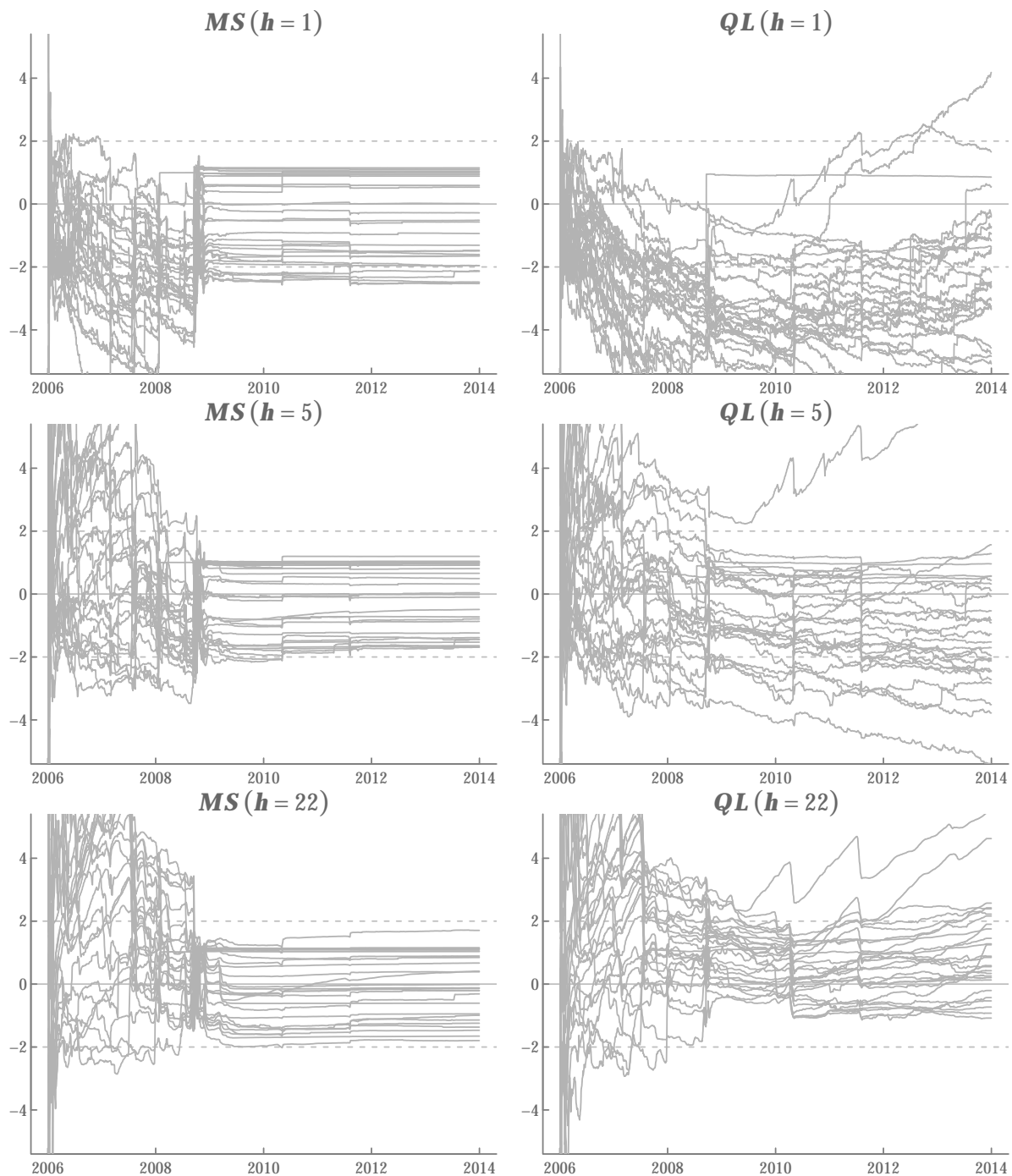
Appendix A.5. Running *t*-Ratios

Figure A5. Running *t*-ratios of Diebold and Mariano [27] test of equal predictive accuracy of *h*-step RV rolling forecasts. Each panel has 27 lines for the 27 stocks in the sample. The baseline forecast is from the HARQ model (no transformation) and the comparison forecast is the exponential transform of the log forecast from model (Section 3.1) assuming log-normality. A positive value indicates better forecast from the comparison model against the baseline HARQ forecasts. *MS* for equality of mean squared error loss and *QL* for QLIKE loss of Patton [24]. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

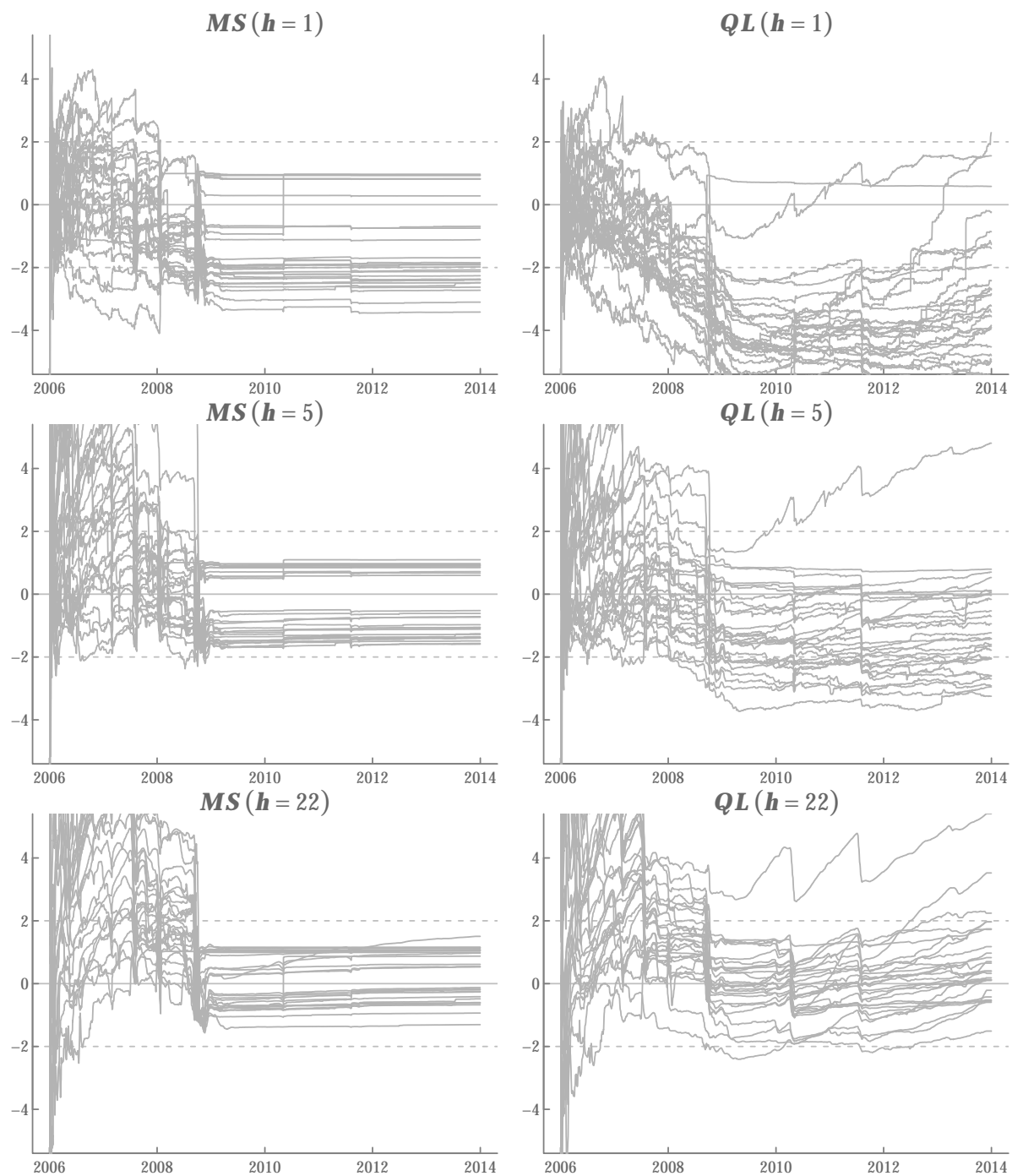


Figure A6. Running t -ratios of Diebold and Mariano [27] test of equal predictive accuracy of h -step RV rolling forecasts. Each panel has 27 lines for the 27 stocks in the sample. The baseline forecast is from the HARQ model (no transformation) and the comparison forecast is a simple exponential transform of the log forecast from model (Section 3.1). A positive value indicates better forecast from the comparison model against the baseline HARQ forecasts. MS for equality of mean squared error loss and QL for QLIKE loss of Patton [24]. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

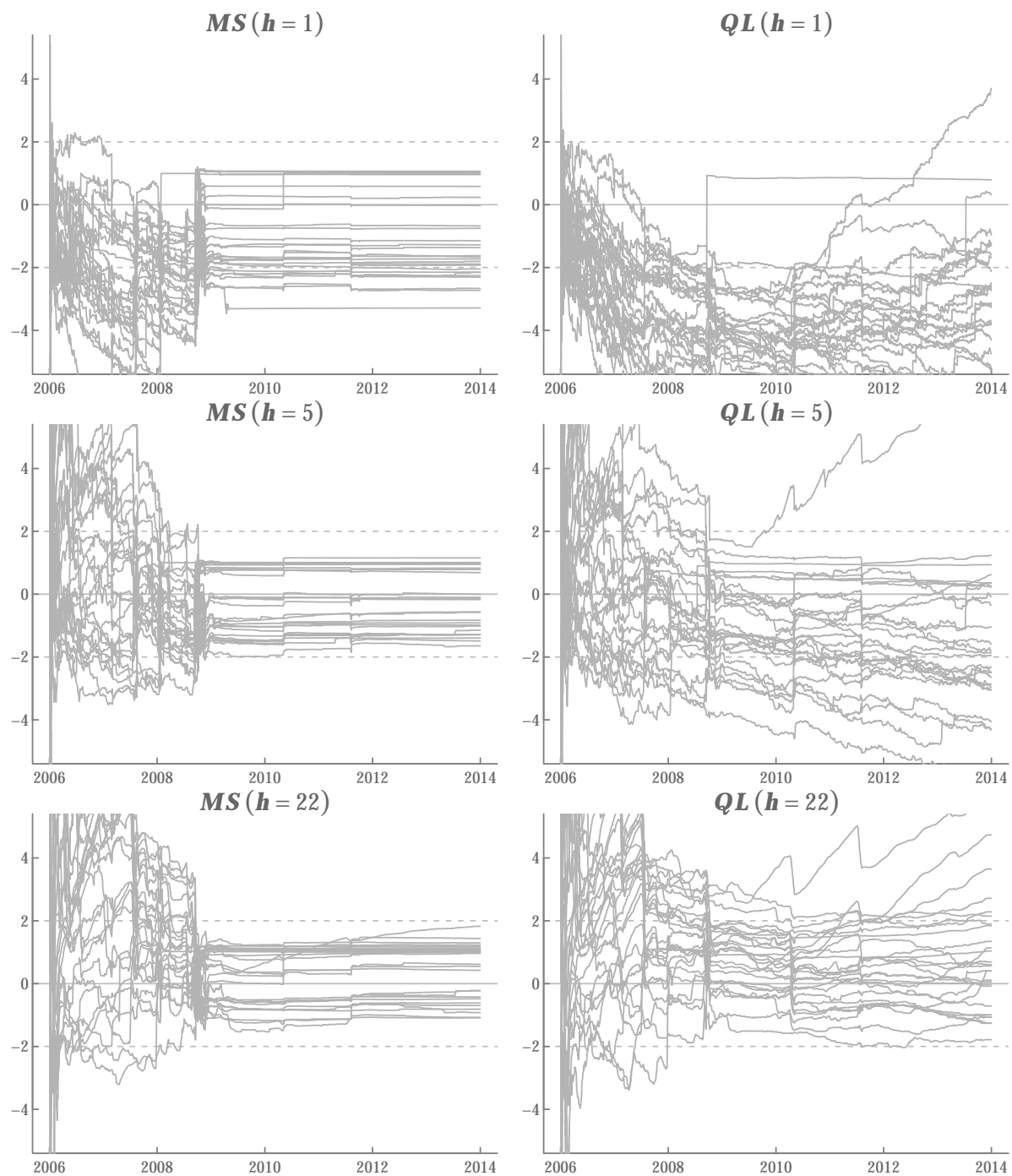


