Clements, Adam & Herrera, Rodrigo
(2019)
Moderate and Extreme Volatility: Do the Magnitude of Returns Matter for Forecasting?
[Working Paper]
(Unpublished)

https://doi.org/10.2139/ssrn.3443259

Moderate and extreme volatility: Do the magnitude of returns matter for forecasting?

Rodrigo Hererra[a], Adam Clements[b,*]

[a]*Facultad de Economía y Negocios, Universidad de Talca, Talca, Chile.*

[b]*School of Economics and Finance, Queensland University of Technology, Australia.*

Abstract

This paper proposes a novel decomposition of realized volatility (RV) into moderate and extreme realized volatility estimates. These estimates behave like long and short term components of volatility, and are very different from either realized semi-variance or the continuous and jump components of volatility. Within the standard linear HAR framework, a forecast comparison exercise using index returns shows that employing the new decomposition leads to forecasts that are often superior to the competing forecasts based on existing realized measures.

JEL classification: C22, C53, C58.

Keywords: Realized volatility, HAR model, moderate volatility, extreme volatility, forecasting.

---

*Corresponding author

Email addresses: `rodriherrera at utalca.cl` (Rodrigo Hererra), `a.clements@qut.edu.au` (Adam Clements)

Preprint submitted to Elsevier *September 9, 2019*

1. Introduction

Forecasting the volatility of financial asset returns is an important issue in the context of risk management, portfolio construction, and derivative pricing. As such, a great deal of research effort has focused on developing and evaluating volatility forecasting models. With the widespread availability of high-frequency financial data, the recent literature has focused on employing realized measurers of volatility to build forecasting models. There is a wide range of realized measures of volatility, building on the seminal work on realized volatility (RV) by Andersen and Bollerslev (1998) and Andersen et al. (2001) among others. The heterogeneous autoregressive (HAR) model of Corsi (2009) was designed to parsimoniously capture the strong persistence typically observed in RV and has become the workhorse of this literature due to its consistently good forecasting performance, and that standard linear regression techniques can be used for its estimation.

Moving beyond simple RV, there have been numerous developments in terms of realized measures of volatility. For instance, Barndorff-Nielsen and Shephard (2006) proposed decomposing RV into its continuous and jump components, and Patton and Sheppard (2015) proposed a decomposition into positive and negative semi-variance. The basic HAR structure has easily been extended to incorporate these additional components of volatility. There have also been many extensions to the HAR structure itself that have considered more complex models, Fengler et al. (2015) proposed a non-parametric model, Audrino et al. (2018) considered more flexible lag structures, and Bollerslev et al. (2016) and Buccheri and Corsi (2017) considered estimation error and time-varying parameters.

The goal here is not to propose another extension to the basic HAR framework, but to propose a novel but simple decomposition of RV to be used within the existing linear HAR framework. A decomposition of RV into moderate and extreme realized volatility estimates is proposed as an alternative to good and bad volatility of Patton and Sheppard (2015). By using moderate and extreme volatility, it is possible to consider if the magnitude of returns, as opposed to their sign, is important for constructing volatility estimates for forecasting purposes.

In comparison to existing RV, it is shown that these components of RV behave like long and short term components of volatility. It is also shown that extreme volatility behaves very differently to the jump component of RV. A forecast comparison exercise shows that employing the new decomposition within the standard HAR framework leads to forecasts that are superior to the competing forecasts based on existing realized measures.

2. Data and realized measures of volatility

The main empirical analysis presented below are based on the SPYDR ETF. Five-minute intraday data for the S&P 500 SPYDR was downloaded from Thomson Reuters Datascope for the period 3 January 2000 to 28 June 2019, representing 4798 trading days. While not reported here in the main paper, results based on other indices, Nasdaq, DAX and FTSE, are also presented in an online appendix.

The following realized measures of volatility are based on a single asset for which the log-price process P within the active part of a trading day evolves in continuous time as:

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

where μ and σ are the instantaneous drift and volatility processes, respectively, and W is a standard Brownian motion (Wiener process). From prices, the i^{th} Δ-period return within day t is defined as:

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

where $M = 1/\Delta$ is the sampling frequency.

