The Contribution of Jump Activity and Sign to Forecasting Stock Price Volatility

Ruijun Bu, Rodrigo Hizmeri, Marwan Izzeldin, Anthony Murphy and Mike G. Tsionas

Working Paper 1902 **April 2019 (revised November 2020)**

Research Department

https://doi.org/10.24149/wp1902r1

The Contribution of Jump Activity and Sign to Forecasting Stock Price Volatility[*]

Ruijun Bu[°], Rodrigo Hizmeri[†], Marwan Izzeldin[‡], Anthony Murphy[§] and Mike G. Tsionas[±]

Original Version: March 2019
Revised: August 2020

Abstract

This paper proposes a novel approach to decompose realized jump measures by type of activity (finite/infinite) and by sign. We also provide noise-robust versions of the ABD jump test (Andersen et al. 2007) and realized semivariance measures for use at high frequency sampling intervals. The volatility forecasting exercise involves the use of different types of jumps, forecast horizons, sampling frequencies, calendar and transaction time-based sampling schemes, as well as standard and noise-robust volatility measures. We find that infinite (finite) jumps improve the forecasts at shorter (longer) horizons; but the contribution of signed jumps is limited. Noise-robust estimators, that identify jumps in the presence of microstructure noise, deliver substantial forecast improvements at higher sampling frequencies. However, standard volatility measures at the 300-second frequency generate the smallest MSPEs. Since no single model dominates across sampling frequency and forecast horizon, we show that model averaged volatility forecasts - using time-varying weights and models from the model confidence set - generally outperform forecasts from both the benchmark and single best extended HAR model.

Keywords: Volatility Forecasts, Realized Volatility, Finite Activity Jumps, Infinite Activity Jumps, Signed Jumps, Noise-Robust Realized Volatility, Model Averaging.

JEL Classification: C22, C51, C53, C58.

[*] We would like to thank Andrew Patton, Peter R. Hansen, Ingmar and Sandra Nolte, Jia Li, George Tauchen, Nour Meddahi, Roberto Renò, Gerald Steele, participants at the 2019 Africa Meeting of the Econometric Society in Rabat, the 6th Annual Conference of the International Association for Applied Econometrics in Nicosia, the 2019 Asian Meeting of the Econometric Society in Xiamen, 2019 Infiniti Conference in Glasgow, 12th International Conference in Computational Statistics and Financial Econometrics in Pisa, the 2018 QuantFest at the Federal Reserve Bank of Chicago, the 11th International Conference in Computational Statistics and Financial Econometrics in London, and seminar participants at the University of Warwick for helpful comments. The views expressed are those of the authors and do not represent the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. The first author acknowledges the financial support of the Economic and Social Research Council (ESRC), grant number ES/P000665/1.

[°] Ruijun Bu, University of Liverpool.

[†] Rodrigo Hizmeri, Lancaster University.

[‡] Marwan Izzeldin, Lancaster University.

[§] Corresponding author: Anthony Murphy, Federal Reserve Bank of Dallas, 2200 N. Pearl St., Dallas, Texas 75201, U.S.A., anthony.murphy@dal.frb.org.

[±] Mike G. Tsionas, Lancaster University.

1. Introduction

Modelling and forecasting asset return volatility is central to asset pricing, portfolio optimization and risk management. The introduction and use of high frequency data provides a framework for directly measuring and capturing the main stylized facts of volatility. Realized volatility (RV), a non-parametric measure calculated as the sum of intra-day squared returns, provides a consistent estimator of quadratic variation when the price process contains discontinuities or jumps[2]

In relation to volatility forecasting, the seminal work of Andersen et al. (2007) suggests that the jump component is both highly important and distinctly less persistent than the continuous component. Thus, treating rough jumps separately results in significant improvements in out-of-sample volatility forecasts, not least because many significant jumps are associated with specific macroeconomic news announcements. However, recent empirical evidence that classifies jumps into finite and infinite activity jumps (Aït-Sahalia and Jacod, 2012), presents a new question as to whether such different types of jumps are equally important in the prediction of future volatility.[3]

A large literature examines the role of jumps in volatility forecasting. However, much of that literature focuses on signed jumps, and does not separate finite jumps from infinite jumps. It also tends to use 300-second returns, rather than higher frequency 5- or 60-second returns, in order to mitigate the impact of the market microstructure noise. Whether for jumps or signed jumps, the literature provides mixed evidence regarding their value added in forecasting. One side of the literature reports gains in forecasting from incorporating jumps. Andersen et al. (2007) find that separating jumps from the continuous volatility component improves out-of-sample forecasts. Corsi et al. (2010) show that the use of a threshold bipower estimator to calculate the jump component improves the out-of-sample forecasts. Patton and Sheppard (2015) argue that volatility is strongly related to the volatility of past negative returns, and show that models incorporating signed jumps lead to significantly better out-of-sample forecast performance. Duong and Swanson (2015) identify large and small jumps using higher order power variations, and find that small jumps are more important for forecasting volatility than large jumps.

---

[2] Early adoption of RV for modelling and forecasting volatility featured in the work of Andersen and Bollerslev (1998) and Andersen et al. (2001, 2003 and 2005) inter alios.

[3] Other research considering the role of infinite jumps can be found in Aït-Sahalia (2004), Huang and Tauchen (2005), Lee and Mykland (2007), Aït-Sahalia and Jacod (2009), Dumitru and Urga (2012) and the extensive references therein. For infinite jumps, see Aït-Sahalia and Jacod (2009, 2014), and Todorov and Tauchen (2010) inter alios.

Another side of the literature finds that jumps do not significantly improve volatility forecasts. For instance, Forsberg and Ghysels (2007), Giot and Laurent (2007), Martens et al. (2009), Busch et al (2011), Sévi (2014) and Prokopczuk et al. (2016) review the use of jumps and signed jumps to forecast future volatility. Their results suggest that the inclusion of jumps and signed jumps improves the in-sample fit of models, but generate no significant out-of-sample forecasting gains.

The current paper contributes to the literature in a number of ways. First, we show how jumps may be decomposed into signed, finite and infinite activity jumps. We identify the finite and infinite jump components using the intersection of the ABD intraday jump test and the SFA finite activity jump test (Andersen et al. (2007b), Aït-Sahalia and Jacod (2011)). Duong and Swanson (2015) use higher order power variations to decompose jumps into large and small jumps, and examine their role in predicting the volatility of returns. By contrast, we use a more robust test based decomposition of days with significant jumps into ones with finite or infinite activity jumps. As noted by Aït-Sahalia and Jacod (2014), estimated jumps based on higher-order power variation are often poor measures of actual jumps. Second, we develop versions of the ABD test and realized semivariances measures that are robust to microstructure noise, and perform well at high frequency. The noise robust realized semivariance measures are modifications of the two-scale realized volatility measure of Zhang et al. (2005). Third, we present new empirical evidence showing the contribution of the various types of signed, finite and infinite activity, jumps to improving volatility forecasts at different forecast horizons. We examine the choice of sampling frequency and sampling scheme, as well as the use of noise-robust realized measures. Volatility forecasts using transaction-time based measures are dominated by those using regular clock-time based measures. Fourth, as most jumps are idiosyncratic, no single forecasting model dominates, so better forecasts are obtained using simple model averages using 300-second jump measures.

Our application uses high-frequency data from 2000 to 2016. Using extended HAR models, we forecast the volatility of SPY, the SPDR S&P 500 ETF, as well as 20 constituents of the S&P 100 index which vary by sector and volume. We show that jumps contribute significantly to the volatility of SPY and the 20 stocks we examine. As expected, we find the SPY volatility forecasts to be more accurate, since aggregation helps to identifies more informative jumps which improves the out-of-sample mean square prediction error (MSPE) performance.

To preview our findings, when jumps classified by sign and activity are used as additional predictors in HAR models, we find significant improvements in both in- and out-of-sample performance. We focus on the MSPE results from pseudo, out-of-sample forecasts using rolling window regressions. In terms of our classification of jumps by activity, infinite jumps are relatively more important at shorter horizons, whereas finite jumps dominate at longer horizons. Adding signed finite and infinite jumps to the forecasting model often generates significantly better forecasts than the standard HAR-RV model. However, no single extended model dominates.

The use of noise-robust estimators substantially improves the out-of-sample performance of our extended HAR models, especially at higher frequencies. The gains are greater for individual stocks than for the SPY index. This is unsurprising since SPY is a very liquid asset with a low level of microstructure noise. One might have expected standard volatility measures to deliver forecasts that are more accurate at the 300-second frequency, since microstructure noise should be small. However, this only holds true for SPY. For individual stocks, the forecasting performance of noise-robust and standard volatility measures is similar. In line with Ghysels and Sinko (2011), noise-robust measures only improve forecasting performance when the level of market microstructure noise is significant.

The greatest gains in real-time forecasting performance are generally found using returns sampled at 300-second intervals, rather than at 5- or 60-second intervals, irrespective of whether noise-robust or standard volatility measures are used.[4] Since the forecasting performance of no single model dominates across sampling frequency and forecast horizon, we investigate model averaging using the model confidence set approach of Hansen et al. (2011) to reduce the set of retained models in the averages. Simple model averaging, including averages using time-varying weights, generally results in significant out-of-sample forecasting performance (e.g. Aiolfi and Timmermann (2006), Timmermann (2006), Aiolfi et al. (2011), Elliott and Timmermann (2016)). These gains arise using both the SPY and individual stocks across different horizons. The gains are greatest using the returns sampled every 300-seconds. We assess the predictive accuracy of model averaging using the pairwise test of Diebold and Mariano (1995). The results show that model averaging produces significantly smaller MSPEs, even at long horizons of 66 days / 3 months.

These results are in line with those of Giacomini and Rossi (2010), where the relative forecasting performance of individual models often changes over time. Here we identify the incidence of cojumps in our data using the co-exceedance rule of Gilder et al. (2014). The cojumps results indicate that the jumps in our data are mainly idiosyncratic, reflecting stock specific differences in the arrival of news and

[4] This result is in line with Liu et al. (2015) who find that 300-second / 5-min RV is very difficult to beat. Across a range of assets in different classes, they found that 5-minute returns volatilities obtained from the two scales realized volatility (TSRV) subsampling approach of Zhang et al. (2005) is the preferred method of estimating daily volatility.

the reaction to that news.[5] The fact that the timing, size and sign of most jumps are stock specific is the main reason why no single forecast model dominates.

As a robustness check, we consider alternative, transaction-time sampled volatility measures. To the best of our knowledge, only Patton and Sheppard (2015) have considered an alternative sampling scheme for forecasting and their focus is on signed jumps. They do not examine the role of finite and infinite jumps, nor do they compare their results with those using the popular clock-time sampling scheme. In the case of SPY, we find that the share of jumps in transaction-time based RV measures is far smaller than for clock-based measures, and any jumps are predominantly finite activity jumps. In terms of forecasting performance, we conclude that forecasts using volatility and jump measures based on transaction sampling are inferior to the forecasts from clock-based sampling.

The remainder of the paper is as follows. The theoretical background is set out in Section 2. The estimation of signed, finite and infinite activity jumps is described in Section 3. Robust-to-noise volatility measures are also discussed. Section 4 sets out the forecasting framework, including the extended HAR forecasting models and forecast evaluation criteria. The data used in this study are described in Section 5, where the incidence of various types of jumps is tabulated. The forecasting gains from adding different types of signed, finite and infinite activity jumps to HAR models are documented in Section 6. Model averaging results are presented in Section 6. Volatility forecasting results using transaction-time sampled volatility measures are discussed in Section 8. Finally, Section 9 summarizes the paper and presents our conclusion.

2. Theoretical Background

Let X_t denote the log of the price of an equity or an equity index. We assume X is an Itô-semimartingale process defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with the following representation

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t, \quad t \in [0, T]$$

where a_t is a locally bounded and predictable drift term and σ_t is the adapted, càdlàg spot volatility, W_t is a standard Brownian motion and J_t is a pure jump process with finite and infinite activity components, $J_t = J_t^F + J_t^I$. The finite activity J_t^F and infinite activity J_t^I jump processes are:

[5] Similar qualitative conclusions are obtained using the multijump test of Caporin et al. (2017). The number of detected cojumps is also similar to the numbers reported in Caporin et al. (2017) and Mukherjee et al. (2020).

$J_t^F = \int_0^t \int_{|x|>\epsilon} x\mu(dx, ds),$

$J_t^I = \int_0^t \int_{|x|\le\epsilon} x(\mu(dx, ds) - \nu(dx)ds),$

where μ is the jump measure of X with compensator ν, and $\epsilon > 0$ is an arbitrary number. For more details on Itô semi-martingale processes, see Aït-Sahalia and Jacod (2014) and the references therein. As Aït-Sahalia and Jacod (2012) note, the continuous part of the X process captures the normal risk of an asset that can be hedged using standard methods. The large, finite jumps capture default risk or big news-related events, while the small jumps capture price moves that impact high frequency prices but wash out at the daily level, e.g. the price impact of large transactions.