Figure A7. Running t -ratios of Diebold and Mariano [27] test of equal predictive accuracy of h -step RV rolling forecasts. Each panel has 27 lines for the 27 stocks in the sample. The baseline forecast is from the HARQ model (no transformation) and the comparison forecast is the exponential transform of the log forecast from model (Section 3.1) adjusted with a smearing factor [25]. A positive value indicates better forecast from the comparison model against the baseline HARQ forecasts. MS for equality of mean squared error loss and QL for QLIKE loss of Patton [24]. The rolling forecasts are generated with a rolling estimation window of 8.67 years (104 months) followed by a one month prediction window shifted each month over 96 windows. The forecast sample is from 3 January 2006 to 31 December 2013 (2013 trading days).

References

1. Corsi, F. A Simple Approximate Long-Memory Model of Realized Volatility. *J. Financ. Econom.* **2009**, *7*, 174–196. [\[CrossRef\]](#)
2. Rockinger, M.; Jondeau, E. Entropy densities with an application to autoregressive conditional skewness and kurtosis. *J. Econom.* **2002**, *106*, 119–142. [\[CrossRef\]](#)
3. Jondeau, E.; Rockinger, M. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *J. Econ. Dyn. Control* **2003**, *27*, 1699–1737. [\[CrossRef\]](#)
4. Bali, T.G.; Mo, H.; Tang, Y. The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *J. Bank. Financ.* **2008**, *32*, 269–282. [\[CrossRef\]](#)
5. Ding, Z.; Granger, C.W.; Engle, R.F. A long memory property of stock market returns and a new model. *J. Empir. Financ.* **1993**, *1*, 83–106. [\[CrossRef\]](#)