In the simplest case, the latent one-day integrated variance defined by:

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

is unobservable, although it can be consistently estimated by the one-day realized variance (RV):

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

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

$$RV_t = IV_t + \eta_t, \quad MN(0, 2\Delta IQ_t),$$

where MN denotes a mixed normal distribution and $IQ_t = \int_{t-1}^{t} \sigma_s^4 ds$ is the integrated quarticity (IQ) which can be consistently estimated by the realized quarticity (RQ):

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

Extending the standard diffusion process in equation 1 to a jump-diffusion process gives:

$$dP_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t, \tag{4}$$

where the $\kappa_t dq_t$ term refers to a pure jump component, where κ_t is the size of a jump and $dq_t = 1$ if there is a jump at time t (and 0 otherwise). Under the jump diffusion process, the IV can be estimated with bi-power variance

$$BV_t \equiv \frac{\pi}{2}(\frac{M}{M-1}) \sum_{j=2}^{M} |r_{t,j}||r_{t,j-1}|. \tag{5}$$

The difference between the two estimators, RV and BV, can be used to estimate the contribution of the jump component to total variance. However, the theoretical justification for equations (3) and (5) is based on the notion of increasingly finer sampled returns, or $M \to \infty$. Of course, any practical implementation with a finite fixed sampling frequency, or $M < \infty$, is invariably subject to measurement error, and hence, it is desirable to treat small jumps as measurement error and only identify significantly large jumps. Barndorff-Nielsen and Shephard (2006) developed such a test (BNS test), which is modified to account for microstructure noise and improve the finite sample performance following Huang and Tauchen (2005) as:

$$Z_t \equiv \Delta^{-1/2} \times \frac{[RV_t - BV_t]RV_t^{-1}}{[(\mu_1^{-4} + 2\mu_1^{-2} - 5)max\{1, TQ_t BV_t^{-2}\}]^{1/2}}, \tag{6}$$

with $\Delta = 1/M$ and TQ_{t+1} the realized tripower quarticity measure, defined as

$$TQ_t = \Delta^{-1}\mu_{4/3}^{-3} \sum_{j=3}^{M} |r_{t,j}|^{4/3}|r_{t,j-1}|^{4/3} r_{t,j-2}|^{4/3}, \tag{7}$$

with $\mu_{4/3} \equiv E(|Z|^{4/3}) = 2^{2/3} \cdot \Gamma(7/6) \cdot \Gamma(1/2)^{-1}$. The BV and TQ measures are generated based on staggered returns to remove microstructure noise. At least one significant jump on day t is identified by realizations of Z_t in excess of some critical value Φ_α, with α representing the significance level of the test applied to a daily frequency, and set to be $\alpha = 1\%$ in the subsequent empirical analysis. The jump size for days with jumps detected are measured by the difference between RV and BV and will be denoted as J_t, with the continuous BV denoted below simply as C_t.

Patton and Sheppard (2015) proposed a very different decomposition of RV. While most earlier realized measures only employed even powers of intraday returns, Patton and Sheppard (2015) built on the semi-variance idea of Barndorff-Nielsen et al. (2010) to consider the impact of signed returns on future volatility by using the realized semi-variance estimators:

$$\begin{aligned} RV_t^- &= \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{r_{t,i}<0} \\ RV_t^+ &= \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{r_{t,i}>0} \end{aligned} \tag{8}$$

The estimates of moderate and extreme volatility represent an alternative decomposition of RV to that of Patton and Sheppard (2015), one based not on the signs of intraday returns but their tail distribution. To begin, take the unconditional volatilities of intraday returns for each intraday period across all days, σ_i, $i = 1, 2 \ldots, M$. Choose a tail probability of α upon which a negative and positive threshold are defined, $r_i^- = F^{-1}(\alpha)\sigma_i$ and $r_i^+ = F^{-1}(1-\alpha)\sigma_i$ where F is the normal CDF. Based on these thresholds, three measures of realized volatility, extreme negative (REX_t^-) and positive (REX_t^+), and moderate (REX_t^m) volatility can be defined as follows:

$$\begin{aligned} REX_t^- &= \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{r_{t,i}<=r_i^-} \\ REX_t^+ &= \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{r_{t,i}>=r_i^+} \\ REX_t^m &= \sum_{i=1}^{M} r_{t,i}^2 \mathbb{I}_{r_i^->r_{t,i}>r_i^+} \end{aligned} \tag{9}$$

The motivation behind this definition of extreme volatilities is as follows. Assume for the moment that during periods of financial turmoil, returns $r_{t,i}$ exhibit heavy tailed behavior as $r_i^+ \to \infty$, i.e.:

$$\mathbb{P}\left(r_{t,i} > r_i^+\right) \sim r_i^{+\,-\gamma} \mathcal{L}\left(r_i^+\right),$$

where $\mathcal{L}\left(r_i^+\right)$ is some regularly varying function such that $\lim_{t\to\infty} \frac{\mathcal{L}(ts)}{\mathcal{L}(t)} = 1$, $s > 0$, and γ is the tail index, which is related to the moments of the distribution function. Then, from Feller's convolution theorem, the tail of the distribution of the sum of $r_{t,i}$ is given by:

$$\sum_{i=1}^{M} \mathbb{P}\left(r_{t,i} > r_i^+\right) = \mathbb{P}\left(\sum_{i=1}^{M} r_{t,i} > r_i^+\right) \sim M r_i^{+-\gamma} \mathcal{L}\left(r_i^+\right),$$

while if these exhibit, for instance, a standard normal distribution then

$$\sum_{i=1}^{M} \mathbb{P}\left(r_{t,i} > r_i^+\right) \geq \mathbb{P}\left(\sum_{i=1}^{M} r_{t,i} > r_i^+\right) = \mathbb{P}\left(\sqrt{M} r_{t,i} > r_i^+\right) \sim \frac{\sqrt{M}}{r_i^+ \sqrt{2\pi}} \exp\left(-\frac{\left(r_i^+\right)^2}{2M}\right).$$