Since volatility is a latent variable, realized measures are widely employed to give consistent estimates of the quadratic variation (QV) of the process using high-frequency data. The quadratic variation of the price process is:

$$QV_t = \underbrace{\int_0^t \sigma_s^2 ds}_{\substack{\text{Integrated} \\ \text{Variance (IV)}}} + \underbrace{\sum_{0<s<t} (\Delta_s X)^2}_{\substack{\text{Jump} \\ \text{Contribution}}}$$

where $\Delta_s X = X_s - X_{s-}$ when X jumps at time s. Under suitable conditions, the widely used realized volatility (RV) measure converges in probability to the quadratic variation as the sampling interval $\Delta_n \to 0$:

$$RV_t = \sum_{i=1}^{M} (\Delta_i^n X)^2 \xrightarrow{p} QV_t,$$

where the day is split into $n = \lfloor 1/\Delta_n \rfloor$ equally spaced intervals of length Δ_n, $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ is the log return in interval i, and $\lfloor x \rfloor$ denotes the integer part of x.

To separate the integrated variation component of QV from the jump component, we use the threshold bipower variation ($TBPV$) measure proposed by Corsi et al. (2010), a modified version of the so-called bipower variation measure of Barndorff-Nielsen and Shephard (2004). The $TBPV$, which is robust to jumps in both the stochastic limit and the asymptotic distribution, converges in probability to the integrated variance as the sampling interval $\Delta_n \to 0$:

$$TBPV_t = \mu_1^{-2} \frac{n}{n-1} \sum_{i=2}^{n} |\Delta_i^n X| 1_{\{(\Delta_i^n X)^2 \le \vartheta_i\}} |\Delta_{i-1}^n X| 1_{\{(\Delta_{i-1}^n X)^2 \le \vartheta_{i-1}\}} \xrightarrow{p} \int_0^t \sigma_s^2 ds,$$

where $\mu_1 = \sqrt{2/\pi \approx 0.7979}$, $n/(n-1)$ is a small sample correction and ϑ is the threshold estimator in Corsi et al. (2010, Appendix B). We also utilize the positive and negative realized semivariance (RS) estimators of Barndorff-Nielsen et al. (2010) to capture upside and downside risk:

$$RS_t^+ = \sum_{i=1}^{n} (\Delta_i^n X)^2 1_{\{\Delta_i^n X>0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\le t} (\Delta X_s)^2 1_{\{\Delta X_s>0\}},$$

$$RS_t^- = \sum_{i=1}^{n} (\Delta_i^n X)^2 1_{\{\Delta_i^n X<0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\le t} (\Delta X_s)^2 1_{\{\Delta X_s<0\}}.$$

3. Identifying and Decomposing Jumps by Sign and Activity

To identify days with significant jumps, we employ the intra-day jump test proposed by Andersen et al. (ABD, 2007b). If the largest intra-daily value of the test exceeds the critical value, we classify the day as a jump day. The $\mathcal{J}_t$ indicator for a day with significant jumps is one if $\max_i |\Delta_i^n X| > \Phi_{1-\beta/2}^{-1}\sqrt{\Delta_n \times TBPV_t}$ and 0 otherwise, where $\Phi_{(\cdot)}^{-1}$ is the inverse of the standard normal cumulative distribution function, α is the significance level and $\beta = 1-(1-\alpha)^{\Delta_n}$ is the Šidàk multiple testing correction. Hence, the estimated continuous and jump and continuous components of QV are:

$$\hat{C}_t = RV_t \times (1-\mathcal{J}_t) + TBPV_t \times \mathcal{J}_t\,,$$

$$\hat{J}_t = \max(RV_t - TBPV_t, 0) \times \mathcal{J}_t\,.$$

To identify days with significant finite or infinite activity jumps, we employ the SFA test proposed by Aït-Sahalia and Jacod (2011). The test statistic is the ratio of two truncated realized power variation measures designed to eliminate large jumps. The realized power variation $B(p, \upsilon_n, \Delta_n) = \sum_{i=1}^{n} |\Delta_i^n X|^p 1_{\{|\Delta_i^n X| \le \upsilon_n\}}$, with $\upsilon_n = \varrho\Delta_n^{\varpi}, \varrho > 0$ and $\varpi \in (0, \frac{1}{2})$, is the sum of truncated absolute returns, $|\Delta_i^n X| \le \upsilon_n$, raised to the power p over different sampling frequencies Δ_n. The SFA test statistics has different limits depending on whether the jumps in X_t are finite or infinite activity jumps. $SFA_t = \frac{B(p,\upsilon_n,k\Delta_n)_t}{B(p,\upsilon_n,\Delta_n)_t} \xrightarrow{p} k^{p/2-1}$ in the finite activity case and 1 in the infinite activity case. Under the finite activity null, the statistic $\left(SFA_t - k^{\frac{p}{2}-1}\right)/\sqrt{\hat{V}_t} \xrightarrow{L} \mathcal{N}(0,1)$, where $\hat{V}_t = N(p,k)\frac{B(2p,\upsilon_n,\Delta_n)_t}{B(p,\upsilon_n,\Delta_n)_t^2}$. For further details of $N(p,k)$, and other settings, see Ait-Sahalia and Jacod (2011). We set $k = 2$ and $p = 4$, and use the indicator $\mathcal{F}_t = 1(SFA_t < k^{\frac{p}{2}-1} - \Phi_{1-\alpha}^{-1}\sqrt{\hat{V}_t})$ to identify days with finite activity jumps.

- Table 1 Here -

Our classification of jumps by sign and activity is set out in **Table 1**. We classify jumps by activity using the jump $\mathcal{J}_t$ and finite activity $\mathcal{F}_t$ indicators. The contributions of positive and negative jumps to overall QV are based on $\max\,(RS_t^+ - \frac{1}{2}TBPV_t, 0) \times \mathcal{J}_t$ and $\max\left(RS_t^- - \frac{1}{2}TBPV_t, 0\right) \times \mathcal{J}_t$ respectively. When forecasting volatility using our extended HAR models, we use the daily (net) signed jump, $\widehat{SJ}_t$, the difference between the positive and negative measures (Patton and Sheppard, 2015). The corresponding positive and negative signed jumps are $\hat{J}_t^+ = \widehat{SJ}_t \times \mathcal{P}_t$ and $\hat{J}_t^- = \widehat{SJ}_t \times (1 - \mathcal{P}_t)$ respectively, where $\mathcal{P}_t = 1\left(\widehat{SJ}_t > 0\right)$. Their finite / infinite activity counterparts are identified using the finite activity $\mathcal{F}_t$ indicator.

3.1 Market Microstructure Noise

Market microstructure noise can distort realized volatility measures, and hence the identification of jumps. We know that the contribution of jumps varies by sampling frequency (**Table 3** below), and that the level of market microstructure noise increases as the sampling interval $\Delta_n \to 0$. As a result, standard high frequency realized volatility measures tend to be biased, distorting jump test statistics (e.g. Hansen and Lunde (2006), Huang and Tauchen (2005). This suggests that robust-to-noise volatility measures should be used at high frequencies (e.g., 5 and 60 seconds), and possibly lower frequencies. Although Aït-Sahalia and Xiu (2019) suggest that improvements in stock market liquidity mean that the common practice of treating the 5-minute returns of S&P 100 constituents as noise-free is a reasonably safe choice for data sampled after 2009, it is problematic before then. They also suggest that the 5-minutes returns of a large portion of S&P 500 index constituents cannot be treated as noise-free.

We assume that the observed log price process, Y_t, is contaminated by additive, microstructure noise[6]:

$$Y_t = X_t + u_t,$$

where X_t is the underlying log price process described above, and u_t is an i.i.d. noise process with mean zero and variance ω^2, stochastically independent of X_t. Jacod et al. (2009) and Christensen et al.

[6] The mechanics of trading generate a diverse array of market microstructure effects including the bid-ask spread and corresponding bounce, the gradual response of prices to a block trade, the strategic component of order flow inventory control effects (Aït-Sahalia and Jacod, 2014). Additive noise is the simplest market microstructure model.

(2014) propose pre-averaging the returns to remove most of the noise. We estimate the pre-averaging returns as a weighted average of returns within a local neighborhood of L log-prices:

$$\Delta_i^n \breve{X} = \sum_{j=1}^{L-1} g\left(\frac{j}{L}\right) \Delta_{i+j}^n Y,$$

where $g(x) = \min(x, 1-x)$, $\Delta_i^N Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$, $L = \theta\sqrt{M}$, and $\theta = \frac{1}{3}$ for 5 and 60 second returns or $\theta = 1$ for 300 second returns. The robust-to-noise estimators for the realized variance and the bipower variation are:

$$RV_t^* = \frac{n}{n-L+2} \frac{1}{L\psi_2^L} \sum_{i=0}^{n-L+1} |\Delta_i^n \breve{X}|^2 - \frac{1}{\theta^2} \frac{\psi_1^L}{\psi_2^L} \widehat{\omega}^2,$$

$$BPV_t^* = \frac{n}{n-2L+2} \frac{1}{L\psi_2^L \mu_1^2} \sum_{i=0}^{n-2L+1} |\Delta_i^n \breve{X}||\Delta_{i+L}^n \breve{X}| - \frac{1}{\theta^2} \frac{\psi_1^L}{\psi_2^L} \widehat{\omega}^2 \quad ,$$

where the leading terms are small sample corrections, the trailing term $\frac{1}{\theta^2}\frac{\psi_1^L}{\psi_2^L}\widehat{\omega}^2$ is a bias-correction to remove the residual noise that is not eliminated by pre-averaging, and $\psi_1^L = L\sum_{j=1}^{L} \left(g\left(\frac{j}{L}\right) - g\left(\frac{j-1}{L}\right)\right)^2$ and $\psi_2^L = \frac{1}{L}\sum_{j=1}^{L-1} g^2\left(\frac{j}{L}\right)$ are constants (Jacod et al., 2009, Appendix A, and Christensen et al., 2014). The unknown noise variance ω^2 can be approximated using either the Bandi and Russell (2006) estimator $\widehat{\omega}_{RV}^2 = \frac{1}{2}\sum_{i=1}^{M} (\Delta_i^n Y)^2$, or the Oomen (2006a) estimator $\widehat{\omega}_{AC}^2 = -\frac{1}{M-1}\sum_{i=2}^{M} \Delta_{i-1}^n Y \Delta_i^n Y$, the negative of the first order autocovariance of log returns. We use the latter estimator.

The ABD jump test in Andersen et al. (2007b) can be modified to yield a test that is robust to the presence of additive market microstructure noise. This is done by replacing raw returns by pre-averaged returns, and using the pre-averaged bipower variation. Days with significant jumps may be identified using this version of the ABD test. Thus, our noise robust indicator of significant jumps, $\mathcal{J}_t^* = 1$ when $\max_i |\Delta_i^n \breve{X}| > \Phi_{1-\beta/2}^{-1}\sqrt{\Delta_n \times BPV_t^*}$ and 0 otherwise.

Noise-robust versions of the realized semivariances, which capture upside and downside risk, are constructed by appropriately modifying the two-time scale realized variance measure of Zhang et al. (2005):

$$TSRS_t^+ = \frac{1}{K}\sum_k RS_t^{(k)+} - \frac{\overline{n}}{n} RS_t^+ \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta X_s)^2 1_{\{\Delta X_s > 0\}},$$

$$TSRV_t^- = \frac{1}{K}\sum_k RS_t^{(k)-} - \frac{\overline{n}}{n} RS_t^- \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta X_s)^2 1_{\{\Delta X_s < 0\}},$$

where $RS_t^{(k)+}$ and $RS_t^{(k)-}$ are subsample, slower time scale, realized semivariance measures; RS_t^{+} and RS_t^{+} are the full sample, faster time scale, realized semivariance measures; $\overline{n} = \frac{n-K+1}{K}$ is the average number of observations in the K subsamples; $K = \lfloor cn^{2/3} \rfloor$ and c is the optimal bandwidth in Zhang et al. (2005). The two-time scale estimators average the realized semivariances over K subsamples, and apply a bias correction from the full sample (the second time scale).[7]

3.2 Noise-Robust ABD Test and Two-Time Scale realized Semivariance Measures - Monte Carlo Results

We examine the performance of our noise-robust ABD test statistic and two-time scale realized semivariance estimators using Monte Carlo simulations. The simulations use a (log-) price process X based on the Heston (1993) model augmented with finite or infinite activity jumps:

$$\begin{aligned} dX_t &= \sqrt{\nu_t} dW_t + \theta_L dL_t, \\ d\nu_t &= \kappa(\eta_\nu - \nu_t)dt + \gamma_\nu \nu_t^{1/2} dB_t, \end{aligned}$$

where W_t and B_t are Wiener process with covariance $\mathbb{E}[dW_t, dB_t] = \rho dt$, and L_t is either a finite activity compound Poisson process or an infinite activity Cauchy process (a β-stable process with $\beta = 1$). Following Aït-Sahalia and Jacod (2011), we set $\kappa = 5, \sqrt{\eta_\nu} = 0.25, \gamma_v = 0.5$ and $\rho = -0.5$. The compound Poisson process has intensity λ, and jumps that are uniformly distributed on $\nu_t^{1/2}\sqrt{m}([-2,-1] \cup [1,2])$. We set $m = 0.7$ and $\lambda = 0.5$ such that, one average, there is one jump every two days. In the finite activity case, we set $\theta_L = 0.5$, while for infinite jumps we set $\theta_L = 0.5$. Following Barndorff-Nielsen et al. (2008), we add noise to the $X_{t,i}$ process:

$$Y_{t,i} = X_{t,i} + u_{t,i},$$

where Y is the noisy, observed log price process, ξ is the noise-to-signal ratio used to simulate market micro-structure noise, $u_{t,i} \sim \mathcal{N}(0, \omega_t^2)$ and $\omega_t^2 = \xi^2 \int_0^t \nu_s ds$. With his design, the variance of the noise is constant throughout the day, but changing from day to day. We simulate the Y process second-by-second for 50 days, with 6.5 hours of trading per day. We replicate the simulations 3,000 times

- **Table 2 Here** -

[7] Aït-Sahalia, Jacod and Xiu (2019) develop a noise-robust, pre-averaging, version of the Aït-Sahalia and Jacod (2009) jump test, while Li and Xia (2016) develop general GMM procedures that address measurement error in realized volatility measures.