6. Comte, F.; Renault, E. Long memory in continuous-time stochastic volatility models. *Math. Financ.* **1998**, *8*, 291–323. [\[CrossRef\]](#)
7. Comte, F.; Coutin, L.; Renault, E. Affine fractional stochastic volatility models. *Ann. Financ.* **2012**, *8*, 337–378. [\[CrossRef\]](#)
8. Gatheral, J.; Jaisson, T.; Rosenbaum, M. Volatility is rough. *Quant. Financ.* **2018**, *18*, 933–939. [\[CrossRef\]](#)
9. Huang, D.; Schlag, C.; Shaliastovich, I.; Thimme, J. Volatility-of-Volatility Risk. *J. Financ. Quant. Anal.* **2019**, *54*, 2423–2452. [\[CrossRef\]](#)
10. Da Fonseca, J.; Zhang, W. Volatility of volatility is (also) rough. *J. Futur. Mark.* **2019**, *39*, 600–611. [\[CrossRef\]](#)
11. Hansen, P.R.; Huang, Z.; Shek, H.H. Realized GARCH: A joint model for returns and realized measures of volatility. *J. Appl. Econom.* **2012**, *27*, 877–906. [\[CrossRef\]](#)
12. Bollerslev, T.; Patton, A.J.; Quaedvlieg, R. Exploiting the errors: A simple approach for improved volatility forecasting. *J. Econom.* **2016**, *192*, 1–18. [\[CrossRef\]](#)
13. Buccheri, G.; Corsi, F. HARK the SHARK: Realized Volatility Modeling with Measurement Errors and Nonlinear Dependencies. *J. Financ. Econom.* **2021**, *19*, 614–649. [\[CrossRef\]](#)
14. Creal, D.; Koopman, S.J.; Lucas, A. Generalized Autoregressive Score Models with Applications. *J. Appl. Econom.* **2013**, *28*, 777–795. [\[CrossRef\]](#)
15. Ding, Y.D. A simple joint model for returns, volatility and volatility of volatility. *J. Econom.* **2021**, *in press*. [\[CrossRef\]](#)