Thus, the probability of observing extreme positive volatility REX_t^+ is higher for returns displaying a heavy tail behavior as $r_i^+ \to \infty$, and close to zero, for the normal distribution because the tails are exponentially bounded. As a consequence, building volatility measures and forecasting models on extreme volatility, reflects dynamic tail behavior while avoiding direct estimation of a tail index which can be a difficult task. Similar results can be obtained for REX_t^-.

Figure 1 plots RV_t in the top panel (with the y-axis limited to highlight the movements during lower periods). This behaviour of equity market RV is well known, and is dominated by the period of the GFC. Periods of higher volatility were also experienced around the collapse of the technology bubble early in the sample (2000-2004) and again later surrounding the European debt crisis (2009-2012). The middle panel of Figure 1 plots REX_t^m. It is clear that REX_t^m behaves very differently from total RV_t and appears to capture a long-term slow moving component of volatility, similar to those contained in component models for volatility proposed by Amado and Teräsvirta (2017). REX_t^m moves with the overall level in RV_t, and was consistently higher during the 2000-2004, 2008-2012 periods, but does not contain any of the larger spikes observed in RV_t. The bottom panel of Figure 1 plots both REX_t^+ (dashed line) and REX_t^- (dot-dashed line). Both REX_t^+ and REX_t^- behave in a similar fashion and are very different from REX_t^m, reflecting the more short-term behaviour of RV_t. In comparison to RV_t, there are no fluctuations in the overall level of REX_t^+ and REX_t^- (this feature is captured by REX_t^m), they are generally very low and then spike during periods of much higher volatility.

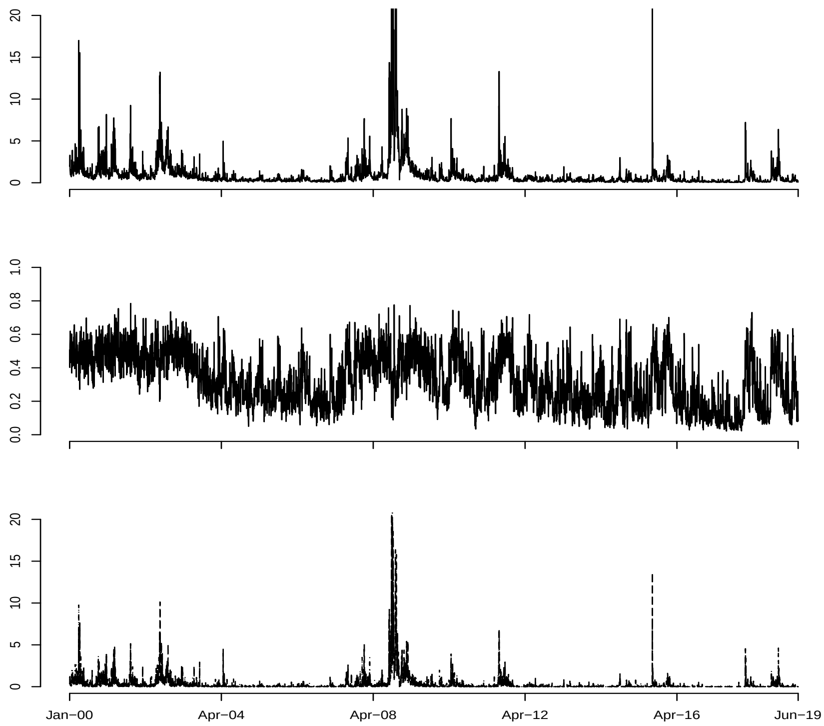

Figure 1: Top panel: RV_t. Middle panel: REX_t^m. Bottom Panel: REX_t^+ (dashed line) and REX_t^- (dot-dashed line).