Table 2 shows the results of our Monte Carlo exercise exploring the size and power of the two versions of the ABD test under finite and infinite jumps, with a moderate and higher level of noise-to-signal ratio. The tests are evaluated at the 5% level. The noise-robust ABD test is more powerful at higher, 5-second and 60-second, frequencies and when the noise-to-signal ratio is higher. At lower frequencies, the size and power of the standard ABD test are better. In the Cauchy (infinite jump) case, the power of the standard ABD test is badly affected by the noise-to-signal levels.

- **Table 3 Here** –

Table 3 compares the finite sample MSEs of the realized semivariance and two time scale realized semivariance measures. The results show that the realized semivariance is very sensitive to market micro-structure noise, resulting in large MSEs even when the noise-to-signal ratio is moderate and the sampling frequency is low. On the other hand, the performance of the two-scale realized semivariance is very good overall.

4. Forecasting Models and Forecast Comparisons

The basic HAR-RV in Corsi (2009) models current and future RV as a linear function of lagged daily, weekly and monthly values of RV. Andersen et al. (2007a) originally added jumps to the HAR-RV model. Our forecasting models extends the HAR-RV model further by adding signed, finite and infinite activity jumps. The baseline HAR-RV model is:

$$RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \varepsilon_{t,t+h-1}$$

where t refers to time in days, h is the forecast horizon, and $RV_{t,t+h-1} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+1-i}$ etc. are period averages of daily RV so the coefficients are on the same scale. We examine nine different, extended HAR models (**Table 4)**. The first three forecasting models include daily, weekly and monthly jumps in addition to the daily, weekly and monthly continuous components of RV. The next three models replace the jump variables in previous models with their finite activity counterparts. The final three models only use infinite activity jumps. We estimate separate models for unsigned, positive and negative jumps.

- **Table 4 Here** -

The realized continuous and jump measures in the models are estimated using the formulae in **Table 1**. We also have an additional nine models where all the right-hand volatility measures are the noise-robust measures discussed in Section 3 above. Although additional variants of these models could be developed and evaluated, we do not believe that it is worthwhile doing so since the model averages should encompass these variants.

Our primary interest is in the performance of pseudo out-of-sample forecasts. We consider forecast horizons h = 1, 5, 22 and 66, corresponding to one day, one week, one month, and one quarter ahead. We use rolling window regressions of size 1000, or approximately four years, to estimate the models. The out-of-sample forecast performance is evaluated using the mean squared prediction error (MSPE) loss function and, to a lesser extent, the out-of-sample R squared (R^2_{oos}). The MSPE, which has been shown to be robust to noise in the proxy for volatility in Patton (2011), is:

$$MSPE^h = \frac{1}{S_h} \sum_{s=1}^{S_h} (RV_s^h - \widehat{RV}_s^h)^2$$

where RV_s^h and RV_s^h are the actual and pseudo out-of-sample forecasts of $RV_{t,t+h-1}$, and S_h is the total number of out-of-sample forecasts from the series of rolling window models. Additionally, we carry out pairwise tests of the null of equal predictive ability using Diebold and Mariano (1995) tests with a MSPE loss criterion and HAC standard errors.

The model confidence set (MCS) procedure of Hansen et al. (2011) is used to identify the subset of models with significantly lower MSPEs than the other models. We use the MCS procedure with a quadratic loss function. The MCS test statistic is $\mathcal{T}_{\mathcal{M}} = max_{i,j\in\mathcal{M}}\left|t^h_{i.j}\right|$, where $\mathcal{M}$ denotes the set of extended HAR models as well as the baseline HAR-RV model, $t^h_{i,j} = \bar{d}^h_{i,j}/\sqrt{\widehat{var}(\bar{d}^h_{i,j})}$ and $\bar{d}^h_{i,j} = MSPE^h_i - MSPE^h_j$ is the difference in the mean squared prediction errors of models i and j. The null hypothesis is that all models have the same expected loss, while the alternative hypotheses is that there is some model i with a MSPE that is greater than the MSPE's of all the other models $j \in \mathcal{M}\backslash i$. When the null is rejected, the worst performing model is eliminated, and this process is iterated until no further model can be eliminated. The surviving models denoted by $\mathcal{M}_{MCS}$ are retained with a confidence level $\alpha = 0.05$. We implement the MCS via a block bootstrap using a block length of 10 days and 5,000 bootstrap replications.[8]

5. Data

For our forecasting exercise, we use return data for the SPDR S&P 500 ETF (SPY) and 20 individual stocks in the S&P 500 index. The data are for the years 2000 to 2016, a total of 4,277 trading days. The 20 stocks were chosen based on their jump activity, and the relative contributions of finite and infinite jumps. The data are sourced from the TickData database.[9] We follow Hansen and Lunde (2006) and use previous tick interpolations to aggregate the ticks to the required sampling frequency.

Mean daily RV ranges from 1.037 to 8.284, while the average number of shares traded per day ranges from 0.875 to 98.972 million. Since we are interested in the role of realized measures using different sample frequencies in forecasting realized volatility, we sample returns every 5, 60, and 300 seconds. The choice of 300 seconds is standard in high frequency finance studies, and is motivated by the trade-off between bias and variance. See Aït-Sahalia et al. (2005), Zhang et al. (2005) and Bandi and Russell (2006) inter alios for a more detailed discussion.

- Table 5 Here -

The contributions of the different types of jumps to QV are shown in **Table 5**. The contribution of jumps decreases as the sampling interval increases from 5 to 300 seconds. For SPY, the share of jumps decreases from 43.2% (5 seconds) to 14.3% (300 seconds). For the 20 stocks, the average jump share decreases from 67.6% to 29.8%. In both cases, the decline is mainly due to the drop in the share of infinite jumps. The share of infinite jumps in SPY drops

---

[8] Qualitatively similar results were obtained using different block sizes (20 and 50 days), and additional bootstrap replications (10,000 and 20,000).

[9] TickData provides pre-cleaned and filtered price series. The algorithmic data filters identify bad prints, decimal errors, transposition errors and other data irregularities. The filters take advantage of the fact that, since the filtered data are not produced in real time, it is possible to look at the tick following a suspected bad tick before deciding whether or not the tick is valid. The filters are proprietary and based upon recent tick volatility, moving standard deviation windows and time of day. For a more detailed explanation, see the high frequency data filtering white paper on the TickData resources page https://www.tickdata.com/resources/white-papers/.

from 32.6% using 5-second returns to 0.1% using 300 second returns, and for the 20 stocks, the average share of infinite jumps drops from 34.2% to 0.2%. Hence, when returns are sampled every 300 seconds, the vast majority of jumps in SPY and the 20 stocks are infinite activity jumps. At this frequency, the small variations that characterize infinite jumps are close to Brownian increments. We find little evidence of asymmetry in the shares of signed jumps. The estimated Blumenthal-Getoor jump activity indices ($\hat{\beta}_{IJA}$), which measure the incidence of finite activity jumps, are consistent with the estimated shares of finite and infinite jump components. In the case of SPY, the index is 1.45 using 5-second returns and 0.78 using 300 second returns, which implies that infinite jumps are more prevalent at higher frequencies.

- Figures 1 and 2 Here -

Figure 1 plots the continuous and jump components of RV for SPY and the three stocks -- AMZN, HD and KO - with the largest, smallest and average RV. The days with jumps are shown in red, and other days in blue. It is clear that there is considerable heterogeneity in the level and timing of volatility. Although the highest spikes in volatility occur around the Dot-Com and Subprime crises (shaded areas), many other spikes in volatility are idiosyncratic. The 5- and 300-second autocorrelation functions of the SPY realized measures are displayed in **Figure 2**. The SPY RV_t and $\hat{C}_t$ measures appear to be long memory processes since their autocorrelations do not decline exponentially. The ACF of the 5-second RV_t and $\hat{C}_t$ measures (left panel) lie below their 300-second counterparts (right panel) - a hint that volatility forecasts using 300-second realized measures may perform better than ones using 5-second realized measures.

6. Extended HAR Model Results

6.1 SPY Forecasting Results

Since we use the HAR-RV model as a baseline for assessing the forecasting performance of our extended HAR models, **Table 6** sets out the in-sample coefficients, as well as the in- and out-of-sample R^2s and *MSPE*s, of the HAR-RV model for four forecast horizons – $h = 1$ (day), $h = 5$ (week), $h = 22$ (month), $h = 66$ (three months) - using returns sampled every 300 seconds. The significance of the coefficients is evaluated using Newey-West HAC-robust standard errors, allowing for serial correlation of up to order 5 ($h = 1$), 10 ($h = 5$), 44 ($h = 22$), or 132 ($h = 66$), since the random error terms in the models are serially correlated at least up to order $h - 1$. In following Andersen et al. (1999) and Patton and Sheppard (2015), we estimate the out-of-sample R squared, $R^2_{(oos)}$, as 1 minus the ratio of the out-of-sample model-based MSPE to the out-of-sample MSPE from a model including only a constant. The MSPE results are based on pseudo out-of-sample, rolling regression forecasts using a 1000 day window

- Table 6 Here -

All the coefficients in Table 6 are significant even at the three month horizon, confirming the high persistence of volatility. The magnitude of the daily and weekly coefficients decrease as we lengthen the forecast horizon. Although, the magnitude of the monthly coefficient changes little with the horizons, it's relative importance increases at longer horizons.[10]

- Table 7 Here -

Summary forecastsing results for extended HAR-CJ (jumps), HAR-CFJ (finite jumps), and the HAR-CIJ (infinite jumps) models are presented in **Table 7**, also using 300 second returns. In- and out of sample R^2s and $MSPE$s are presented for unsigned jumps, positive signed jumeps and negative signed jumps. Full results are available on request. A few points about the coefficient estimates are worth noting. The restrictions that the coefficients on finite and infinite jumps are the same, and that the coefficients on positive and negative jumps are the same, are decisively rejected. In line with Andersen et al (2007a) and Patton and Sheppard (2013), overall jumps tend to reduce future volatility, while negative jumps tend to increase it and positive jumps decrease it. Finite (infinite) jumps tend to decrease (increase) future volatility.

Unsurprisingly, the in-sample R-squared statistics (R^2_{in}) in **Table 7** suggest that incorporating jumps as predictors results in a better fit for our models. outperforming the baseline HAR-RV across the four horizons under examination. The out-of-sample R-squared statistics (R^2_{oos}) show that extended HAR models outperform the baseline model at one day and one week horizons, and about half the time at longer horizons. The models with positive jumps have higher R^2_{oos}'s at all horizons. Turning to the $MSPE$ results, the forecasting performance of the extended HAR models is significantly better at the one day and one week horizons, and better (significantly better) about half (one quarter) of the time at the one

[10] These results are been well documented in the literature (see Andersen et al 2007a, Corsi 2029 and Corsi et al. 2010 among others).

month- and three-month horizons. Note that no single extended HAR model outperforms all the others, a finding also reported in Patton and Sheppard (2009), which suggests that model averages combining the information contained in the different volatility forecasting models may generate forecast gains. See Section 7 below.

6.2 SPY Forecasting Results Using Standard and Robust-to-Noise Realized Measures

We know that microstructure noise is important at high frequencies, and the resulting attentuation bias may generate less accurate volativity forecasts than forecasts using noise-robust measures, such as the ones discussed in Section 3 above. We examined this issue in detail. **Table 8** compares the forecasting performance of SPY extended HAR volatility models using standard versus noise-robust realized measures, identifing models with significantly lower MSPEs than the baseline HAR-RV model. The entries in the top panel are based on forecasts using standard realized jump measures as explanatory variables; the bottom panel entries are based on robust-to-noise measures. The entries are relative MSPEs – the ratio of the MSPE of the proposed model to the MSPE of the corresponding baseline HAR-RV model – so ratios below one indicate more accurate roling regression forecasts.[11] Models with significantly lower MSPE than the baseline HAR-RV model, based on pair-wise Diebold-Mariano (DM, 1995) tests, are starred. The DM tests show that many of the extended HAR models in **Table 8** forecast as well as, or better, than the baseline HAR-RV models, although there is considerable variation across sampling frequencies and time horizon.

