Working Paper in Economics

 # 202109

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

Ruijun Bu
Rodrigo Hizmeri
Marwan Izzeldin
Anthony Murphy
Mike Tsionas

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

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

[†]*Lancaster University*
[‡]*Federal Reserve Bank of Dallas*
[§]*University of Liverpool*

Original Version: March, 2019
Revised: June, 2020

Abstract

This paper proposes a novel approach to decompose realized jump measures by type of activity (infinite/finite) and by sign. It also provides noise-robust versions of the ABD jump test (Andersen et al., 2007b) 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: Realized volatility, Signed Jumps, Finite Jumps, Infinite Jumps, Volatility Forecasts, Noise-Robust Volatility, Model Averaging.
JEL classification: C22, C51, C53, C58.

[◇]We would like to thank Torben Andersen, Andrew Patton, Peter R. Hansen, Ingmar and Sandra Nolte, Jia Li, George Tauchen, Nour Meddahi, Roberto Renò, Aleksey Kolokolov, 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 and 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 second author acknowledges the financial support of the Economic and Social Research Council (ESRC), Grant Number: ES/P000665/1.

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

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 provide 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 intraday squared returns, provides a consistent estimator of the quadratic variation when the price process contain discontinuities or jumps.[1]

In relation to volatility forecasting, the seminal work of Andersen et al. (2007a) 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 announcements. However, recent empirical evidence that classifies jumps into finite and infinite activity jumps (e.g. 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.[2]

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 frequencies such as 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. (2007a) 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 affords substantial out-of-sample gains. Patton and Sheppard (2015) argue that volatility is strongly related to the volatility of past negative returns, and show that models incorporating signed jumps

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

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

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.

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), 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 jump test and the SFA finite activity jump tests (Aït-Sahalia and Jacod, 2011; Andersen et al., 2007b). 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 tests based decomposition of days with significant jumps into ones with finite or infinite activity jumps. As noted by Aït-Sahalia and Jacod (2014), in finite samples the estimated jumps based on higher-order power variations are often poor measure of actual jumps. Second, we develop versions of the ABD test and realized semivariance measures that are robust to microstructure noise, and perform well at high-frequency. The noise robust semivariance measures are modifications of the two-scale realized variance 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 with 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 identify 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 with 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 the most liquid asset with a low level of microstructure noise. One might have expected standard volatility measures to deliver more accurate forecasts at the 300-second frequency, since microstructure noise should be small. However, this only holds true for SPY. For individual stocks, the forecasting gains are quite similar using noise-robust and standard volatility measures. 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.[3] Since the forecasting

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

performance of no single model dominate across sampling frequency and forecasting 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 with time-varying weights, generally results in significant out-of-sample forecasting performance (e.g. Aiolfi et al., 2011; Aiolfi and Timmermann, 2006; Elliott and Timmermann, 2016; Timmermann, 2006). These gains arise using both 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 longer horizons of 66 days / 3 months.

These results are in line with 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 the reaction to that news.[4] 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

[4]Similar qualitative conclusion 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).

Section 3. Noise-robust volatility measures are also discussed. Section 4 sets out the forecasting framework, including the extended HAR forecasting model 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 7. Volatility forecasting results using transaction-time sampled volatility measures are presented in Section 8. Finally, Section 9 summarizes the paper and presents our conclusion.

2 Theoretical Background

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

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t, \quad t \in [0, T] \tag{1}$$

where a is a locally bounded and predictable drift term, σ 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:

$$J_t^F := \int_0^t \int_{|x|>\varepsilon} x\mu(dx, ds), \tag{2}$$

$$J_t^I := \int_0^t \int_{|x|\leq\varepsilon} x(\mu(dx, ds) - \nu(dx)ds), \tag{3}$$

where μ is the jump measure of X with compensator ν, and $\varepsilon > 0$ is an arbitrary number. For more details on Itô-semimartingale 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 small jumps capture price moves which 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 QV of the price process is defined as:

$$QV_t = \underbrace{\int_0^t \sigma_s^2 ds}_{\text{Integrated Variation (IV)}} + \underbrace{\sum_{0<s\leq t} (\Delta_s X)^2}_{\text{Jump Contribution}} \tag{4}$$

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

$$RV_t = \sum_{i=1}^{n} (\Delta_i^n X)^2 \xrightarrow{p} QV_t, \tag{5}$$

where the day is split into $n = \lfloor 1/\Delta_n \rfloor$ equally spaced intervals of length Δ_n with 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 $\Delta_n \to 0$:

$$TBPV_t = \mu_1^{-2} \frac{n}{n-1} \sum_{i=2}^{n} |\Delta_i^n X| \mathbb{1}_{\{(\Delta_i^n X)^2 \leq \vartheta_i\}} |\Delta_{i-1}^n X| \mathbb{1}_{\{(\Delta_{i-1}^n X)^2 \leq \vartheta_{i-1}\}} \xrightarrow{p} \int_0^t \sigma_s^2 ds, \tag{6}$$

where $\mu_1 = \sqrt{2/\pi} \approx 0.7979$, $n/(n-1)$ is a small sample correction, and ϑ is the threshold estimator defined as in Corsi et al. (2010, appendix B).