To gain a deeper understanding of the behaviour of the different realized measures, Figure 2 plots the AutoCorrelation Functions (ACF) for the different series out to a maximum of 50. In each panel, the ACF for RV_t is shown as a solid line as a point of reference. In the top panel, the ACFs for C_t (dashed line) and J_t (dotted line) are shown along with that for RV_t. This shows that the persistence in the continuous diffusive volatility, C_t is very similar to that in total RV_t, while the jump component J_t exhibits consistently positive, though a very small degree of autocorrelation. The middle panel plots the ACFs for RV_t^-

(dashed line) and RV_t^+ (dotted line), where it is clear that RV_t^- exhibits a very similar degree of persistence to RV_t while the persistence in RV_t^+ is consistently lower. Finally, the bottom panel plots the ACFs for REX_t^- (dotted line), REX_t^m (dot-dashed line) and REX_t^+ (dashed line) against that from RV_t which reveals a very interesting pattern. REX_t^m is consistently more persistent than RV_t, with the ACF decaying much slower than RV_t. On the other hand, the persistence in both REX_t^- and REX_t^+ is consistently lower than that of RV_t, with REX_t^+ exhibiting the lowest persistence of all the components. These differences in persistence, as reflected in the ACFs, indicate that by decomposing RV into moderate and extreme volatility, longer- and shorter-term components of total volatility are revealed.

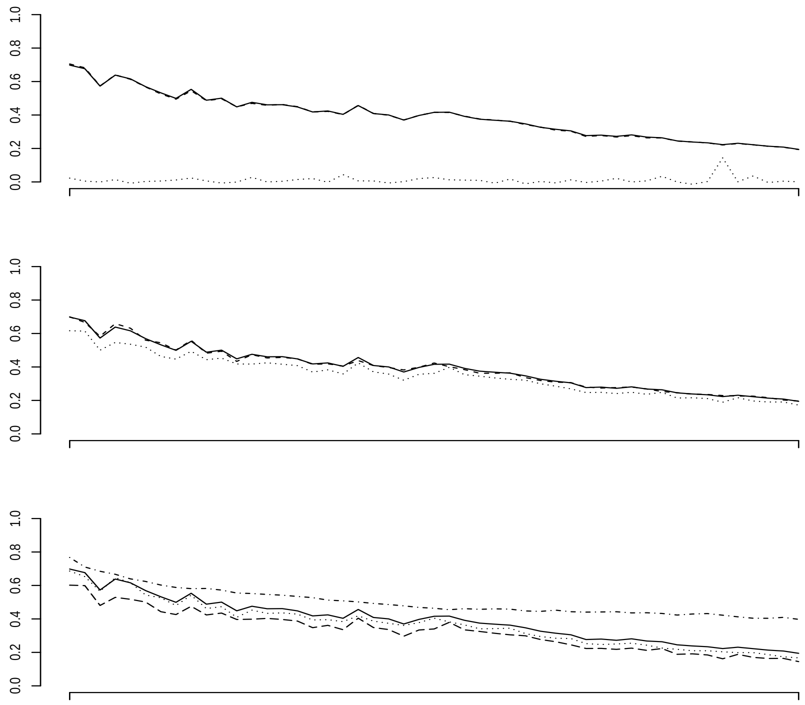

Figure 2: Autocorrelation function (ACF) plots to a maximum lag of 50. In each panel, the ACF for RV_t is shown as a solid line as a point of reference. Top panel: ACF for C_t (dashed line) and J_t (dotted line). Middle Panel: ACF for RV_t^- (dashed line) and RV_t^+ (dotted line). Bottom panel: ACF for REX_t^- (dotted line), REX_t^m (dot-dashed line) and REX_t^+ (dashed line).

3. Methodology

This section outlines the HAR models that utilise the realized measures discussed in the previous section. This is followed by a brief description of the approach used for forecast comparison.

3.1. HAR models

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

The original HAR model specifies RV as a linear function of daily, weekly and monthly realized variance components, and can be expressed (in logarithmic form here) as:

$$\ln(RV_t) = \beta_0 + \beta_1 \ln(RV^d_{t-1}) + \beta_2 \ln(RV^w_{t-1}) + \beta_3 \ln(RV^m_{t-1}) + u_t, \qquad (10)$$

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

Bollerslev et al. (2016) recently proposed an easily implemented, and by OLS estimated, extension of the HAR model dubbed the HARQ model, which accounts for the error with which RV is estimated by using RQ. Bollerslev et al. (2016) find that, at least for short-term forecasting, a simplified version of the full HARQ model is:

$$\ln(RV_t) = \beta_0+(\beta_1+\beta_{1Q}(RV^d_{t-1}/RQ^d_{t-1}))\ln(RV^d_{t-1})+\beta_2\ln(RV^w_{t-1})+\beta_3\ln(RV^m_{t-1})+u_t, \tag{11}$$

where RQ^d_{t-1} is the lagged quarticity, and RV^d_{t-1}/RQ^d_{t-1} represents the estimation error associated with the logarithm of RV. Bollerslev et al. (2016) find that this simplified model is useful as most of the attenuation bias in the forecasts (due to RV being less persistent than unobserved IV) is due to the estimation error in RV^d_{t-1}. Overall, this framework allows for less weight to be placed on historical observations of RV when the measurement error is higher. This model will be denoted below as the HAR^Q model.