- Table 8 Here -

At the 5 and 60 second frequencies, the forecasts from models using noise-robust realized jump measures are somewhat more accurate than forecasts based on regular realized jump measures. Many models using 5 and 60 second regular volatility measures are excluded from the MCS at longer horizons, confirming the importance of taking account of micro-structure noise at higher frequencies. Nevertheless, the MSPE numbers for the baseline HAR-RV model in the final row of **Table 8** suggest that models using 300-second volatility measures tend to give better forecasts than models using 5 or 60-second returns, irrespective of whether standard or robust-to-noise volatility measures are used.

---

[11] The MSPE results are based on pseudo out-of-sample, rolling regression forecast using a 1,000 day window. Most models are retained in the model confidence set (MCS); the small number of entries for models that are not retained in the MCS are identified with the suffix x. The MCS results are generated using a 10-day block bootstrap and 5,000 replications.

6.3 Extended HAR Model Forecasting Results for the 20 S&P Stocks

Some results for the 20 S&P stocks are presented in **Table 9**. The relative MSPE entries (averaged across the 20 stocks) are shown in the body of the table, while the average MSPEs for the baseline HAR-RV models using standard realized measures are shown in the final row of the table. The entries for models which are not retained in the MCS at least 15 times (out of 20) are suffixed withthe letter x. The relative MSPE entries for the 20 stocks are more clustered around one than the entries for SPY in **Table 8**.[12] In addition, with the majority of the models are retained in the MCS at least 15 times, this indicates that the improvement in the forecasting performance of extended models with jumps is less clearcut for the 20 stocks, than it is for SPY. At the 5 and 60-second frequencies, the results show that noise-robust volatility measures work best. However, consistent with the results for SPY, forecasts using 300 second volatility measures are generally better than forecast using 5 or 60 second measures. In addition, the relative MSPEs of the standard volatility measures are often lower than those of the robust-to-noise measures.

-- Table 9 and Figure 3 Here -

No single extended HAR model with jumps dominates all the other models – the main reason being the small number of systematic jumps across the 20 stocks.[13] We find that, on average, cojumps only contribute to 9% of the total jump component, which means that most jumps are idiosyncratic. For instance, the left panel of **Figure 3** shows the returns on May 26th 2010, the day of the so called Flash Crash, one of the few days when the stocks jumped together. The movement in returns on that date is very different from returns on a typical day such as December 23rd 2003 (right-panel) in which only idiosyncratic jumps are present. Since the idiosyncratic jumps are stock specific reactions to news, what it is perceived as negative news for one stock might be positive news for another stock, generating jumps of different sizes and directions. Aït-Sahalia and Xiu (2016) suggest that co-jumps stem from surprising news announcements that occur primarily before the opening of the US market. Amengual and Xiu (2018) note that downward volatility jumps in the S&P500 are associated with a resolution of policy uncertainty, mostly through statements from FOMC meetings and speeches by the chair of the Federal Reserve. Aït-Sahalia, Kalnina and Xiu (2020) find that idiosyncratic jumps are related to idiosyncratic events such as

[12] The entries are also less dispersed because we are reporting averages.

[13] We identify jumps using the co-exceedance procedure of Gilder et al. (2014), which relies on the intersection of the univariate jump tests.

earnings disappointments, with earnings disappointments have a larger effect than positive earnings surprises. Given the rich information content of the different jump classifications and since no single extended HAR model dominates, the next section focuses on whether model averaged forecasts consistently outperform the forecasts from the baseline HAR-RV and the best extended HAR models across sampling frequencies and forecasting horizons.

7. The Gains from Model Averaging

Thus far, we have shown that a variety of extended HAR volatility models, that account for the nature and sign of jumps generate, significant improvements in forecasting performance. However, no single specification consistently outperforms the other models across horizons and frequencies, which suggests that model averaging might generate further forecasting gains. Four simple approaches to assigning model averaging weights are considered.[14] The aim of model averaging is to exploit relevant information embedded in the different forecasts, and produce an ensemble model that outperforms the benchmark HAR-RV model and, more importantly, the best single, extended HAR-RV jump model. Our approaches follow the literature closely (see, for instance, Bates and Granger 1969, Aiolfi and Timmermann 2006, Aiolfi et al. 2011, Elliott and Timmermann 2016, and the references therein).

- **Table 10 Here** -

We present model averaging results for the four sets of weights tabulated below - weights minimizing the estimated variance of the prediction errors, inverse MSPE weights, inverse MSPE rank weights and equal weights (**Table 10**). In the first three cases, the weights are recalculated every time a new set of rolling window forecasts are generated, and we prune the set of models under consideration by only averaging models that are retained in the model confidence set.

- Table 11 Here -

The model averaging results for SPY and four stocks, chosen by the level of their jump activity, are set out in **Table 1**. The table shows the relative MSPEs for the best extended HAR-RV model and the four model averaging approaches. The MSPEs for each index or stock and forecast horizon are measured relative to the MSPE of the corresponding HAR-RV model. All the stocks have estimated Blumenthal-

[14] We experimented with more complicated model averaging procedures, but the results were similar to those presented here. To conserve space, we do not report these experiments, but the details are available on request.

Getoor indices in the range 0 to 1, so their returns include finite and infinite activity jumps, with finite jumps dominating. BA and KO with jump activity of 0.58 and 0.91 respectively are the extreme cases.

The bold entries in **Table 11** are model averages with lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. The starred entries denote models averages with significantly lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. The daggered entries denote models with significant lower MSPEs than the HAR-RV model, but not the best extended HAR model. The four model averages generate forecasts that typically outperform the benchmark model for the four forecast horizons examined: h = 1 (one day), h = 5 (one week), h = 22 (one month) and h = 66 (three months). For example, in the case of SPY with 300-second returns, the one-week relative MSPE of the best extended HAR model is 0.753 versus a range of 0.693 to 0.715 for the four model averages. The largest MSPE reductions are generally observed at the one-week horizon, followed by the one-month horizon.

We also compared the model averaging results for SPY using 60 and 300 second returns. The 300-second model average forecasts dominate the forecasts using 60-second returns, generating significantly lower MSPEs. The 300-second forecasts also dominate the unreported model average forecasts using 5-second returns. These results also hold for the four stocks reported here, and for the other 16 stocks. The 300-second model average MSPEs are generally lower than the MSPEs of both the benchmark HAR-RV and best extended HAR models. In about a quarter of the cases, the MSPEs from the 300-second model average are significantly lower than the MSPEs of the best extended HAR model.

In conclusion, model averaging the forecasts from extended HAR-RV models generally result in lower MSPEs. Forecasts using 300-second returns dominate forecasts using higher frequency returns. The MCS procedure for pruning dominated models and the use of time varying weights for the model averages are helpful. Simple weighting schemes, e.g. the use of inverse MSPEs or inverse MSPE ranks, work as well as schemes that are more complicated (see Patton and Sheppard, 2009).

8. Results Using Transaction-Time Sampled Volatility Measures

In this section, we examine the volatility forecasting performance of alternative jump measures based on a transaction-based sampling scheme. Relatively few studies consider alternative sampling schemes. For instance, Griffin and Oomen (2008) and Oomen (2006b) study the properties of alternative RV measures using clock/calendar, transaction and business time sampling, but they do not consider jumps. To the best of our knowledge, only Patton and Sheppard (2015) examine the forecasting performance of jump measures using transaction time sampling, but they do not compare the clock and transaction time-based volatility components and the forecasting performance thereof. We contribute to this literature in two ways. Firstly, we decompose clock and transaction-based RV measures into their continuous and jump components, including their signed and finite/infinite activity jump components. Secondly, we compare the volatility forecasting performance of the clock and transaction time-based measures, using our extended HAR model and model averaging frameworks.

For brevity, we only report results for SPY. The transaction-based volatility measures are calculated using a 78 intraday return sampling scheme as in Patton and Sheppard (2015). This is the transaction-based equivalent of the 300 second / 5 minute return sampling scheme, which is widely used in the literature. Intra-trading day returns are calculated by fixing opening and closing prices, and recording sprices at business time $\lfloor ik \rfloor$, where $i = 1, \dots, 79$, $k = \frac{S-1}{79}$, S is the number of unique date stamps per day, and $\lfloor . \rfloor$ denotes rounding down to the nearest integer.[15]

- Table 12 Here -

Table 12 shows that the transaction-based RV measure is primarily driven by its continuous part: the contribution of jumps to total QV is about 4.6% versus 12.3% for the clock-based measures. Almost all of the jumps are finite jumps, the same as for clock time, and there is little difference in the contribution of positive and negative jumps. Although most jumps are finite activity jumps, the smaller contribution of transaction time based jumps to total QV implies a somewhat smaller jump activity index $\hat{\beta}_{IJA}$ (0.708 versus 0.778).

- Table 13 Here -

The relative MSPEs in **Table 13** suggest that the forecasting performance of extended HAR models using transaction-based measures is comparable to that of the baseline HAR-RV model, in sharp contrast to forecasting performance of extended HAR models using clock-based measures. Similar to the clock-time results, the MSPEs of most of the extended models are lower than the MSPE of the baseline HAR-RV model at the one-day horizon, although only three forecasts have significantly lower MSPEs. By contrast, as the horizon increases, we only obtain a handful of statistically significant reductions in MSPEs. Consequently, the model confidence set now includes all the models; since the forecasting performance of all of the models is broadly similar, we cannot identify a set of superior models.

[15] Note that the clock- and transaction-based RV descriptive statistics for SPY very similar.

- Table 14 Here -

A comparison of clock- and transaction-time based SPY model averaging results is presented **Table 14**. Results are presented for daily, weekly, monthly, and quarterly horizons. With transaction-based sampling, simple model averaging procedures (using MSPE, rank or equal weights) generate statistically significant improvements in the MSPEs. However, the MSPE improvements are far smaller than the improvements with clock-based sampling, so the transaction-time based MSPEs are always higher than their clock-based counterparts. Based on these SPY results, as well as results for the 20 stocks that are not reported, we conclude that forecasts using volatility measures from transaction-based sampling of returns are inferior to forecasts from clock-based sampling.

9. Summary and Conclusion

We examine the gains in forecasting the volatility of equity prices by decomposing jumps by activity (finite/infinite) and by sign using high-frequency data for SPY and 20 individual stocks. Our key findings are as follows. Quadratic variation contains a significant jump component, even at the 300-second frequency. The contribution of infinite jumps is greater than that of finite jumps at higher frequencies. However, at the 300-second frequency, jumps are mainly of finite activity.

Extended HAR models, incorporating a variety of jump activity and sign measures, generate statistically significant in- and out-of-sample improvements for both SPY and the 20 individual stocks we examined. The use of noise-robust realized measures improve the forecasts of future volatility at higher frequencies. However, since market microstructure noise declines as the sampling frequency decreases, the forecasting advantage of the noise-robust jump volatility measures also diminishes.

The rolling window, out-of-sample forecast results suggest that the lowest MSPE forecasts are obtained using returns sampled every 300 seconds, rather than every 5 or 60 seconds. This result holds for all of the horizons we examined -- a day, a week, a month and a quarter -- irrespective of whether noise-robust volatility measures are, or are not, used. In terms of MSPEs, there is little to choose between standard or robust-to-noise measures at this frequency.

We also examine the volatility forecasting performance of alternative jump measures based on a transaction time-based sampling scheme. The transaction-based RV measures are mainly driven by their continuous component, and finite jumps dominate infinite jumps. Using transaction-based volatility measures, the overall forecasting performance of extended HAR models is similar to that of the baseline HAR-RV model. Our conclusion is that forecasts using realized volatility and jump measures based on transaction sampling are inferior to forecasts using clock-based sampling measures. As our findings relate to the role of jumps using transaction time versus calendar time based sampling, this underscores the importance of the appropriate choice of sampling scheme.

In the absence of a single dominant forecasting model, we investigate whether various model averaging procedures generate significant forecasting gains. In many cases, we prune the set of models using the MCS procedure of Hansen et al. (2011) to eliminate dominated models. We find that simple model averaging procedures generally result in significant gains in forecasting performance vis-á-vis the single best extended HAR model, which in turn outperforms the baseline HAR-RV model. For example, model averaged results using equal weights, or the normalized inverse MSPE weights in Bates and Granger (1969) perform as well as model averaged results where the weights minimize the variance of the prediction error.

References

Aiolfi, Marco, Carlos Capistrán, and Allan Timmermann. 2011. “Forecasts Combinations”. In The Oxford Handbook of Economic Forecasting, ed. by Michael P Clements and David F Hendry. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195398649.013.0013.