Barndorff-Nielsen et al. (2010) introduced positive and negative realized semivariance

(RS) estimators to capture upside and downside risk:

$$RS_t^+ = \sum_{i=1}^{n} (\Delta_i^n X)^2 \mathbb{1}_{\{\Delta_i^n X>0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta_s X)^2 \mathbb{1}_{\{\Delta_s X_0\}} \tag{7}$$

$$RS_t^- = \sum_{i=1}^{n} (\Delta_i^n X)^2 \mathbb{1}_{\{\Delta_i^n X<0\}} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta_s X)^2 \mathbb{1}_{\{\Delta_s X<0\}}. \tag{8}$$

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. (2007b, ABD). 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 1 if max $\left(|\Delta_i^n X| \big/ \sqrt{\Delta_n TBPV} > \Phi_{1-\beta/2}\right)$ and 0 otherwise, where $\Phi_{(\cdot)}^{-1}$ is the inverse of the standard normal 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 components of QV are:

$$\widehat{C}_t = RV_t \cdot (1 - \mathcal{J}_t) + TBPV_t \cdot \mathcal{J}_t, \tag{9}$$

$$\widehat{J}_t = (RV_t - TBPV_t, 0)^+ \cdot \mathcal{J}_t. \tag{10}$$

To identify days with significant finite and 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 to eliminate the large jumps. The truncated realized power variation $B(p, \upsilon_n, \Delta_n)_t = \sum_{i=1}^n |\Delta_i^n X|^p \mathbb{1}_{\{|\Delta_i^n X \leq \upsilon_n\}}$, with $\upsilon_n = \varrho \Delta_n^{\varpi}$, $\varrho > 0$, $\varpi \in (0, 1/2)$, is the sum of truncated absolute returns, $|\Delta_i^n X| \leq \upsilon_n$, 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 $\left(SFA_t - k^{p/2-1}\right) \Big/ \sqrt{\widehat{V}_t} \xrightarrow{L} \mathcal{N}(0,1)$, where $\widehat{V}_t = N(p,k)\frac{B(2p,\upsilon_n,\Delta_n)_t}{B(p,\upsilon_n,\Delta_n)_t^2}$. For further details on $N(p,k)$, and other settings, see Aït-Sahalia and Jacod (2011). We set $k = 2$ and

$p = 4$, and use the indicator function $F_t = \mathbb{1}\left\{SFA_t < k^{p/2-1} - \Phi^{-1}_{1-\alpha}\sqrt{\widehat{V}_t}\right\}$ to identify days with finite activity jumps.

Our classification of jumps by sign and activity is described below.

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

We classify jumps by activity using the jump $\mathcal{J}_t$ and finite activity F_t indicators. The contribution of positive and negative jumps to overall QV are based on $(RS^+_t - \frac{1}{2}TBPV_t, 0)^+ \cdot \mathcal{J}_t$ and $(RS^-_t - \frac{1}{2}TBPV_t, 0)^+ \cdot \mathcal{J}_t$ respectively. When forecasting volatility using our extended HAR models, we use daily (net) signed jumps, $\widehat{SJ}_t$, the difference between the positive and negative measures (e.g. Patton and Sheppard, 2015). The corresponding positive and negative signed jumps are $\widehat{J}^+_t = \widehat{SJ} \cdot \mathcal{P}_t$ and $\widehat{J}^-_t = \widehat{SJ} \cdot (1 - \mathcal{P}_t)$ respectively, where $\mathcal{P}_t = \mathbb{1}\left\{\widehat{SJ}_t > 0\right\}$. Their finite/infinite counterparts are identified using the finite activity 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), 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).[5] This suggests that noise-robust 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-minute returns of a large portion of the 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]

$$Y_t = X_t + u_t, \tag{11}$$

where X_t is the process described in equation (1), u_t is an i.i.d. noise process with $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[u_t^2] = \omega^2$, and $u_t \perp\!\!\!\perp X_t$. Jacod et al. (2009) and Christensen et al. (2014) propose pre-averaging estimators for the RV and a consistent estimator of the IV. The pre-averaging returns are estimated as a weighted average of returns within a local neighborhood of L log-prices:

$$\Delta_i^n X^* = \sum_{j=1}^{L-1} g\left(\frac{j}{L}\right) \Delta_{i+j}^n Y, \tag{12}$$

where $g = \min(x, 1-x)$, and set $L = \theta\sqrt{n}$ with $\theta = 1/3$ for 5 and 60 seconds return or $\theta = 1$ for 300 seconds returns. With these choices, the noise-robust estimator for the

[5]The bias is due to $\mathbb{E}[|\Delta_i^n X|] \leq \mathbb{E}[|\Delta_i^n X + \eta_i|]$, where $\eta_i = u_i - u_{i-1}$, and its presence produces poor measures of the true volatility, as well as induces an attenuation bias in the autoregressive estimates (e.g. Bollerslev et al., 2016).