Patton and Sheppard (2015) use the concept of realized semi-variance (RSV) of Barndorff-Nielsen et al. (2010) for forecasting total RV. Here, a fully flexible version using all lags of the negative (RV^-) and positive (RV^+) semi-variances is employed:

$$\begin{aligned}\ln(RV_t) &= \beta_0 + \beta_1\ln(RV^{d-}_{t-1}) + \beta_2\ln(RV^{w-}_{t-1}) + \beta_3\ln(RV^{m-}_{t-1}) \\ &+ \beta_4\ln(RV^{d+}_{t-1}) + \beta_5\ln(RV^{w+}_{t-1}) + \beta_6\ln(RV^{m+}_{t-1}) + u_t \end{aligned} \tag{12}$$

where lags of the semiavariances are constructed in the same way as RV, as discussed in the context of equation 10. This model will be denoted below as the HAR^{RSV} model.

Andersen et al. (2007) utilised the continuous and jump components of RV within the HAR framework for the purposes of forecasting total RV. The model estimated here based on the continuous component C_t and jumps J_t is:

$$\begin{aligned}\ln(RV_t) &= \beta_0 + \beta_1 \ln(C^d_{t-1}) + \beta_2 \ln(C^w_{t-1}) + \beta_3 \ln(C^m_{t-1}) \\ &+ \beta_4 \ln(J^d_{t-1}) + \beta_5 \ln(J^w_{t-1}) + \beta_6 \ln(J^m_{t-1}) + u_t. \qquad (13)\end{aligned}$$

Lags of C_t are again constructed in the same way as RV in the context of equation 10. Lags of J_t are constructed in the following manner, $\ln(j^d_{t-1}) = \ln(J_{t-1}+1)$, $\ln(J^w_{t-1}) = \frac{1}{5}\sum_{i=1}^{5} \ln(J_{t-i}+1)$, $\ln(J^m_{t-1}) = \frac{1}{22}\sum_{i=1}^{22} \ln(J_{t-i}+1)$. This model will be denoted below as the HAR^{CJ} model.

To harness the moderate and extreme volatility estimates from equation 9, the following HAR structure is employed

$$\begin{aligned}\ln(RV_t) &= \beta_0 + \beta_1 \ln(REX^{d-}_{t-1}) + \beta_2 \ln(REX^{w-}_{t-1}) + \beta_3 \ln(REX^{m-}_{t-1}) \\ &+ \beta_4 \ln(REX^{dm}_{t-1}) + \beta_5 \ln(REX^{wm}_{t-1}) + \beta_6 \ln(REX^{mm}_{t-1}) \\ &+ \beta_7 \ln(REX^{d+}_{t-1}) + \beta_8 \ln(REX^{w+}_{t-1}) + \beta_9 \ln(REX^{m+}_{t-1}) + u_t. \qquad (14)\end{aligned}$$

where again lags of REX^-_t, REX^m_t and REX^+_t are constructed in the same manner. This model will be denoted below as the HAR^{RE} model.

Define extreme (irrespective of positive or negative) volatility as $REX_t = REX^-_t + REX^+_t$. Based on total extreme volatility, a simplified version of equation 14 is estimated which utilises REX_t, it is:

$$\begin{aligned}\ln(RV_t) &= \beta_0 + \beta_1 \ln(REX^{dm}_{t-1}) + \beta_2 \ln(REX^{wm}_{t-1}) + \beta_3 \ln(REX^{mm}_{t-1}) \\ &+ \beta_4 \ln(REX^{d}_{t-1}) + \beta_5 \ln(REX^{w}_{t-1}) + \beta_6 \ln(REX^{m}_{t-1}) + u_t. \qquad (15)\end{aligned}$$

Comparing the performance of equation 15 against that of equation 14 will indicate if the signed extreme volatility (losses versus gains) contain valuable information for forecasting. This model will be denoted below as the HAR^{RE*} model.

In the subsequent forecasting exercise, longer horizon forecasts are generated by using direct forecasts based on the average of log RV over the horizon t to $t+k-1$.

3.2. Comparing forecasts

Following (Patton, 2011; Patton and Sheppard, 2009), the empirical quasi-likelihood (QLIKE) will be used to assess out-of-sample forecast accuracy and is defined as:

$$\text{QLIKE} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{RV_t}{F_t} - \log\frac{RV_t}{F_t} - 1\right), \tag{16}$$

where T is the number of forecasts and F_t denotes a forecast of RV_t.[1] Equation (16) is easily modified for weekly, or longer horizon, volatility forecasts.