Aiolfi, Marco, and Allan Timmermann. 2006. “Persistence in forecasting performance and conditional combination strategies”. Journal of Econometrics 135(1-2): 31–53. https//doi.org/10.1016/j.jeconom.2005.07.015.

Aït-Sahalia, Yacine. 2004. “Disentangling diffusion from jumps”. Journal of Financial Economics 74(3): 487–528. https://doi.org/10.1016/j.jfineco.2003.09.005.

Aït-Sahalia, Yacine, and Jean Jacod. 2009. “Estimating the degree of activity of jumps in high frequency data”. The Annals of Statistics 37(5A): 2202–2244. https://doi.org/10.1214/08-aos640.

Aït-Sahalia, Yacine, and Jean Jacod. 2011. “Testing whether jumps have finite or infinite activity”. The Annals of Statistics 39(3): 1689–1719. https://doi.org/10.1214/11-aos873.

Aït-Sahalia, Yacine, and Jean Jacod. 2012. “Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data”. Journal of Economic Literature 50(4): 1007–50. https://doi.org/ 10.1257/jel.50.4.1007.

Aït-Sahalia, Yacine, and Jean Jacod. 2014. High-Frequency Financial Econometrics. Princeton University Press. https://doi. org/10.23943/princeton/9780691161433.001.0001.

Aït-Sahalia, Yacine, Jean Jacod, and Jia Li. 2012. “Testing for jumps in noisy high frequency data”, Journal of Econometrics 168(2): 207-222. https://doi.org/10.1016/j.jeconom.2011.12.004

Aït-Sahalia, Yacine, Ilze Kalnina, and Dacheng Xiu. 2020. “High-frequency factor models and regressions”, Journal of Econometrics 216(1): 86-105. https://doi.org/10.1016/j.jeconom.2020.01.007

Aït-Sahalia, Yacine, and Jean Jacod. 2009. “Testing for jumps in a discretely observed process”. Annals of Statistics 37(1): 184-222. https://doi:10.1214/07-aos568

Aït-Sahalia, Yacine, Per A. Mykland, and Lan Zhang. 2005. “How often to sample a continuous- time process in the presence of market microstructure noise”. The Review of Financial Studies 18(2): 351–416. https://doi.org/10.1093/rfs/hhi016.

Aït-Sahalia, Yacine, and Dacheng Xiu. 2016. “Increased correlation among asset classes: Are volatility or jumps to blame, or both?”. Journal of Econometrics 194(2): 205-219. https://doi.org/10.1016/j.jeconom.2016.05.002

Aït-Sahalia, Yacine, and Dacheng Xiu. 2019. “A Hausman test for the presence of market microstructure noise in high frequency data”. Journal of Econometrics 211(1): 176-205. https://doi.org/10.1016/j.jeconom.2018.12.013

Amengual, Dante, and Dacheng Xiu. 2018. “Resolution of policy uncertainty and sudden declines in volatility”. Journal of Econometrics 203(2): 297-315. https://doi.org/10.1016/j.jeconom.2017.12.003

Andersen, Torben G., and Tim Bollerslev. 1998. “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts”. International Economic Review 39(4): 885– 905. https://doi.org/10.2307/2527343.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2007a. “Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility”. The Review of Economics and Statistics 89(4): 701–720. https://doi.org/10.1162/rest.89.4.701.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold and Paul Labys. 2001. “The distribution of realized exchange rate volatility”. Journal of the American Statistical Association 96(453): 42–55. https://doi.org/10. 1198/016214501750332965.

Andersen, Torben G., Tim Bollerslev, Francis X Diebold and Paul Labys. 2003. “Modeling and forecasting realized volatility”. Econometrica 71(2): 579–625. https://doi.org/10.1111/1468-0262.00418.

Andersen, Torben G., Tim Bollerslev, and Dobrislav Dobrev. 2007b. “No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications”. Journal of Econometrics 138(1): 125–180. https://doi.org/10.1016/j.jeconom.2006.05.018.

Andersen, Torben G., Tim Bollerslev, and Steve Lange. 1999. “Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon”. Journal of Empirical Finance 6(5): 457–477. https://doi.org/10.1016/s0927-5398(99)00013-4.

Andersen, Torben G., Tim Bollerslev, and Nour Meddahi. 2005. “Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities”. Econometrica 73(1): 279–296. https://doi.org/10.1111/j.1468-0262.2005.00572.x.

Bandi, Federico M., and Jeffrey R. Russell. 2006. “Separating microstructure noise from volatility”. Journal of Financial Economics 79 (3): 655–692. https://doi. org/10.1016/j.jfineco.2005.01.005.

Barndorff-Nielsen, Ole E., Silja Kinnebrock, and Neil Shephard. 2010. Measuring downside risk-realised semivariance. In Bollerslev, Tim, Jeffrey Russell, and Mark Watson, (eds), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199549498.003.0007.

Barndorff-Nielsen, Ole E., Peter R. Hansen, Asger Lunde, and Neil Shephard. 2008. “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise”. Econometrica 76(6): 1481–1536. https://doi.org/10.3982/ecta6495.

Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. “Power and bipower variation with stochastic volatility and jumps”. Journal of Financial Econometrics 2(1): 1–37. https://doi.org/10.1093/jjfinec/nbh001.

Bates, John M., and Clive W. J. Granger. 1969. “The combination of forecasts”. Journal of the Operational Research Society 20(4): 451–468. https://doi.org/10.1057/jors.1969.103.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg. 2016. "Exploiting the errors: A simple approach for improved volatility forecasting". Journal of Econometrics 192(1): 1-18. https://doi.org/10.1016/j.jeconom.2015.10.007.

Busch, Thomas, Bent J. Christensen, and Morten Ø. Nielsen. 2011. "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets". Journal of Econometrics 160(1): 48–57. https://doi.org/ 10.1016/j.jeconom.2010.03.014.

Caporin, Massimiliano, Aleksey Kolokolov, and Roberto Renò. 2017. "Systemic co-jumps". Journal of Financial Economics 126(3): 563–591. https://doi.org/10.1016/j.jfineco.2017.06.016.

Christensen, Kim, Ulrich Hounyo, and Mark Podolskij. 2018. "Is the diurnal pattern sufficient to explain intraday variation in volatility? A nonparametric assessment". Journal of Econometrics 205(2); 336-362. https://doi.org/10.1016/j.jeconom.2018.03.016.

Christensen, Kim, Roel C. A .Oomen, and Mark Podolskij. 2014. "Fact or fiction: Jumps at ultra high frequency". Journal of Financial Economics 114(3): 576–599. https://doi.org/10.1016/j.jfineco.2014.07.007.

Corsi, Fulvio. 2009. "A simple approximate long-memory model of realized volatility". Journal of Financial Econometrics 7(2): 174–196. https://doi.org/10.1093/jjfinec/ nbp001.

Corsi, Fulvio, Davide Pirino, and Roberto Renò. 2010. "Threshold bipower variation and the impact of jumps on volatility forecasting". Journal of Econometrics 159(2): 276– 288. https://doi.org/10.1016/j.jeconom.2010.07.008.

Diebold, Francis X., and Robert S Mariano. 1995. "Comparing predictive accuracy". Journal of Business & Economic Statistics 20(1): 134–144. https://doi.org /10.1198/073500102753410444.

Dumitru, Ana-Maria, and Giovanni Urga. 2012. "Identifying jumps in financial assets: a comparison between nonparametric jump tests". Journal of Business & Economic Statistics 30(2): 242–255. https://doi.org/10.1080/07350015.2012.663250.

Duong, Diep, and Norman R. Swanson. 2015. "Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction". Journal of Econometrics 187(2): 606–621. https://doi.org/10.1016/j.jeconom.2015.02.042.

Elliott, Graham, and Allan Timmermann. 2016. Economic Forecasting. Princeton University Press.

Forsberg, Lars, and Eric Ghysels. 2007. "Why do absolute returns predict volatility so well?". Journal of Financial Econometrics 5(1): 31–67. https://doi.org/10.1093/jjfinec/nbl010.

Ghysels, Eric, and Arthur Sinko. 2011. "Volatility forecasting and microstructure noise". Journal of Econometrics 160(1): 257–271. https://doi.org/10.1016/j.jempfin.2009.03.002.

Giacomini, Raffaella, and Barbara Rossi. 2010. "Forecast comparisons in unstable environments". Journal of Applied Econometrics 25(4): 595–620. https://doi.org/10.1002/ jae.1177.

Gilder, Dudley, Mark B Shackleton, and Stephen J. Taylor. 2014. "Cojumps in stock prices: Empirical evidence". Journal of Banking & Finance 40: 443–459. https://doi.org/10.1016/j.jbankfin.2013.04.025.

Giot, Pierre, and Sébastien Laurent. 2007. "The information content of implied volatility in light of the jump/continuous decomposition of realized volatility". Journal of Futures Markets 27(4): 337–359. https://doi.org/10.1002/fut.20251.

Griffin, Jim E., and Roel C. A. Oomen. 2008. "Sampling returns for realized variance calculations: tick time or transaction time?" Econometric Reviews 27(1-3): 230–253. https://doi.org/10.1080/07474930701873341.

Hansen, Peter R., and Asger Lunde. 2006. "Realized variance and market microstructure noise". Journal of Business & Economic Statistics 24(2): 127–161. https://doi.org/10.1198/073500106000000071.

Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. "The model confidence set". Econometrica 79(2): 453–497. https://doi.org/10.3982/ecta5771.

Heston, Steven L. 1993. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". The Review of Financial Studies 6(2): 327–43. https://doi.org/10.1093/rfs/6.2.327.

Huang, Xin, and George Tauchen. 2005. "The relative contribution of jumps to total price variance". Journal of Financial Econometrics 3(4): 456–499. https://doi.org/10.1093/jjfinec/nbi025.

Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. 2009. "Microstructure noise in the continuous case: the pre-averaging approach". Stochastic Processes and Their Applications 119(7): 2249–2276. https://doi.org/10.1016/j.spa.2008.11.004.

Jing, Bing-Yi, Xin-Bing Kong, Zhi Liu, and Per Mykland. 2012. "On the jump activity index for semimartingales". Journal of Econometrics 166(2): 213–223. https://doi.org/10.1016/j.jeconom.2011.09.036.

Lee, Suzanne S., and Per A. Mykland. 2007. "Jumps in financial markets: A new nonparametric test and jump dynamics". The Review of Financial Studies 21(6): 2535–2563. https://doi.org/10.1093/rfs/hhm056.

Li, Jia, and Dacheng Xiu. 2016. "Generalized method of integrated moments for high-frequency data". Econometrica 84(4): 1613-1633. https://doi.org/10.3982/ecta12306

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard. 2015. "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes". Journal of Econometrics 187(1): 293–311. https://doi.org/10.1016/j.jeconom.2015.02.008.

Martens, Martin, Dick Van Dijk, and Michiel De Pooter. 2009. "Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements". International Journal of Forecasting 25(2): 282–303. https://doi.org/10.1016/j.ijforecast.2009.01.010.

Mukherjee, Arpita, Wejia Peng, Norman R. Swanson, and Xiye Ying. 2020. "Financial econometrics and big data: A survey of volatility estimators and tests for the presence jumps and co-jumps". In Vinod, Hrishikesh, and C R Rao (eds.), Financial, Macro and Micro Econometrics Using R, Handbook of Statistics 42, North-Holland, Amsterdam.

Oomen, Roel C. A. 2006a. "Realized Variance and Market Microstructure Noise: Comment." Journal of Business & Economic Statistics 24(2): 195–202. https://doi.org/10.1198/073500106000000125.

Oomen, Roel C. A. 2006b. "Properties of realized variance under alternative sampling schemes". Journal of Business & Economic Statistics 24(2): 219–237. https://doi.org/10.1198/073500106000000044.

Patton, Andrew J. 2011. "Volatility forecast comparison using imperfect volatility proxies". Journal of Econometrics 160(1): 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034.

Patton, Andrew J. and Kevin Sheppard. 2009. "Optimal combinations of realised volatility estimators". International Journal of Forecasting 25(2): 218-238. https://doi.org/10.1016/j.ijforecast.2009.01.011.

Patton, Andrew J., and Kevin Sheppard. 2015. "Good volatility, bad volatility: Signed jumps and the persistence of volatility". Review of Economics and Statistics 97(3): 683–697. https://doi.org/10.1016/j.jeconom.2010.03.034.

Prokopczuk, Marcel, Lazaros Symeonidis, and Chardin Wese Simen. 2016. "Do jumps matter for volatility forecasting? Evidence from energy markets". Journal of Futures Markets 36(8): 758–792. https://doi.org/10.1002/fut.21759.

Sévi, Benoît. 2014. "Forecasting the volatility of crude oil futures using intraday data". European Journal of Operational Research 235(3): 643–659. https://doi.org/10.1016/j.ejor.2014.01.019.

Timmermann, Allan. 2006. "Forecast Combinations". In Handbook of Economic Forecasting, ed. by Graham Elliot, Clive W. J. Granger, and Allan Timmermann, 1: 135–196. Elsevier. https//doi.org/10.1016/S1574-0706(05)01004-9.