[6]The mechanics of trading generate a diverse array of market microstructure effects including 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 and most common market microstructure model.

realized variance and the bipower variation are:[7]

$$RV_t^* = \frac{M}{M-L+2}\frac{1}{L\psi_2^L}\sum_{i=0}^{M-L+1}|\Delta_i^n X^*|^2 - \frac{\psi_1^L\hat{\omega}_{AC}^2}{\theta^2\psi_2^L} \tag{13}$$

$$BPV_t^* = \frac{M}{M-2L+2}\frac{1}{L\psi_2^L\mu_1^2}\sum_{i=0}^{M-2L+1}|\Delta_i^n X^*||\Delta_{i+L}^n X^*| - \frac{\psi_1^L\hat{\omega}_{AC}^2}{\theta^2\psi_2^L}, \tag{14}$$

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

The ABD test in Andersen et al. (2007b) can be modified to yield a test that is robust to the presence of market microstructure noise. To do this we use the asymptotic distribution of pre-averaged returns (see, for instance Christensen et al., 2014; Jacod et al., 2009; Podolskij and Vetter, 2009, and the referencves therein):

$$n^{1/4}\Delta_i^n X^*\big|\mathcal{F}_{i/n} \sim \mathcal{N}\left(0, \frac{\theta\sigma^2}{12} + \frac{\omega^2}{\theta}\right). \tag{15}$$

Thus, we can define a threshold for identifying jumps as follows:

$$\tau = \frac{q_\beta}{n^{\varpi}}\sqrt{\psi_2^L\theta\hat{\sigma}^2 + \frac{\hat{\omega}^2}{\theta}}, \tag{16}$$

where $q_\beta = \Phi_{1-\beta/2}^{-1}$ is the inverse of the standard normal distribution, α is the significance level, and $\beta = 1-(1-\alpha)^{\Delta_n}$ is the Šidàk multiple testing correction. We use the BPV_t^* to estimate $\hat{\sigma}^2$ and $\hat{\omega}_{AC}^2$ to estimate $\hat{\omega}^2$. We set $\varpi = 1/4$ and $\theta = 1/3$. Therefore, we reject the null of no jumps whenever $\max\left(|\Delta_i^n X^*|\right) > \tau$.

Noise-robust versions of the realized semivariances, which capture upside and downside

[7]We also tried the threshold bipower variation measure proposed by Christensen et al. (2018) but the differences were negligible.

risk, are constructed by appropriately modifying the two-scale realized variance measure of Zhang et al. (2005):

$$TSRS_t^+ = \frac{1}{K}\sum_{k=1}^{K} RS_{t,k}^+ - \frac{n}{\bar{n}} RS_t^+ \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta_s X)^2 \mathbb{1}_{\{\Delta_s X>0\}}, \tag{17}$$

$$TSRS_t^- = \frac{1}{K}\sum_{k=1}^{K} RS_{t,k}^- - \frac{n}{\bar{n}} RS_t^- \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{0<s\leq t} (\Delta_s X)^2 \mathbb{1}_{\{\Delta_s X<0\}}, \tag{18}$$

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; $\bar{n} = \frac{n-K+1}{K}$ is the average number of observations in the subsamples; $K = \lfloor cn^{2/3} \rfloor$ and c is the optimal bandwidth as in Zhang et al. (2005). The two-time scale estimators average the realized semivariances over K subsamples, and apply a bias correction from the highest possible frequency.[8]

3.2 Noise-Robust ABD Test and Two-Time Scale Realized Semi-variance – 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, where the log-price X is simulated as:

$$\begin{aligned} dX_t &= \sqrt{\nu_t} dW_t + \theta_L dL_t \\ d\nu_t &= \kappa(\eta_\nu - \nu_t)dt + \gamma_\nu \nu_t^{1/2} dB_t, \end{aligned} \tag{19}$$

where W_t and B_t are standard Brownian motions 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 $\kappa = 5$, $\eta_\nu = 1/16$, $\rho = -0.5$. 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 $\lambda = 1.0$ such that there is on average one

[8] Aït-Sahalia et al. (2012) develop a noise-robust, pre-averaging, version of the Aït-Sahalia and Jacod (2009b) jump test, while Li and Xiu (2016) develop general GMM procedures that address measurement error in realized volatility measures.

jump every day. When jumps are of finite activity we set $\theta_L = 1$, 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, ξ is the noise-to-signal ratio used to simulate market microstructure noise, $u_{t,i} \sim \mathcal{N}(0, \omega_t^2)$ and $\omega_t^2 = \xi^2 \int_0^t \nu_s ds$. With this design, the variance of the noise is constant throughout the day, but changing from day to day.

The price process is simulated via an Euler scheme where we normalize one second to be $\Delta_n = 1/23,400$. Thus, the interval $[0, 1]$ contains the usual 6.5hrs of trading activity. To generate the observed prices, we discretize $[0, 1]$ into a number $n = 23,400$ of intervals. We then contaminate the prices with market microstructure noise and aggregate the data to the 5-, 60- and 300 seconds, which are equivalent to 4,680, 390 and 78 intraday observations per day. We simulate 5 trading days and use 5,000 replications.

Table 1 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. The standard ABD test is undersized (oversized) at higher (lower) frequencies, irrespective of the level of noise-to-signal ratio, whereas the noise-robust test displays very decent size levels which decreases with the sampling frequency. This result is expected as the level of microstructure noise decreases when the data is sampled more sparsely and therefore pre-averaged methods are less efficient.

The second and third panels show the power of the tests under finite and infinite activity jumps. When jumps are of finite activity and the noise-to-signal ratio is small, both tests perform quite well with the noise-robust (standard) test outperforming its standard (noise-robust) version at higher (lower) frequencies. Finally, when jumps are infinite activity, the standard ABD test is badly affected by the noise-to-signal levels.

Table 2 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 microstructure 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-time scale realized semivariance is very good overall.