Statistically significant differences in forecast performance will be assessed using the model confidence set (MCS) introduced by Hansen et al. (2011). The MCS procedure avoids the specification of a benchmark model, and starts with a collection of competing models (or approaches), $\mathcal{M}^0$, indexed by $i = 1, \ldots, m_0$. QLIKE based loss differentials $d_{ij,t}$ between models i and j are computed, and $H_0 : E(d_{ij,t}) = 0$ for all i, j (the null hypothesis of EPA) is tested. If the null hypothesis is rejected at the significance level α, the worst performing model is eliminated and the process is repeated until non-rejection occurs with the set of surviving models being the MCS, $\widehat{\mathcal{M}}^*_{1-\alpha}$. By using the same significance level for all tests, $\widehat{\mathcal{M}}^*_{1-\alpha}$ contains the best model(s) from $\mathcal{M}^0$ with a limiting $(1-\alpha)$ level of confidence.[2] Here the results are reported based on both the range and squared (SQ) statistics as described in Hansen et al. (2003).[3]

[1]Using the commonly employed empirical mean squared error (MSE) is also a possibility, however, simulation based evidence by Patton and Sheppard (2009) suggests the use of QLIKE rather than MSE due to the formers higher power in Diebold and Mariano (1995) and West (1996) type tests for equal predictive accuracy (EPA).

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

[3]MCS code from the MFE toolbox of Kevin Sheppard, `https://www.kevinsheppard.com/MFE_Toolbox`, was used.

4. Results

This section presents the in-sample estimation results for the different HAR models from Section 3.1. A comparison across the estimation results for the different models will provide a deeper understanding of the information contained in the different realized measures, including the new measures proposed here. MCS results highlighting the out-of-sample forecasting performance will follow.

4.1. In-sample estimation

Table 1 reports in-sample estimation results for each of the HAR models discussed in Section 3.1 for a $k = 1$-day horizon. To begin, the coefficient estimates for the original HAR^{RV} follow a familiar pattern. All three coefficients are significant and reducing in magnitude from β_1 through to β_3. Estimates for the HAR^{Q} model reflect a similar degree of persistence in β_1 through β_3 relative to the HAR^{RV} model. In addition, the estimate of β_{1Q} is negative and significant, which is consistent with Bollerslev et al. (2016) and captures the attenuation bias due to the estimation error in RQ. Estimation results for the HAR^{RSV} model show that the coefficients β_1 through β_3 on the RV_t^- are much stronger than β_4 through β_6 which relate to RV_t^+. This result is consistent with those from Patton and Sheppard (2015) who found that negative volatility is more important than positive volatility. These results are consistent with the differences between the ACFs presented earlier showing that RV_t^- is more persistent, and hence is found to be more important for explaining future volatility. Results for the HAR^{CJ} show that the continuous component is important with estimates of β_1 through β_3 very similar to those from the HAR^{RV} or HAR^{Q} models. Only one of the coefficient estimates, β_6, on the past jump components is significant. Here the importance of C_t relative to J_t is consistent with the ACFs presented earlier in that there is much less persistence in J_t and that in the continuous component C_t is similar to RV_t.

To begin with the results for the HAR^{RE} model (with moderate and positive and negative extreme volatility) consider the coefficient estimates for β_4 through β_6 which relate to REX_t^m. These estimates are stinkingly similar to those from the standard HAR^{RV} model implying that the dominant feature in RV_t driving the standard model is similar to REX_t^m. The large and significant β_1 and β_2 coefficient estimates from the day and week lags of REX_t^- indicate there is a very strong short-term effect from extreme negative volatility. In contrast, the effect of REX_t^+ is less pronounced with only β_7 being significant. For the final model HAR^{RE*}, the coefficient estimates for β_1 through β_3 on REX_t^m are very similar to those from the HAR^{RE} and standard HAR^{RV} models. The estimates on β_4 through β_6 relating to lags of REX_t, combined positive and negative extreme volatility, are all significant with much stronger effects evident at the 1-day lag as evident with β_4.

In summary, moderate and extreme volatility appear to play very different roles in explaining future RV. These estimation results indicate that moderate volatility may be the dominant feature in total RV within the standard HAR^{RV} model. The influence of extreme volatility is felt more at shorter lags indicating that it may capture more short-term effects in volatility, a pattern broadly consistent with the ACFs presented earlier. The question of whether this novel decomposition is of value in the context of out-of-sample forecasting will now be addressed.