Todorov, Viktor, and George Tauchen. 2010. "Activity signature functions for high-frequency data analysis". Journal of Econometrics 154(2): 125–138. https://doi.org/10.1016/ j.jeconom.2009.06.009.

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia. 2005. "A tale of two time scales: Determining integrated volatility with noisy high-frequency data". Journal of the American Statistical Association 100(472): 1394–1411. https://doi.org/10.1198/016214505000000169.

Table 1: Realized Jump Measures

Use	Measure	Formula
QV Contributions	Finite Activity Jumps	$\widehat{FJ}_t = \hat{J}_t \times \mathcal{F}_t$
	Infinite Activity Jump	$\widehat{IJ}_t = \hat{J}_t \times (1 - \mathcal{F}_t)$
	Positive Jumps	$\widehat{PJ}_t = \max\left(RS_t^+ - \frac{1}{2}TBPV_t, 0\right) \times \mathcal{J}_t$
	Negative Jumps	$\widehat{NJ}_t = \max\left(RS_t^- - \frac{1}{2}TBPV_t, 0\right) \times \mathcal{J}_t$
Forecasting Models	Signed Jumps	$\widehat{SJ}_t = \widehat{PJ}_t - \widehat{NJ}_t$
	Positive Signed Jumps	$\hat{J}_t^+ = \widehat{SJ}_t \times \mathcal{P}_t$
	Negative Signed Jumps	$\hat{J}_t^- = \widehat{SJ}_t \times (1 - \mathcal{P}_t)$
	Positive Signed Finite Activity Jumps	$\widehat{FJ}_t^+ = \hat{J}_t^+ \times \mathcal{F}_t$
	Negative Signed Finite Activity Jumps	$\widehat{FJ}_t^- = \hat{J}_t^- \times \mathcal{F}_t$
	Positive Signed Infinite Activity Jumps	$\widehat{IJ}_t^+ = \hat{J}_t^+ \times (1 - \mathcal{F}_t)$
	Negative Signed Infinite Activity Jumps	$\widehat{IJ}_t^- = \hat{J}_t^- \times (1 - \mathcal{F}_t)$

Notes: $TBPV$, RS_t^+ and RS_t^1 are the threshold bipower variation measure of Corsi et al. (2010) and the realized semivariance measures of Barndorff-Nielsen et al. (2010) respectively. $\mathcal{J}$ is a 0/1 indicator for days with significant jumps based on the Andersen et al. (ABD, 2007b) test. $\mathcal{F}$ is an indicator for days with finite activity jumps based on the SFA test of Aït-Sahalia and Jacod (2011). $\mathcal{P} = 1(\widehat{SJ} > 0)$ is an indicator for days with net positive signed jumps. Noise-robust versions of the realized jump measures are described in the text.

Table 2: Noise-Robust ABD Test - Size and Power Simulations

Noise-to-Signal Ratio	$\xi = 0.01$			$\xi = 0.1$		
Sampling Frequency	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
			Size			
ABD	0.030	0.055	0.128	0.029	0.045	0.082
ABD Noise-Robust	0.070	0.041	0.015	0.049	0.011	0.006
			Power – Compound Poisson (Finite Activity)			
ABD	0.985	0.989	0.986	0.337	0.484	0.586
ABD Noise-Robust	1.000	0.991	0.703	0.963	0.905	0.458
			Power – Cauchy Process (Infinite Activity)			
ABD	0.732	0.774	0.764	0.361	0.415	0.463
ABD Noise-Robust	0.948	0.784	0.410	0.670	0.572	0.342

Notes: The table report the empirical size and power of the ABD test of Andersen et al. (2007b), and our modified, noise-robust version. ξ is the noise-to-signal ratio used to simulate market microstructure noise. The theoretical size of the tests is 5% ($\alpha = 0.05$). The models and Monte Carlo settings are described Section 3.3 of the paper.

Table 3: Realized and Noise-Robust Two-Scale Realized Semivariances – Finite Sample MSE Performance in a Heston (1993) Model with Compound Poisson Jumps

Noise-to-Signal Ratio	$\xi = 0.01$			$\xi = 0.1$		
Sampling Frequency	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
RS^{+}	9.680	0.067	0.003	963.848	6.718	0.277
RS^{-}	9.704	0.069	0.004	964.483	6.779	0.290
$TSRS^{+}$	0.001	0.001	0.002	0.113	0.014	0.008
$TSRS^{-}$	0.001	0.001	0.002	0.112	0.015	0.009

Notes: The table entries are the MSEs of the realized and two-scale realized semivariances in the simulation exercise described in Section X of the paper. The data generation process is a Heston (1993) model augmented with finite activity, compound Poisson jumps. ξ represents the noise-to-signal ratio used to simulate market micro-structure noise. Second-by-second prices were simulated 3,000 times for 50 days with 6½ trading hours per day.

Table 4: Extended HAR Models

Jumps: Total, Positive & Negative	HAR-CJ	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta_{J,d}\hat{J}_{t-1} + \beta_{J,w}\hat{J}_{t-5,t-1} + \beta_{J,m}\hat{J}_{t-22,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CJ$^+$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta_{J,d}\hat{J}^+_{t-1} + \beta_{J,w}\hat{J}^+_{t-5,t-1} + \beta_{J,m}\hat{J}^+_{t-5,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CJ$^-$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta_{J,d}\hat{J}^-_{t-1} + \beta_{J,w}\hat{J}^-_{t-5,t-1} + \beta_{J,m}\hat{J}^-_{t-5,t-1} + \varepsilon_{t,t+h-1}$
Finite Jumps: Total, Positive & Negative	HAR-CFJ	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta_{FJ,d}\widehat{FJ}_{t-1} + \beta_{FJ,w}\widehat{FJ}_{t-5,t-1} + \beta_{FJ,m}\widehat{FJ}_{t-22,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CFJ$^+$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta^+_{FJ,d}\widehat{FJ}^+_{t-1} + \beta^+_{FJ,w}\widehat{FJ}^+_{t-5,t-1} + \beta^+_{FJ,m}\widehat{FJ}^+_{t-5,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CFJ$^-$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta^-_{FJ,d}\widehat{FJ}^-_{t-1} + \beta^-_{FJ,w}\widehat{FJ}^-_{t-5,t-1} + \beta^-_{FJ,m}\widehat{FJ}^-_{t-22,t-1} + \varepsilon_{t,t+h-1}$
Infinite Jumps: Total, Positive & Negative	HAR-CIJ	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta_{IJ,d}\widehat{IJ}_{t-1} + \beta_{IJ,w}\widehat{IJ}_{t-5,t-1} + \beta_{IJ,m}\widehat{IJ}_{t-22,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CIJ$^+$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta^+_{IJ,d}\widehat{IJ}^+_{t-1} + \beta^+_{IJ,w}\widehat{IJ}^+_{t-5,t-1} + \beta^+_{IJ,m}\widehat{IJ}^+_{t-22,t-1} + \varepsilon_{t,t+h-1}$
	HAR-CIJ$^-$	$RV_{t,t+h-1} = \beta_0 + \beta_{C,d}\hat{C}_{t-1} + \beta_{C,w}\hat{C}_{t-5,t-1} + \beta_{C,m}\hat{C}_{t-22,t-1} + \beta^-_{IJ,d}\widehat{IJ}^-_{t-1} + \beta^-_{IJ,w}\widehat{IJ}^-_{t-5,t-1} + \beta^-_{IJ,m}\widehat{IJ}^-_{t-22,t-1} + \varepsilon_{t,t+h-1}$

Note: The table lists the extended HAR models used in this paper.

Table 5: Estimated Contributions of Signed, Finite and Infinite Activity Jumps to QV (%)

Shares of QV		SPY			Avg. 20 Stocks		AMZN	BA	BFB	CAT	CHL	COST	CVX
	5s	60s	300s	5s	60s	300s	300s	300s	300s	300s	300s	300s	300s
	%	%	%	%	%	%	%	%	%	%	%	%	%
Continuous	56.80	88.47	85.72	32.40	65.61	70.20	73.43	72.59	55.14	74.90	62.18	69.53	80.28
Jumps	43.20	11.53	14.28	67.60	34.39	29.80	26.57	27.41	44.86	25.10	37.82	30.48	19.72
Pos. Jumps	21.85	6.45	8.26	33.95	16.54	14.99	15.21	14.36	22.47	12.57	17.98	15.96	9.85
Neg. Jumps	21.36	5.08	6.02	33.65	17.85	14.81	11.37	13.05	22.38	12.53	19.84	14.51	9.87
Finite Jumps	10.60	10.42	14.16	33.39	32.42	29.60	26.41	27.23	44.65	24.85	37.31	30.36	19.58
Infinite Jumps	32.60	1.11	0.12	34.21	1.97	0.21	0.17	0.19	0.21	0.25	0.50	0.12	0.15
Pos. Finite Jumps	5.58	5.94	8.22	17.03	15.54	14.88	15.13	14.25	22.38	12.47	17.68	15.89	9.77
Neg. Finite Jumps	5.02	4.48	5.94	16.37	16.88	14.71	11.28	12.98	22.27	12.39	19.63	14.47	9.81
Pos. Infinite Jumps	16.26	0.51	0.04	16.92	1.00	0.11	0.08	0.11	0.09	0.11	0.30	0.07	0.08
Neg. Infinite Jumps	16.34	0.60	0.08	17.29	0.98	0.10	0.08	0.07	0.12	0.14	0.21	0.05	0.06
Memo: Jump Activity, $\hat{\beta}_{IJA}$	1.454	1.056	0.778	1.455	1.040	0.723	0.461	0.576	0.802	0.621	0.763	0.697	0.748

Notes: The table reports the estimated percentage contribution of the different realized jump measures to QV. Results using 5- second, 60-second and 300-second returns are shown for SPY and the average of the 20 stocks. The results for the individual stocks were estimated using 300-second returns. $\hat{\beta}_{IJA}$ is the estimated Blumenthal-Getoor index of jump activity.

Table 5 (Continued): Estimated Contributions of Signed, Finite and Infinite Activity Jumps to QV (%)

QV Shares	DOW 300s	EXC 300s	GILD 300s	GS 300s	HD 300s	JNJ 300s	JPM 300s	KO 300s	OKE 300s	PG 300s	SO 300s	UPS 300s	WMT 300s
	%	%	%	%	%	%	%	%	%	%	%	%	%
Continuous	68.88	69.49	63.20	75.98	73.94	70.61	76.12	74.21	59.17	71.15	70.79	68.29	74.10
Jumps	31.12	30.51	36.80	24.02	26.07	29.39	23.88	25.79	40.83	28.85	29.21	31.71	25.90
Pos. Jumps	30.85	30.40	36.46	23.94	25.94	29.28	23.82	25.52	40.60	28.78	28.64	31.53	25.80
Neg. Jumps	0.27	0.11	0.34	0.08	0.13	0.11	0.06	0.27	0.23	0.08	0.57	0.18	0.10
Finite Jumps	15.03	15.51	18.91	12.31	13.88	12.92	12.93	12.50	19.06	15.42	14.49	15.48	13.01
Infinite Jumps	16.09	15.01	17.89	11.71	12.19	16.47	10.95	13.29	21.77	13.44	14.72	16.23	12.89
Pos. Finite Jumps	14.83	15.43	18.67	12.30	13.84	12.83	12.90	12.34	18.98	15.37	14.27	15.37	12.97
Neg. Finite Jumps	16.02	14.97	17.79	11.64	12.10	16.45	10.92	13.18	21.62	13.41	14.37	16.15	12.83
Pos. Infinite Jumps	0.20	0.07	0.24	0.01	0.03	0.09	0.03	0.16	0.08	0.05	0.21	0.10	0.05
Neg. Infinite Jumps	0.07	0.04	0.10	0.07	0.09	0.02	0.03	0.12	0.15	0.03	0.36	0.08	0.05
Memo: Jump Activity, $\hat{\beta}_{IJA}$	0.58	0.73	0.52	0.61	0.67	0.97	0.61	0.91	0.65	0.96	0.88	0.90	0.82

Notes: The table reports the estimated percentage contribution of the different realized jump measures to QV. Results using 5- second, 60-second and 300-second returns are shown for SPY and the average of the 20 stocks. The results for the individual stocks were estimated using 300-second returns. $\hat{\beta}_{IJA}$ is the estimated Blumenthal-Getoor index of jump activity.