4 Forecasting Models and Forecast Comparisons

The 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 extend the HAR-RV model further by adding signed, finite and infinite activity jumps. The benchmark HAR-RV model is

$$RV_{t,t+h} = \beta_0 + \beta_d RV_t + \beta_w RV_{t-5,t} + \beta_m RV_{t-22,t} + \epsilon_{t+h}, \tag{20}$$

where h is the forecast horizon, and $RV_{t,t+h-1} = \frac{1}{h}\sum_{i=1}^{h} RV_{t+1-i}$. We examine nine different, extended HAR models. The first three forecasting models include daily, weekly and monthly jumps in addition to the daily, weekly and monthly continuous component of RV. The next three models replace the jump variables in previous models with their finite activity counterparts. The final three models replace the jump part with their infinite activity jumps. We estimate separate models for unsigned, positive and negative jumps:

Jumps, Signed and Unsigned Models:

HAR-CJ: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d}\widehat{J}_t + \beta_{J_w}\widehat{J}_{t-5,t} + \beta_{J_m}\widehat{J}_{t-22,t} + \epsilon_{t,t+h}$

HAR-CJ$^+$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d^+}\widehat{J}_t^+ + \beta_{J_w^+}\widehat{J}_{t-5,t}^+ + \beta_{J_m^+}\widehat{J}_{t-22,t}^+ + \epsilon_{t,t+h}$

HAR-CJ$^-$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{J_d^-}\widehat{J}_t^- + \beta_{J_w^-}\widehat{J}_{t-5,t}^- + \beta_{J_m^-}\widehat{J}_{t-22,t}^- + \epsilon_{t,t+h}$

Finite Jumps, Signed and Unsigned Models:

HAR-CFJ: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d}\widehat{FJ}_t + \beta_{FJ_w}\widehat{FJ}_{t-5,t} + \beta_{FJ_m}\widehat{FJ}_{t-22,t} + \epsilon_{t,t+h}$

HAR-CFJ$^+$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d^+}\widehat{FJ}_t^+ + \beta_{FJ_w^+}\widehat{FJ}_{t-5,t}^+ + \beta_{FJ_m^+}\widehat{FJ}_{t-22,t}^+ + \epsilon_{t,t+h}$

HAR-CFJ$^-$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{FJ_d^-}\widehat{FJ}_t^- + \beta_{FJ_w^-}\widehat{FJ}_{t-5,t}^- + \beta_{FJ_m^-}\widehat{FJ}_{t-22,t}^- + \epsilon_{t,t+h}$

Infinite Jumps, signed and Unsigned Models:

HAR-CIJ: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d}\widehat{IJ}_t + \beta_{IJ_w}\widehat{IJ}_{t-5,t} + \beta_{IJ_m}\widehat{IJ}_{t-22,t} + \epsilon_{t,t+h}$

HAR-CIJ$^+$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d^+}\widehat{IJ}_t^+ + \beta_{IJ_w^+}\widehat{IJ}_{t-5,t}^+ + \beta_{IJ_m^+}\widehat{IJ}_{t-22,t}^+ + \epsilon_{t,t+h}$

HAR-CIJ$^-$: $RV_{t,t+h} = \beta_0 + \beta_{C_d}\widehat{C}_t + \beta_{C_w}\widehat{C}_{t-5,t} + \beta_{C_m}\widehat{C}_{t-22,t} + \beta_{IJ_d^-}\widehat{IJ}_t^- + \beta_{IJ_w^-}\widehat{IJ}_{t-5,t}^- + \beta_{IJ_m^-}\widehat{IJ}_{t-22,t}^- + \epsilon_{t,t+h}$

The realized continuous and jump measures in the models are estimated using the formulae outlined in Section 3. We also have an additional nine models where all the right-hand volatility measures are the noise-robust measures discussed in Section 3.1. Although additional variants of these models could be developed and evaluated, we do not believe that it is worthwhile doing so since model averages should encompass these variants.

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

$$MSPE = S_h^{-1} \sum_{s=1}^{S_h} \left(RV_s^h - \widehat{RV}_s^h \right)^2, \tag{21}$$

where RV_s^h and $\widehat{RV}_s^h$ are respectively the actual and pseudo out-of-sample forecasts of $RV_{t,t+h}$, 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, DM,hereafter) 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. We denote by $\mathcal{M}$ the set of all the HAR models. We define $d_{h,i,j} = L(RV_{t,t+h}, \widehat{RV}^{(i)}_{t,t+h}) - L(RV_{t,t+h}, \widehat{RV}^{(j)}_{t,t+h})$ as the difference in the loss of model i and model j. We use a quadratic loss function as L. Finally, we construct the average loss difference, $\bar{d}_{h,i,j}$, and define the test statistics as follows

$$t^h_{i,j} = \frac{\bar{d}_{h,i,j}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{h,i,j})}}, \quad \forall i, j \in \mathcal{M} \tag{22}$$

The MCS test statistics are given by $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{i,j}^h|$ and have the null hypothesis, H_0 that all models have the same expected loss. The alternative hypothesis 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 5000 bootstrap replications.[9]

5 Data

For our forecasting exercise, we use 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 4277 trading days. The 20 individual stocks were chosen based on their jump activity index, and the relative contributions of finite and infinite jumps. The data are sourced from the TickData database.[10] We follow Hansen and Lunde (2006) and use previous tick interpolations to aggregate the ticks to the required frequency.

Mean daily RV for SPY and the 20 stocks 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 sampling 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; Bandi and Russell, 2006; Zhang et al., 2005, inter alios for a more detailed discussion).

The contribution of the different types of jumps to total QV are shown in Table 3. The contribution of jumps decreases as the sampling interval increases from 5 to 300 seconds.