16. Ghysels, E.; Santa-Clara, P.; Valkanov, R. There is Risk-Return Trade-off After all. *J. Financ. Econ.* **2005**, *76*, 509–548. [\[CrossRef\]](#)
17. Corsi, F.; Mittnik, S.; Pigorsch, C.; Pigorsch, U. The Volatility of Realized Volatility. *Econom. Rev.* **2008**, *27*, 46–78. [\[CrossRef\]](#)
18. Andersen, T.G.; Bollerslev, T.; Diebold, F.X.; Labys, P. Modeling and Forecasting Realized Volatility. *Econometrica* **2003**, *71*, 529–626. [\[CrossRef\]](#)
19. Geweke, J.; Porter-Hudak, S. The Estimation and Application of Long Memory Time Series Model. *J. Time Ser. Anal.* **1983**, *4*, 221–238. [\[CrossRef\]](#)
20. Hansen, P.R.; Huang, Z. Exponential GARCH Modeling With Realized Measures of Volatility. *J. Bus. Econ. Stat.* **2016**, *34*, 269–287. [\[CrossRef\]](#)
21. Engle, R.F.; Lilien, D.M.; Robins, R.P. Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. *Econometrica* **1987**, *55*, 391–407. [\[CrossRef\]](#)
22. Nelson, D.B. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **1991**, *59*, 347–370. [\[CrossRef\]](#)
23. Barndorff-Nielsen, O.E.; Shephard, N. How Accurate is the Asymptotic Approximation to the Distribution of Realised Variance? In *Identification and Inference for Econometric Models*; Andrews, D.W.K., Stock, J.H., Eds.; Cambridge University Press: Cambridge, UK, 2005; Chapter 13, pp. 306–331.
24. Patton, A.J. Volatility forecast comparison using imperfect volatility proxies. *J. Econom.* **2011**, *160*, 246–256. [\[CrossRef\]](#)