Table 6: HAR-RV Benchmark - SPY, 300 Second Returns

$$RV_{t,t+h-1} = \beta_0 + \beta_d RV_{t-1} + \beta_w RV_{t-5,t-1} + \beta_m RV_{t-22,t-1} + \varepsilon_{t,t+h-1}$$

	Forecast Horizon (*h*)			
	1 Day	5 Days	22 Days	66 Days
β_0	0.095*	0.148**	0.288***	0.527***
β_d	0.246**	0.184***	0.103***	0.061***
β_w	0.422***	0.347***	0.322***	0.200***
β_m	0.238**	0.323***	0.290***	0.215***
R^2_{in}	0.512	0.629	0.562	0.337
R^2_{oos}	0.443	0.673	0.707	0.470
$MSPE$	3.102	1.322	0.944	1.262

Notes: The table reports the OLS coefficient estimates of the HAR-RV forecasting equation, and the in- and out-of-sample R^2s for the SPY RV at daily (h =1), weekly (h = 5), monthly (h = 22) and quarterly (h = 66) horizons. The RV measures are calculated using 300-second returns. The significance of the coefficients is based on Newey-West HAC standard errors, allowing for serial correlation up to order 5, 10, 44 or 132 for horizons h = 1, 5, 22 and 66 trading days. The superscripts *, ** and *** denote statistical significance at the 10%, 5% and 1% levels. The out-of-sample R-squared, R^2_{oos} , is calculated as one minus the ratio of the $MSPE$ from the HAR-RV model to the $MSPE$ from a model that only has an intercept.

Table 7: SPY Extended HAR Volatility Forecasting Regressions – Summary Performance Measures

Models / Horizons	h = 1	h = 5	h = 22	h = 66	h = 1	h = 5	h = 22	h = 66	h = 1	h = 5	h = 22	h = 66
Signed Jumps		HAR-CJ				HAR-CJ$^+$				HAR-CJ$^-$		
R^2_{in}	**0.555**	**0.666**	**0.572**	**0.338**	**0.541**	**0.668**	**0.578**	**0.341**	**0.523**	**0.664**	**0.612**	**0.362**
R^2_{oos}	**0.493**	**0.747**	**0.728**	0.465	**0.450**	**0.754**	**0.739**	**0.489**	**0.511**	**0.724**	0.690	0.445
$MSPE$	**2.821***	**1.017***	**0.872***	1.274	**3.059**	**0.995***	**0.840***	**1.218***	**2.720***	**1.110***	0.994	1.318
Signed Finite Jumps		HAR-CFJ				HAR-CFJ$^+$				HAR-CFJ$^-$		
R^2_{in}	**0.555**	**0.666**	**0.572**	**0.338**	**0.541**	**0.668**	**0.577**	**0.341**	**0.523**	**0.665**	**0.614**	**0.363**
R^2_{oos}	**0.493**	**0.747**	**0.728**	0.464	**0.449**	**0.753**	**0.734**	**0.478**	**0.511**	**0.724**	0.684	0.446
$MSPE$	**2.822***	**1.018***	**0.874***	1.276	**3.066**	**0.998***	**0.850***	**1.243**	**2.721***	**1.112***	0.994	1.317
Signed Infinite Jumps		HAR-CIJ				HAR-CIJ$^+$				HAR-CIJ$^-$		
R^2_{in}	**0.512**	**0.630**	**0.563**	**0.340**	**0.512**	**0.630**	**0.576**	**0.381**	**0.512**	**0.629**	**0.563**	**0.339**
R^2_{oos}	**0.493**	**0.747**	**0.728**	0.464	**0.449**	**0.753**	**0.734**	**0.478**	**0.511**	**0.724**	0.684	0.446
$MSPE$	**2.722***	**1.173***	1.151	1.316	**2.731***	**1.168***	1.125	1.264	**2.714***	**1.162***	1.121	1.299

Notes: See notes to Table 4. The bold in-sample and out-of-sample R-squared entries indicate that the fit of the proposed models is better that of the baseline HAR-RV model in Table 4. Bold MSPE entries are lower than the MSPEs of the benchmark models. Significantly lower MSPE entries at the 5% level are starred. The complete table of coefficient estimates is available on request.

Table 8: SPY Relative MSPEs by Frequency – Standard vs. Noise-Robust Measures

Forecast Horizon		h = 1 (day)			h = 5 (week)			h = 22 (month)			h = 66 (quarter)		
Sampling Frequency		5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
Standard ~~Raw~~ Jump Measures	HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000x	1.000x	1.000	1.000	1.000x	1.000	1.000
	HAR-CJ	1.253	0.755*	0.909*	1.029	0.990	0.770*	0.980*	1.172	0.924*	0.968	1.167x	1.010
	HAR-CJ$^+$	0.871*	0.752*	0.910*	1.181	0.992	0.770*	1.051	1.178	0.926*	1.010x	1.171x	1.011
	HAR-CJ$^-$	1.124	1.060	0.878*	1.022	1.034	0.888*	0.969*	1.001	1.220	0.940*	0.993	1.043
	HAR-CFJ	0.903*	0.993	0.986	1.165	0.969	0.753*	1.147x	0.894*	0.891*	1.074x	0.977	0.965*
	HAR-CFJ$^+$	0.848*	0.969	0.877*	1.124	1.017	0.840*	0.841*	0.936*	1.053	0.917*	1.020	1.045
	HAR-CFJ$^-$	0.925*	0.993	0.988*	1.175	0.971	0.755*	1.198x	0.877*	0.908*	1.096x	0.959	0.985
	HAR-CIJ	0.915*	0.969	0.877*	1.215	1.035	0.841*	0.982	0.959*	1.054	1.035x	1.020*	1.044
	HAR-CIJ$^+$	0.910*	1.055	0.881*	1.151	1.020	0.884*	1.086x	0.964*	1.192	1.136x	0.940*	1.002
	HAR-CIJ$^-$	0.729*	1.059	0.875*	0.996	1.030	0.879*	1.054x	0.921*	1.189	0.939*	0.977*	1.029
Noise-Robust Jump Measures	HAR-RV	0.843*	0.907*	1.009	0.882*	0.976	0.962	0.821*	1.031	1.154	0.893*	1.013	1.014
	HAR-CJ	0.768*	0.966	1.015	0.865*	1.010	0.962	0.977	1.044	1.145	0.988	0.996	0.906*
	HAR-CJ$^+$	0.775*	0.960*	1.015	0.867*	1.060	0.958*	0.987	1.031	1.143	0.921*	0.925*	1.032
	HAR-CJ$^-$	0.791*	0.980	1.018	0.890*	1.025	0.965	0.803*	1.073	1.179	0.875*	1.016	0.998
	HAR-CFJ	0.851*	0.684*	1.015	0.884*	0.930	0.960	0.838*	0.907*	1.145	0.926*	1.037	0.991
	HAR-CFJ$^+$	0.870*	0.852*	1.013	0.828*	0.889	0.953*	0.772	0.912	1.135	0.899*	0.968	0.997
	HAR-CFJ$^-$	0.866*	0.677*	1.015	0.895*	0.889	0.960	0.861*	0.938*	1.145	0.919*	1.037	0.990
	HAR-CIJ	1.111	0.852*	1.013	0.882*	0.894	0.953*	0.786*	0.902*	1.135	0.931*	0.953	0.753*
	HAR-CIJ$^+$	0.794*	0.972	1.026	0.875*	1.005	0.977	0.841*	1.166*	1.164	0.930*	1.038	0.994
	HAR-CIJ$^-$	1.009	0.958	1.016	0.793*	1.015	0.961	0.794*	0.947*	1.137	0.852*	0.941*	1.000
Memo: HAR-RV MSPE		3.364	4.550	3.102	1.553	1.350	1.322	1.443	1.025	0.944	1.778	1.344	1.262

Notes: The relative MSPE ratios are the ratios of the MSPEs of the extended HAR models using standard volatility measures (top panel) or robust-to-noise measures (bottom panel) to the MSPEs of HAR-RV models employing standard measures. The starred MSPE entries indicate statistically significant reductions in the MSPEs at the 5% level. Entries with an "x" suffix denote models not in the MCS. The MSPE and MCS results are based on rolling regression using 1,000 observations and a 10-day block bootstrap with 5,000 replications respectively.

Table 9: Twenty Stock Relative MSPE Averages – Standard vs. Noise-Robust Measures

Forecast Horizon		h = 1 (day)			h = 5 (week)			h = 22 (month)			h = 66 (quarter)		
Sampling Frequency		5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
Standard Jump Measures	HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000x	1.000	1.000
	HAR-CJ	0.999	0.991	0.972	0.950	0.916	0.933	0.929	0.942	0.970	0.928	0.958	0.995
	HAR-CJ$^+$	1.057	0.984	0.973	1.048x	0.916	0.934	1.064x	0.943	0.974	1.043	0.952	0.997
	HAR-CJ$^-$	1.010	0.973	0.940	0.986	0.955	0.942	1.035	1.010	1.063	1.007x	1.014x	1.037
	HAR-CFJ	1.044	1.000	0.968	1.098x	0.939	0.945	1.203x	0.994	1.038	1.127x	1.004x	1.033
	HAR-CFJ$^+$	1.063	1.018	0.932	1.038	0.943	0.934	1.144x	0.970	1.026	1.078x	0.997x	1.018
	HAR-CFJ$^-$	1.055x	0.999	0.969	1.153x	0.940	0.945	1.267x	0.994	1.038	1.173x	1.004	1.031
	HAR-CIJ	1.103x	0.984	0.932	1.115x	0.938	0.937	1.228x	0.970	1.030	1.144x	0.997x	1.016
	HAR-CIJ$^+$	1.044	0.979	0.939	1.090x	0.966	0.946	1.189x	1.010	1.080	1.129x	1.004x	1.042
	HAR-CIJ$^-$	1.011	0.982	0.947	1.071x	0.960	0.945	1.213x	1.005	1.091	1.137x	1.006x	1.062
Noise-Robust Jump Measures	HAR-RV	0.966	0.916	0.969	0.975	1.017	0.998	0.975	1.081	1.138	0.956x	1.050	1.032x
	HAR-CJ	0.958	0.935	0.975	0.934	0.975	0.990	0.958	1.077	1.135	0.949	1.040	0.962
	HAR-CJ$^+$	0.980	0.939	0.976	0.962	1.003	0.996	0.966	1.082	1.075	0.882	0.963	0.994
	HAR-CJ$^-$	0.969	0.926	0.970	0.956	1.022	0.985	0.943	1.064	1.122	0.905	1.042	1.021
	HAR-CFJ	0.955	0.986	0.978	0.962	1.008	0.991	0.981	1.082	1.092	0.956	1.042	1.017x
	HAR-CFJ$^+$	0.973	0.943	0.961	0.950	0.984	0.994	0.938	1.043	1.126	0.936x	1.030	1.019
	HAR-CFJ$^-$	0.947	0.987	0.980	0.952	1.010	0.993	0.967	1.086	1.091	0.924x	1.044	1.014
	HAR-CIJ	0.963	0.938	0.962	0.962	0.984	0.994	0.948	1.047	1.107	0.945x	1.031	1.024
	HAR-CIJ$^+$	0.972	0.926	0.950	0.957	1.022	0.994	0.966	1.073	1.091	0.952x	1.045	1.008
	HAR-CIJ$^-$	0.964	0.935	0.948	0.948	1.025	0.986	0.969	1.061	1.116	0.943x	1.037	1.033
Memo: HAR-RV MSPE		373.14	54.886	22.744	85.581	16.968	9.926	27.084	8.812	6.393	17.267	7.901	6.292

Notes: The relative MSPE ratios are the 20 stock average ratios of the MSPEs of the extended HAR models using standard volatility measures (top panel) or robust-to-noise measures (bottom panel) to the MSPEs of HAR-RV models employing standard measures. The entries with an "x" suffix denote models that are retained in the MCS for fewer than 15 stocks. The MSPE and MCS results are based on rolling regression using 1,000 observations and a 10-day block bootstrap with 5,000 replications respectively.

Table 10: Model Averaging Weights

Weights	Formula	Models
Min. Prediction Error Variance	$w_t^h = argmin_w \; w'\hat{\Sigma}_t^h w$ s.t. $\iota' w = 1$	MCS
Inverse MSPE	$w_{t,i}^h = \frac{(MSPE_{t,i}^h)^{-1}}{\sum_{i \in \mathcal{M}_t^h} (MSPE_{t,i}^h)^{-1}}$	MCS
Inverse Rank	$w_{t,i}^h = \frac{(Rank_{t,i}^h)^{-1}}{\sum_{i \in \mathcal{M}_t^h} (Rank_{t,i}^h)^{-1}}$	MCS
Equal Weights	$w_{i,t}^h = N^{-1}$	All

Note: $\hat{\Sigma}_t^h$ is the estimated, rolling window variance-covariance matrix of the set of MCS retained forecasting models ($\mathcal{M}_t^h$) for horizon h at time t. ι is a vector of ones with dimension equal to number of retained models. $MSPE_{t,i}^h$ and $Rank_{t,i}^h$ are the rolling window MSPEs and MCS ranks of the retained models. Nine extended HAR models are used in this study so $N = 9$.