4.2. Forecasting results

Table 2 presents the MCS results for three different horizons, $k = 1, 5$ and 22 days ahead, with the HAR^{RE} and HAR^{RE*} based on REX_t^-, REX_t^m and REX_t^+ being determined using $\alpha = 0.025$. This is the most extreme case considered here. Beginning at the 1−day horizon, based on a 5% critical level, all forecasts from all models with the exception of HAR^{CJ} are included in the MCS, with the HAR^Q ranking first in terms of QLIKE loss. A very similar result is also

	HAR^{RV}	HAR^{Q}	HAR^{RSV}	HAR^{CJ}	HAR^{RE}	HAR^{RE*}
β_1	0.4428 (0.0171)	0.4625 (0.0188)	0.2995 (0.0167)	0.4526 (0.0170)	0.6537 (0.0588)	0.4700 (0.0233)
β_{1Q}		−0.0026 (0.0010)				
β_2	0.3536 (0.0257)	0.3502 (0.0257)	0.3235 (0.0440)	0.3331 (0.0254)	0.7403 (0.1487)	0.3113 (0.0362)
β_3	0.1619 (0.0199)	0.1601 (0.0199)	0.1327 (0.0881)	0.1575 (0.0199)	0.1103 (0.2805)	0.1837 (0.0309)
β_4			0.1140 (0.0187)	−0.1062 (0.0884)	0.4797 (0.0231)	0.5772 (0.0367)
β_5			0.0448 (0.0438)	0.4369 (0.2163)	0.3472 (0.0357)	0.4426 (0.0576)
β_6			0.0420 (0.0859)	0.2766 (0.4124)	0.1962 (0.0300)	0.1570 (0.0493)
β_7					0.1561 (0.0551)	
β_8					−0.1269 (0.1384)	
β_9					0.1332 (0.2688)	

Table 1: Estimated HAR coefficients and associated standard errors for a 1-day forecast horizon. REX_t^-, REX_t^m and REX_t^+ are determined based on $\alpha = 0.025$.

found at the 5−day horizon. Moving to the longer 22−day horizon reveals a very different pattern. The HAR^{RE*} model ranks best and is the only member of the MCS if both test statistics are considered. The more flexible HAR^{RE} model and HAR^{RSV} are narrowly included in the MCS under the Range statistic, and are only rejected marginally under the SQ statistic.

$k = 1$			$k = 5$			$k = 22$		
	Range	SQ		Range	SQ		Range	SQ
HAR^{CJ}	0.0000	0.0000	HAR^{CJ}	0.0020	0.0010	HAR^{CJ}	0.0490	0.0080
HAR^{RE}	0.1580	0.1340	HAR^{RE}	0.0910	0.0990	HAR^{Q}	0.0490	0.0150
HAR^{RE*}	0.3450	0.3170	HAR^{RE*}	0.5220	0.5120	HAR^{RV}	0.0540	0.0180
HAR^{RSV}	0.3450	0.3420	HAR^{RV}	0.6920	0.7330	HAR^{RSV}	0.0540	0.0360
HAR^{RV}	0.3450	0.3420	HAR^{Q}	0.9240	0.9240	HAR^{RE}	0.0540	0.0360
HAR^{Q}	1.0000	1.0000	HAR^{RSV}	1.0000	1.0000	HAR^{RE*}	1.0000	1.0000

Table 2: MCS results, p-values are reported. REX_t^-, REX_t^m and REX_t^+ are determined based on $\alpha = 0.025$.

It is clear that a choice regarding the value of α needs to be made when constructing REX_t^-, REX_t^m and REX_t^+. To check the robustness of the forecasting results to the choice of α, Tables 3 and 4 present MCS results for the cases when REX_t^-, REX_t^m and REX_t^+ are based on $\alpha = 0.05$ and $\alpha = 0.1$ respectively. Overall, these two sets of results are consistent with those in Table 2. The HAR^{RE} and HAR^{RE*} models continue to produce the best ranked forecasts at the longer horizon. In fact, based on $\alpha = 0.1$, HAR^{RE} and HAR^{RE*} produce forecasts that are statistically superior to all other models. This result indicates using somewhat less extreme, extreme volatility (and moderate volatility being correspondingly wider) may be beneficial. The optimal choice of α however is an interesting avenue for future research. Here the fact that the HAR^{RE} and HAR^{RE*} forecasts are consistently the most accurate forecasts, across the full range of forecast horizons, indicates the flexibility of the decomposition of total RV into shorter and longer-term components is beneficial.

Results in the online appendix for other indices show very similar results for the Nasdaq and DAX indices, though not quite as strong for the FTSE. Experiments were also undertaken with a number five large U.S. individual stocks. While HAR^{RE} was often found to be the most accurate forecast, the forecast accuracy of HAR^{RE} (and HAR^{RE*}) were not found to be statistically superior to the competing models such as HAR^{Q}. The differences stronger performance across in the context of indices relative to individual stocks seems to indicate that there are important differences in the dynamics between the two. The most likely reason is that the more persistent moderate volatility component is only important in cases where there is an strong persistent common factor, present in a well diversified index.