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

[10] 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 we are not producing data in real time, we have the capacity to look at the tick following a suspected bad tick before we decide whether or not the tick is valid. The filters are proprietary and are based upon recent tick volatility, moving standard deviation windows, and time day. For a more detailed explanation, see the high-frequency data filtering white paper on the TickData resources page TickData.

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 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 finite activity jumps. At this frequency, the small variations that characterize jumps are close to Brownian increments. We find little evidence of asymmetry in the shares of signed jumps. The Blumenthal-Getoor index or jump activity index ($\hat{\beta}_{IJA}$),[11] which measures the activity of small increments, 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 important at higher frequencies.

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 sub-prime crises (shaded areas), many other spikes in volatility are idiosyncratic. The 5- and 300-second autocorrelation functions of the SPY realized measures based on noise-robust and standard measures are displayed in Figure 2. The SPY RV_t and $\widehat{C}_t$ measures appear to be long memory processes since their autocorrelations do not decline exponentially. The ACF of the 5-second RV_t and $\widehat{C}_t$ 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.

[11] The jump activity index is estimated as in Jing et al. (2012).

6 Empirical Findings

6.1 SPY Forecasting Results

Since we use the HAR-RV model as a benchmark for assessing the forecasting performance of our extended HAR models, Table 4 sets out the in-sample coefficients, as well as the in- and out-of-sample R^2s and MSPEs, of the HAR-RV model for four forecast horizons – $h = 1$ (day), $h = 5$ (week), $h = 22$ (month), $h = 66$ (quarter), 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 5 ($h = 1$), 10 ($h = 5$), 44 ($h = 22$), and 132 ($h = 66$), since the random error term in the models is serially correlated at least up to order $h - 1$. In following Andersen et al. (1999) and Patton and Sheppard (2015), we estimate R^2_{oos} as 1 minus the ratio of the out-of-sample models-based MSPE to the out-of-sample MSPE from a forecast including only a constant. The MSPE results are based on a pseudo out-of-sample rolling regression forecast using a 1000 day window.

All the coefficients in Table 4 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, its relative importance increases at longer horizons.[12]

Summary forecasting results for extended HAR-CJ (jumps), HAR-CFJ (finite jumps), and the HAR-CIJ (infinite jumps) models are presented in Table 5, also using 300 second returns. In- and out-of-sample R^2s and the MSPEs are presented for unsigned jumps, positive signed jumps and negative signed jumps. Full results are available on request. A few points about the coefficients 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 (2015), overall jumps tend to reduce future volatility,

[12] These results are well-documented in the literature, see Andersen et al. (2007a), Corsi (2009), and Corsi et al. (2010) among others.

negative jumps tend to increase it and positive jumps to decrease it. Finite (infinite) jumps tend to decrease (increase) future volatility.

Unsurprisingly, the in-sample R-squared statistics (R^2_{is}) in Table 5 suggest that incorporating jumps as predictors results in a better fit for our models, outperforming the benchmark HAR-RV across the four forecasting horizons under examination. The out-of-sample R-squared statistics (R^2_{oos}) show that extended HAR models outperform the benchmark 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 one day and one week horizons, and better (significantly better) about half (one quarter) of the time at the one-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 further forecast gains. See Section 7 below.

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

We know that microstructure noise is important at higher frequencies, and the resulting attenuation bias may generate less accurate volatility forecasts than forecasts using noise-robust measures, such as the ones discussed in Section 3.1 above. We examined this issue in detail. Table 6 compares the forecasting performance of SPY extended HAR volatility models using standard versus noise-robust realized measures identifying models with significantly lower MSPEs than the benchmark 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 noise-robust measures. The entries are relative MSPEs –The ratio of the MSPE of the proposed model to the MSPE of the corresponding benchmark model– so ratios below one indicate more accurate rolling regression forecasts.[13] Models with significantly lower MSPE than the benchmark model, based on pair-wise Diebold and Mariano (1995, DM) tests, are starred. The DM tests show that many of the extended HAR models in Table 6 forecast as well as, or better, than the HAR-RV models, although there is considerable variation across sampling frequencies and time horizon.

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 standard volatility measures are excluded from the MCS at longer horizons, confirming the importance of taking account of microstructure noise at higher frequencies. Nevertheless, the MSPE numbers for the benchmark HAR-RV model in the final row of Table 6 suggest that models using 300-second volatility measures tend to give better forecast than models using 5- or 60-second returns, irrespective of whether standard or noise-robust volatility measures are used.

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

Some results for the 20 S&P 500 stocks are presented in Table 7. The relative MSPE entries (averaged across the 20 stocks) are shown in the body of the table, while the average MSPEs for the benchmark 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 with a dagger (†). The relative MSPE entries are more clustered around one than in Table 6.[14] In addition, with the majority of the models 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 the SPY. At the 5 and 60-second frequencies, the results show that noise-robust volatility measures work best. This is because noise-robust measures provide more

[13] The MSPE results are based on pseudo out-of-sample, rolling regression forecast using 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 a dagger (†). The MCS results are generated using a 10-day block bootstrap and 5,000 replications.

[14] The entries are also less dispersed, in part because we are reporting averages.

efficient estimators of the latent volatility process, thereby reducing the attenuation bias on the autoregressive coefficients (see, e.g. Andersen et al., 2005; Bollerslev et al., 2016). However, consistent with the results for SPY, forecasts using 300 second volatility measures are generally better than forecasts using 5 or 60 second-based volatility measures.[15] In addition, the relative MSPEs of the standard volatility measures are often lower than those of the noise-robust measures.