25. Duan, N. Smearing Estimate: A Nonparametric Retransformation Method. *J. Am. Stat. Assoc.* **1983**, *78*, 605–610. [\[CrossRef\]](#)
26. Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, 7th ed.; Cengage Learning: Boston, MA, USA, 2020.
27. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263.
28. Barndorff-Nielsen, O.E.; Hansen, P.R.; Lunde, A.; Shephard, N. Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica* **2008**, *76*, 1481–1536.
29. Bickel, P.J.; Li, B.; Tsybakov, A.B.; van de Geer, S.A.; Yu, B.; Valdés, T.; Rivero, C.; Fan, J.; van der Vaart, A. Regularization in statistics. *Test* **2006**, *15*, 271–344. [\[CrossRef\]](#)
30. Hastie, T. Ridge Regularization: An Essential Concept in Data Science. *Technometrics* **2020**, *62*, 426–433. [\[CrossRef\]](#)
31. Barndorff-Nielsen, O.E.; Shephard, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 253–280. [\[CrossRef\]](#)
32. Barndorff-Nielsen, O.E.; Shephard, N. Estimating Quadratic Variation Using Realised Variance. *J. Appl. Econom.* **2002**, *17*, 457–477. [\[CrossRef\]](#)
33. Barndorff-Nielsen, O.E.; Shephard, N. Power and Bipower Variation with Stochastic Volatility and Jumps. *J. Financ. Econom.* **2004**, *2*, 1–37. [\[CrossRef\]](#)
34. Barndorff-Nielsen, O.E.; Shephard, N.; Winkel, M. Limit theorems for multipower variation in the presence of jumps. *Stoch. Process. Appl.* **2006**, *116*, 796–806. [\[CrossRef\]](#)
35. Andersen, T.G.; Bollerslev, T.; Diebold, F.X. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Rev. Econ. Stat.* **2007**, *89*, 701–720. [\[CrossRef\]](#)