	$k=1$			$k=5$			$k=22$	
	Range	SQ		Range	SQ		Range	SQ
HAR^{CJ}	0.0000	0.0000	HAR^{CJ}	0.0020	0.0040	HAR^{CJ}	0.0930	0.0220
HAR^{RE}	0.3780	0.3100	HAR^{RE}	0.1130	0.1430	HAR^{RV}	0.0960	0.0490
HAR^{RE*}	0.4510	0.3940	HAR^{RE*}	0.4610	0.3930	HAR^{Q}	0.0960	0.0700
HAR^{RSV}	0.4510	0.3940	HAR^{RV}	0.6600	0.7100	HAR^{RSV}	0.0960	0.0900
HAR^{RV}	0.4510	0.3940	HAR^{Q}	0.9170	0.9170	HAR^{RE}	0.0960	0.0900
HAR^{Q}	1.0000	1.0000	HAR^{RSV}	1.0000	1.0000	HAR^{RE*}	1.0000	1.0000

Table 3: MCS results, p-values are reported. REX_t^-, REX_t^m and REX_t^+ are determined based on $\alpha = 0.05$.

	$k=1$			$k=5$			$k=22$	
	Range	SQ		Range	SQ		Range	SQ
HAR^{CJ}	0.0000	0.0000	HAR^{CJ}	0.0020	0.0030	HAR^{CJ}	0.0300	0.0080
HAR^{RSV}	0.3180	0.3880	HAR^{RV}	0.3750	0.3700	HAR^{Q}	0.0300	0.0140
HAR^{RE}	0.3180	0.3880	HAR^{RSV}	0.3750	0.3700	HAR^{RV}	0.0300	0.0140
HAR^{RV}	0.3590	0.4780	HAR^{Q}	0.3750	0.3700	HAR^{RSV}	0.0300	0.0140
HAR^{Q}	0.7890	0.7890	HAR^{RE}	0.3750	0.3700	HAR^{RE}	0.5360	0.5360
HAR^{RE*}	1.0000	1.0000	HAR^{RE*}	1.0000	1.0000	HAR^{RE*}	1.0000	1.0000

Table 4: MCS results, p-values are reported. REX_t^-, REX_t^m and REX_t^+ are determined based on $\alpha = 0.1$.

5. Conclusion

This paper proposed a novel decomposition of realized volatility, into moderate and extreme volatility on the basis of the magnitude of intraday returns. This differs from existing approaches that are either based on formal statistical tests to differentiate between the continuous and jump components of volatility, or the sign of intraday returns. A simple inspection of the autocorrelations of the realized measures showed that the proposed moderate and extreme components behave quite differently to the continuous and jump components. Moderate volatility was found to be more persistent than total volatility and behaved like a long-term component of volatility, while extreme volatility reflected short-term persistence in volatility. Within the HAR forecasting framework, relative to models containing existing realized measures, models based on moderate and extreme volatility often produced forecasts of equal quality at short horizons, and forecasts of superior quality at longer horizons. Given the forecasts were consistently among the most accurate forecasts, across the full range of forecast horizons, the flexibility of the proposed decomposition into moderate and extreme (short- and long-term components) volatility seems beneficial. While the forecasting results were robust across a range of thresholds for defining extreme volatility, a method for formally choosing this threshold (or a more general structure), is a potentially interesting agenda for future research.

Amado, C., Teräsvirta, T., 2017. Specification and testing of multiplicative time-varying garch models with applications. Econometric Reviews 36 (4), 421–446.

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

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

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

Audrino, F., Huang, C., Ostap, O., 2018. Flexible har model for realized volatility. Studies in Nonlinear Dynamics and Econometrics.

Barndorff-Nielsen, O., Kinnebrock, S., Shephard, N., 2010. Measuring downside risk - Realised semivariance in Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press.

Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of Royal Statisitcal Society, Series B 64, 253–280.

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

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

Buccheri, G., Corsi, F., 2017. HARK the SHARK: Realized volatility modelling with measurement errors and nonlinear dependencies. SSRN Electronic Journal.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility.

Journal of Financial Econometrics 7 (2), 174–196.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economics Statistics 13, 253–263.

Fengler, M. R., Mammen, E., Vogt, G., 2015. Specification and structural break tests for additive models with applications to realized variance data. Journal of Econometrics 7 (2), 174–196.

Hansen, P., Lunde, A., Nason, J., 2003. Choosing the best volatility models: the model confidence set approach. Oxford Bulletin of Economics and Statistics 65, 839–861.

Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. Econometrica 79, 453–497.

Huang, X., Tauchen, G., 2005. The realtive contribution of jumps to total price variation. Journal of Financial Econometrics 3, 456–499.

Muller, U., Dacorogna, M., Dave, R., Olsen, R., Pictet, O., Weizsacker, J., 1997. Volatilities of different time resolutions – analysing the dynamics of market components. Journal of Empirical Finance 4, 213–239.

Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A., Sheppard, K., 2009. Evaluating volatility forecasts. In: Andersen, T. G., Davis, R. A., Kreiss, J. P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, Berlin.

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

West, K., 1996. Asymptotic inference about predictive ability. Econometrica 64, 1067–1084.