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.[16] We find that, on average, cojumps only contribute to 9% of the total jump component, which means that most jumps are idiosyncratic. To illustrate, the left panel of Figure 3 shows the returns on May 06, 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 23, 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, so generating jumps of different size and directions. Aït-Sahalia and Xiu (2016) suggest that co-jumps stem from surprising news announcements that occur primarily before the opening of the U.S. market. Amengual and Xiu (2018) note that downward intraday volatility jumps in the S&P 500 index are often associated with a resolution of policy uncertainty, mostly through statements from the FOMC meetings and speeches by the chair of the Federal Reserve. Aït-Sahalia et al. (2020) find that idiosyncratic jumps are related to idiosyncratic events such as earning disappointments. 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 averages forecasts consistently outperform the forecasts from the benchmark HAR-RV and the best extended HAR models across sampling frequencies and forecasting horizons.

[15]The improvements of the 300-second based realized measures vis-à-vis 5- and 60-second returns are due to noise-robust measures are sometimes derived under some (strong) assumptions about the microstructure noise, and whenever (some of) these assumptions are not met in practice, the estimators turn out to be inconsistent. Therefore, the 300-second returns offers enough statistical power that seems to avoid distortions that could arise from microstructure noise.

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

7 The Gains from Model Averaging

Hitherto, 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.[17] 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, e.g. Aiolfi et al., 2011; Aiolfi and Timmermann, 2006; Bates and Granger, 1969; Elliott and Timmermann, 2016, and the references therein).

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. In the first three cases, the weights are recalculated every time a new set of rolling forecasts are generated, and we prune the set of models under consideration by only averaging models that are retained in the model confidence set.

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

Note: $\widehat{\Sigma}_t^h$ is the estimated, rolling window variance-covariance matrix of the set of MCS retained horizon h volatility forecasting models at time t. ι is a vector of ones representing each retained model. $MSPE_{t,i}^h$ and $Rank_{t,i}^h$ are the rolling window MSPEs and MCS Ranks for the MCS retained horizon h forecasting model at time t. Finally, N represents all the jump specifications used in this study.

We present model averaging results for SPY and four individual stocks chosen by the

[17] 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.

level of their jump activity. All the stocks have estimated Blumenthal-Getoor index 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 are the extreme cases.

The relative MSPEs for the best extended HAR-RV model and the four model averaging approaches are shown in Table 8. The MSPEs for each index or stock and forecast horizon are measured relative to the MSPE of the corresponding 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 model averages with significantly lower MSPEs than the MSPEs of the HAR-RV models. Double starred entries identify models whose MSPEs are significantly lower than the MSPEs of both the benchmark HAR-RV and the best extended HAR model. The four model averages generate forecasts that typically outperform the benchmark model for the four forecast horizons examined: $h = 1$ (on-day), $h = 5$ (one week), $h = 22$ (one month), $h = 66$ (one quarter). 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 as compared with a range of 0.693 to 0.715 for the four model averages. The largest MSPE reductions are generally found at the one-week horizon, followed by the one-month horizon.

We also compare 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 averaged 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. Forecasting 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 of inverse MSPE ranks, work as well as schemes that are more complicated (e.g. Patton and Sheppard, 2009).

8 A Robustness Check 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 have considered 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 averaging frameworks.

For brevity, we only report results for SPY. The transaction-based volatility measures are calculated using a 78 intraday returns sampling scheme as in Patton and Sheppard (2015). This is the transaction-based equivalent of the 300-second/5-minute sampling scheme, which is widely used in the literature. Intraday returns are calculated by fixing the opening and closing prices, and recording the prices at business time $\lfloor ik \rfloor$, where $i = 1, \ldots, 79$, $k = \frac{N-1}{79}$, N is the number of unique date stamps per day, and $\lfloor . \rfloor$ denotes rounding down to the nearest integer.[18]

Table 9 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 14.3% for the

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

clock-based measures. Almost all 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).

The relative MSPEs in Table 10 suggest that the forecasting performance of extended HAR models using transaction-based measures is comparable to that of the benchmark 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 benchmark 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 the models is broadly similar, we cannot identify a set of superior models.

A comparison of clock- and transaction-time based SPY model averaging results is presented in Table 11. 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 those obtained with clock-based sampling, so the transaction-time based MSPEs are always higher than their close-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 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 style 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 stock 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 interval increases, 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 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 noise-robust 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 benchmark 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 the 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 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-à-vis the single best extended HAR model, which in turn outperforms the benchmark 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, M., Capistrán, C., and Timmermann, A. (2011). Forecasts combinations. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*. OUP.

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1-2):31–53.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics*, 74(3):487–528.

Aït-Sahalia, Y. and Jacod, J. (2009a). Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics*, 37(5A):2202–2244.

Aït-Sahalia, Y. and Jacod, J. (2009b). Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37(1):184–222.

Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether jumps have finite or infinite activity. *The Annals of Statistics*, pages 1689–1719.

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, 50(4):1007–50.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics.* Princeton University Press.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168(2):207–222.

Aït-Sahalia, Y., Kalnina, I., and Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics.*

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (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.

Aït-Sahalia, Y. and Xiu, D. (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics*, 194(2):205–219.

Aït-Sahalia, Y. and Xiu, D. (2019). A hausman test for the presence of market microstructure noise in high frequency data. *Journal of econometrics*, 211(1):176–205.