Table 11: Model Averaging Results - Relative MSPEs at Different Horizons for SPY, BA, BFB, COST and KO

	SPY – 300 seconds				SPY – 60 seconds			
	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
Best Extended HAR	0.875†	0.753†	0.891†	0.965†	0.752†	0.969	0.877	0.940†
Avg. – Min. Var. Weights	0.987	**0.693***	0.895†	0.966†	0.812†	0.977	0.940†	0.971†
Avg. – Inverse MSPE Weights	0.879	**0.706***	**0.862***	**0.919***	0.875†	**0.914***	**0.850†**	0.965†
Avg. – Inverse Rank Weights	0.910†	**0.715†**	**0.845***	**0.873***	0.880†	**0.923†**	**0.846†**	0.986
Avg. – Equal Weights	**0.873†**	**0.712†**	**0.876†**	**0.928†**	0.877†	**0.914***	**0.852†**	0.964†
Memo: HAR-RV MSPE	3.102	1.322	0.944	1.262	4.550	1.350	1.025	1.344
	BA – 300 seconds				BFB – 300 seconds			
	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
Best Extended HAR	0.981	0.937	0.993	0.864†	0.924†	0.836†	0.822†	0.876†
Avg. – Min. Var. Weights	0.992	**0.905***	1.083	1.001	0.969†	0.845†	**0.751***	**0.812***
Avg. – Inverse MSPE Weights	**0.972†**	**0.906†**	**0.915***	0.959†	0.926†	**0.823†**	**0.814†**	**0.856***
Avg. – Inverse Rank Weights	**0.976†**	**0.923†**	**0.928***	0.980	0.936†	**0.820†**	**0.810***	**0.847***
Avg. – Equal Weights	**0.972†**	**0.906†**	**0.919***	0.961†	0.926†	**0.823†**	**0.816†**	0.878†
	COST – 300 seconds				KO – 300 seconds			
	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
Best Extended HAR	0.958†	0.879†	0.925†	0.957†	0.814†	0.709†	0.882†	0.939†
Avg. – Min. Var. Weights	1.016	0.985	**0.881***	**0.950†**	0.923†	**0.695***	**0.837***	**0.916†**
Avg. – Inverse MSPE Weights	0.962†	**0.871†**	**0.920†**	0.958†	0.817†	0.713†	0.888†	0.975†
Avg. – Inverse Rank Weights	0.969†	**0.856†**	**0.907***	**0.945***	**0.811†**	**0.686†**	**0.829***	0.950†
Avg. – Equal Weights	0.962†	**0.873†**	**0.922†**	0.960†	0.817†	0.723†	0.914†	0.983†

Notes: The table reports the relative MSPE, the ratio of MSPE of the model indicated in the first column to the MSPE of the baseline HAR-RV, in both cases using standard volatility measures as opposed to robust-to-noise measures. The best extended HAR models refers to the min. MSPE model from the set of extended HAR models set out in Section 4. The bold entries are model averages with lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. The starred entries denote models averages with significantly lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. Models with significant lower MSPEs than the HAR-RV model, but not the best extended HAR model, are labelled with a dagger (†).

Table 12: The Contribution of Jumps to SPY QV – Comparison of Clock and Transaction Based Sampling Results

	Clock-Based Sampling	Transaction-Based Sampling
	%	%
Continuous	85.725	95.413
Jumps	14.275	4.587
Pos. Jumps	8.257	2.279
Neg. Jumps	6.018	2.308
Finite Jumps	14.156	4.503
Infinite Jumps	0.118	0.084
Pos. Finite Jumps	8.219	2.232
Neg. Finite Jumps	5.937	2.271
Pos. Infinite Jumps	0.038	0.047
Neg. Infinite Jumps	0.080	0.038
$\hat{\beta}_{IJA}$	0.778	0.708

Notes: The table reports the contribution of the different realized jumps to SPY's QV using clock (300 second) and transaction-based (78 ticks per interval) sampling.

Table 13: SPY Extended HAR Volatility Forecast Results Using Transaction-Based Realized Measures

Model	Relative MSPEs			
	$h = 1$ (day)	$h = 5$ (week)	$h = 22$ (month)	$h = 66$ (quarter)
HAR-RV	1.000	1.000	1.000	1.000
HAR-CJ	0.973*	1.114	1.030	1.023
HAR-CJ $^+$	1.037	1.119	0.956*	0.971*
HAR-CJ $^-$	0.990	1.003	1.036	1.012
HAR-CFJ	0.973*	1.114	1.030	1.022
HAR-CFJ $^+$	1.037	1.119	0.956*	0.971*
HAR-CFJ $^-$	0.990	1.003	1.036	1.012
HAR-CIJ	0.981	0.999	1.061	1.017
HAR-CIJ $^+$	0.981*	0.996	1.052	1.011
HAR-CIJ $^-$	0.980*	0.997	1.064	1.016
Memo: HAR-RV MSPE	3.724	1.500	1.071	1.349

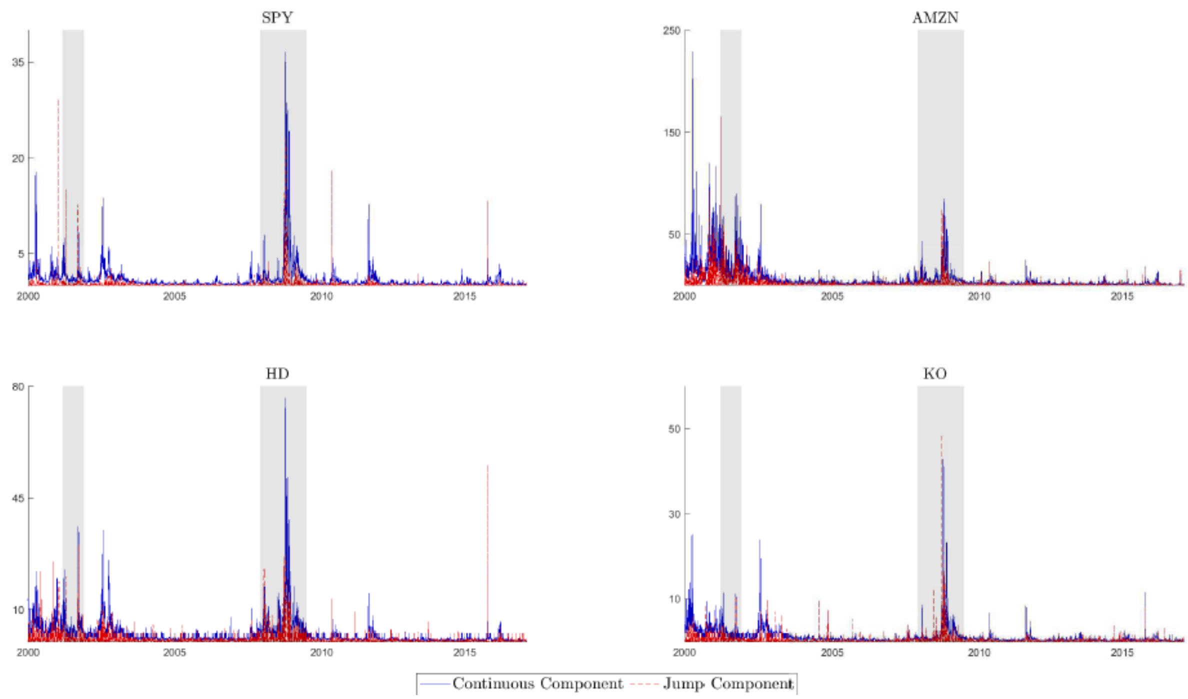
Notes: The table reports the relative MSPEs of the extended HAR SPY volatility forecasting models at different horizons. The relative MSPEs are the ratio of the MSPEs of the extended HAR models relative to the benchmark HAR-RV model. The starred entries indicate statistically significant reductions in MSPE identified by the Diebold and Mariano (1995) test using a 5% significance level and robust standard errors.

Table 14: SPY Model Averaging – Comparison of Clock and Transaction-Based Sampling Results

Model / Model Average	Relative MSPEs							
	Clock-Based Sampling (300 Sec.)				Transaction-Based Sampling			
	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
Best Extended HAR	0.875†	0.753†	0.891†	0.965†	0.973†	0.996	0.956†	0.971†
Avg.: Min. Var. Weights	0.987	**0.693***	0.895†	0.966†	1.009	0.995	**0.921***	1.001
Avg.: Inverse MSPE Weights	0.879†	**0.706***	**0.862***	**0.919***	**0.926***	**0.950***	**0.889***	**0.961†**
Avg.: Inverse MSPE Rank Weights	0.910	**0.715†**	**0.845***	**0.873***	**0.969†**	**0.957***	**0.855***	**0.943***
Avg.: Equal Weights	**0.873†**	**0.712†**	**0.876†**	**0.928***	**0.937***	**0.954***	**0.914***	**0.963†**
Memo: HAR-RV MSPE	3.102	1.322	0.944	1.262	3.724	1.500	1.071	1.349

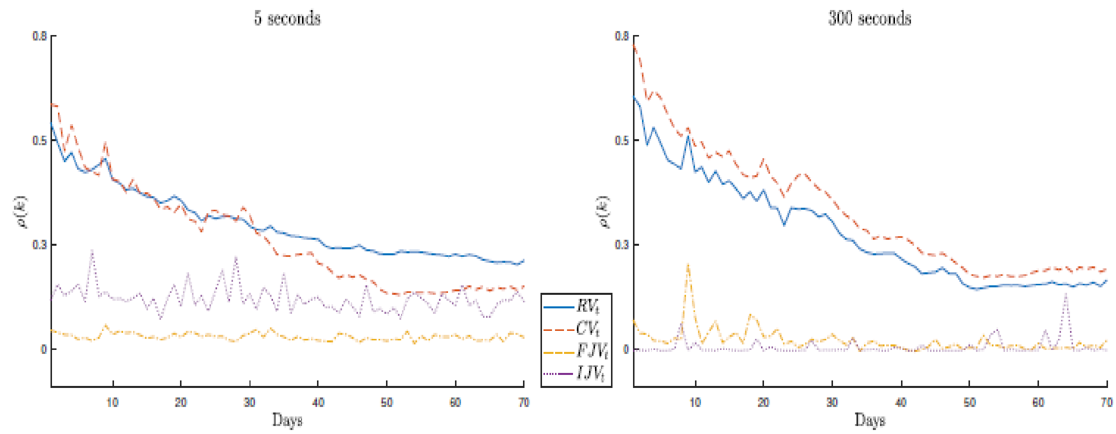
Notes: The table compares reports the forecasting performance of the extended HAR SPY volatility forecasting models at different horizons *h* using clock and transaction based realized measures. The clock-based results use 300 second returns. The relative MSPEs are the ratio of the MSPEs of the extended HAR models relative to the benchmark HAR-RV model. The bold entries are model averages with lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. The starred entries denote models averages with significantly lower MSPEs than the MSPEs of both the HAR-RV and best extended HAR models. Models with significant lower MSPEs than the HAR-RV model, but not the best extended HAR model, are labelled with a dagger (†).

Figure 1: Time Series of SPY, AMZN, HD and KO Realized Volatility – Jump and Continuous Components



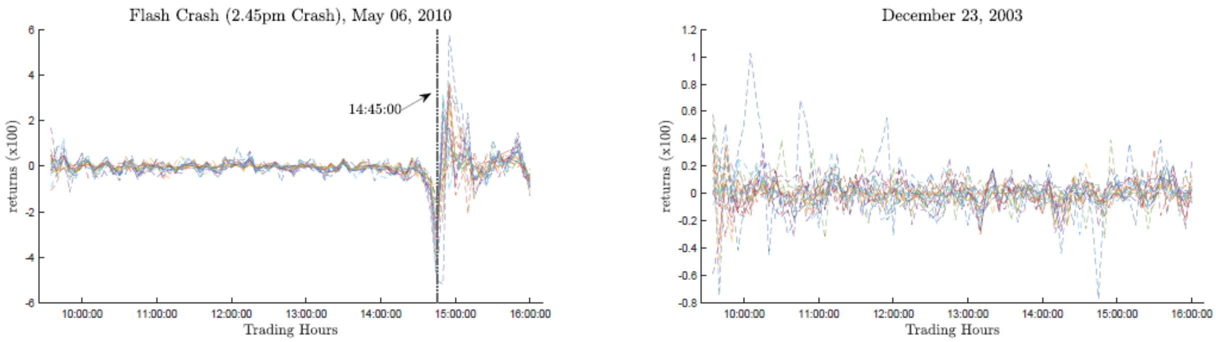
Note: This figure shows the continuous and jump components of RV for SPY, AMZN, HD and KO, using 300 second realized measures. The three stocks have the largest, smallest and average RV. NBER dated U.S. recession are shaded grey.

Figure 2: Autocorrelation Function of SPY Realized Measures - 5 and 300 Second Returns



Note: The figure graphs the autocorrelation function of SPY's RV and its continous, finite and infinite activity jump components. The autocorrelations at the 5 and 300 second frequencies were estimated using noise-robust and raw estimators, respectively.

Figure 3: Systematic Versus Idiosyncratic Jumps in Returns - May 6, 2010 (Flash Crash) and December 23, 2003 (Normal Day)



Note: The two panels plot the 5-minute returns for SPY and the 20 stocks used in this paper on two different days. The left panel shows the co-jump in returns on May 6, 2010 when U.S. stocks experienced a flash crash at 2.32 p.m. ET lasting approximately 36 minutes. The right panel shows returns on December 23, 2003, a typical day with no identified co-jumps.