Amengual, D. and Xiu, D. (2018). Resolution of policy uncertainty and sudden declines in volatility. *Journal of Econometrics*, 203(2):297–315.

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

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (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.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bollerslev, T., and Dobrev, D. (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.

Andersen, T. G., Bollerslev, T., and Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.

Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692.

Barndorff-Nielsen, O., Kinnebrock, S., and Shephard, N. (2010). Volatility and time series econometrics: Essays in honor of Robert F. Engle, chapter measuring downside risk-realised semivariance.

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

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

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468.

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

Busch, T., Christensen, B. J., and Nielsen, M. Ø. (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.

Caporin, M., Kolokolov, A., and Renò, R. (2017). Systemic co-jumps. *Journal of Financial Economics*, 126(3):563–591.

Christensen, K., Hounyo, U., and Podolskij, M. (2018). Is the diurnal pattern sufficient to explain intraday variation in volatility? a nonparametric assessment. *Journal of Econometrics*.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Pirino, D., and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.

Dumitru, A.-M. and Urga, G. (2012). Identifying jumps in financial assets: a comparison between nonparametric jump tests. *Journal of Business & Economic Statistics*, 30(2):242–255.

Duong, D. and Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics*, 187(2):606–621.

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5(1):31–67.

Ghysels, E. and Sinko, A. (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics*, 160(1):257–271.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance*, 40:443–459.

Giot, P. and Laurent, S. (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.

Griffin, J. E. and Oomen, R. C. (2008). Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, 27(1-3):230–253.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

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

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

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

Jing, B.-Y., Kong, X.-B., Liu, Z., and Mykland, P. (2012). On the jump activity index for semimartingales. *Journal of Econometrics*, 166(2):213–223.

Lee, S. S. and Mykland, P. A. (2007). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563.

Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, 84(4):1613–1633.

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

Martens, M., Van Dijk, D., and De Pooter, M. (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.

Mukherjee, A., Peng, W., Swanson, N. R., and Yang, X. (2020). Financial econometrics and big data: A survey of volatility estimators and tests for the presence jumps and co-jumps. *Financial, Macro and Micro Econometrics Using R*, 42:1.

Oomen, R. C. A. (2006a). Comment on 2005 jbes invited address "realized variance and market microstructure noise" by peter r. hansen and asger lunde. *Journal of Business & Economic Statistics*, 24(2):195–202.

Oomen, R. C. A. (2006b). Properties of realized variance under alternative sampling schemes. *Journal of Business & Economic Statistics*, 24(2):219–237.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Patton, A. J. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.

Patton, A. J. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.

Podolskij, M. and Vetter, M. (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and their Applications*, 119(9):2803–2831.

Prokopczuk, M., Symeonidis, L., and Wese Simen, C. (2016). Do jumps matter for volatility forecasting? evidence from energy markets. *Journal of Futures Markets*, 36(8):758–792.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235(3):643–659.

Timmermann, A. (2006). Forecast combinations. In Elliot, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, pages 135–196. Elsevier.

Todorov, V. and Tauchen, G. (2010). Activity signature functions for high-frequency data analysis. *Journal of Econometrics*, 154(2):125–138.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (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.

A Tables and Figures

Table 1: Noise-Robust ABD Test – Size and Power Simulations

	$\xi = 0.01$			$\xi = 0.1$		
	5-Sec.	60-Sec.	300-Sec.	5-Sec.	60-Sec.	300-Sec.
	Size					
ABD Noise-robust	0.059	0.047	0.035	0.051	0.021	0.016
ABD	0.030	0.055	0.128	0.029	0.046	0.084
	Power – Compound Poisson (Finite Jumps)					
ABD Noise-robust	0.999	0.991	0.941	0.963	0.910	0.892
ABD	0.989	0.992	0.988	0.394	0.546	0.622
	Power – Cauchy Process (Infinite Jumps)					
ABD Noise-robust	0.956	0.815	0.746	0.910	0.717	0.546
ABD	0.736	0.770	0.768	0.482	0.572	0.616

Note: The table reports 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 laid out in Section 3.2 of the paper.

Table 2: Standard vs. Noise-Robust Realized Semivariances – Finite Sample MSE Performance

	$\xi = 0.01$			$\xi = 0.1$		
	5-Sec.	60-Sec.	300-Sec.	5-Sec.	60-Sec.	300-Sec.
RS^+	9.568	0.067	0.003	967.498	6.737	0.274
RS^-	9.589	0.069	0.004	968.441	6.801	0.287
$TSRS^+$	0.001	0.001	0.002	0.112	0.014	0.008
$TSRS^-$	0.001	0.001	0.002	0.113	0.016	0.009

Note: The table entries are the MSEs of the realized and two-scale realized semivariances in the simulation described in Section 3.2 of the paper. The DGP is a Heston model augmented with a finite activity, compound Poisson jumps. ξ represents the noise-to-signal ratio used to simulate the market microstrcture noise. Second-by-second prices were simulated 5,000 times for 5 days with 6.5 trading hours per day.

Table 3: Estimated Contribution of Signed, Finite and Infinite Activity Jumps to QV

	SPY			Avg. Stocks			AMZN	BA	BFB	CAT	CHL	COST	CVX
	5s	60s	300s	5s	60s	300s	300s	300s	300s	300s	300s	300s	300s
Continuous	56.798	88.474	85.725	32.399	65.612	70.198	73.426	72.586	55.143	74.899	62.182	69.525	80.277
Jumps	43.202	11.526	14.275	67.601	34.388	29.802	26.574	27.414	44.857	25.101	37.818	30.475	19.723
Pos. Jumps	21.847	6.450	8.257	33.946	16.535	14.992	15.208	14.362	22.474	12.574	17.978	15.963	9.849
Neg. Jumps	21.355	5.075	6.018	33.653	17.853	14.810	11.366	13.052	22.383	12.527	19.841	14.512	9.874
Finite Jumps	10.602	10.419	14.156	33.394	32.417	29.597	26.410	27.228	44.649	24.852	37.314	30.357	19.576
Infinite Jumps	32.600	1.106	0.118	34.207	1.971	0.205	0.165	0.187	0.208	0.249	0.504	0.118	0.147
Pos. Finite Jumps	5.584	5.941	8.219	17.028	15.539	14.883	15.127	14.248	22.380	12.465	17.681	15.892	9.766
Neg. Finite Jumps	5.017	4.478	5.937	16.366	16.878	14.714	11.283	12.979	22.269	12.387	19.633	14.465	9.810
Pos. Infinite Jumps	16.263	0.509	0.038	16.918	0.996	0.108	0.081	0.114	0.093	0.110	0.296	0.070	0.083
Neg. Infinite Jumps	16.338	0.597	0.080	17.287	0.975	0.096	0.084	0.073	0.115	0.140	0.208	0.047	0.064
$\hat{\beta}_{IJA}$	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

	DOW	EXC	GILD	GS	HD	JNJ	JPM	KO	OKE	PG	SO	UPS	WMT
	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s	300s
Continuous	68.881	69.488	63.203	75.979	73.935	70.611	76.122	74.208	59.168	71.147	70.791	68.292	74.102
Jumps	31.119	30.512	36.797	24.021	26.065	29.389	23.878	25.792	40.832	28.853	29.209	31.708	25.898
Pos. Jumps	15.029	15.506	18.911	12.311	13.875	12.919	12.926	12.498	19.059	15.416	14.486	15.477	13.013
Neg. Jumps	16.090	15.006	17.886	11.710	12.190	16.470	10.952	13.294	21.773	13.438	14.723	16.231	12.885
Finite Jumps	30.849	30.400	36.458	23.941	25.940	29.279	23.822	25.519	40.602	28.777	28.642	31.527	25.802
Infinite Jumps	0.270	0.112	0.339	0.080	0.125	0.111	0.056	0.273	0.230	0.076	0.568	0.181	0.096
Pos. Finite Jumps	14.830	15.434	18.670	12.297	13.843	12.832	12.899	12.341	18.982	15.365	14.274	15.373	12.968
Neg. Finite Jumps	16.019	14.966	17.788	11.644	12.097	16.447	10.923	13.178	21.620	13.413	14.368	16.154	12.834
Pos. Infinite Jumps	0.198	0.072	0.241	0.014	0.032	0.088	0.028	0.157	0.077	0.051	0.213	0.104	0.045
Neg. Infinite Jumps	0.071	0.040	0.098	0.066	0.093	0.023	0.029	0.116	0.153	0.025	0.355	0.077	0.051
$\hat{\beta}_{IJA}$	0.579	0.725	0.522	0.610	0.665	0.971	0.606	0.913	0.645	0.955	0.878	0.895	0.824

Note: The table reports the estimated percentage contribution of the different jump measures to QV. Results using 5-, 60-, 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 (see, Jing et al., 2012, for more details and settings).

Table 4: HAR-RV Benchmark – SPY, 300 Second Returns

	$h=1$	$h=5$	$h=22$	$h=66$
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

Note: The table reports the OLS coefficient estimates and in- and out-of-sample R-squared for HAR-RV forecasting regressions for SPY RV at the 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% or 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 5: SPY Extended HAR Regressions Using Total, Positive and Negative Signed Jumps

	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
	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
	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***	**0.857***	**1.243**	**2.721***	**1.112***	0.994	1.317
	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**
$R^2_{(oos)}$	**0.511**	**0.709**	0.644	0.452	**0.509**	**0.711**	0.652	**0.475**	**0.512**	**0.712**	0.651	0.454
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

Note: See Notes to Table 4. Bold in-sample and out-of-sample R-squared entries indicate that the fit of the proposed models is better than that of the benchmark 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 6: SPY Relative MSPEs by Frequency – Standard vs. Noise-Robust Measures

	$h = 1$ (day)			$h = 5$ (week)			$h = 22$ (month)			$h = 66$ (quarter)		
	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
						Panel A: Standard Jump Measures						
HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000†	1.000†	1.000	1.000	1.000†	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	1.167†	1.010
HAR-CFJ	0.871*	0.752*	0.910*	1.181	0.992	0.770*	1.051	1.178	0.926*	1.010†	1.171†	1.011
HAR-CIJ	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-CJ$^+$	0.903*	0.993	0.986	1.165	0.969	0.753*	1.147†	0.894*	0.891*	1.074†	0.977	0.965*
HAR-CJ$^-$	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*	1.198†	0.877*	0.908*	1.096†	0.959	0.985
HAR-CFJ$^-$	0.915*	0.969	0.877*	1.215	1.035	0.841*	0.982	0.959*	1.054	1.035†	1.020	1.044
HAR-CIJ$^+$	0.910*	1.055	0.881*	1.151	1.020	0.884*	1.086†	0.964*	1.192	1.136†	0.940*	1.002
HAR-CIJ$^-$	0.729*	1.059	0.875*	0.996	1.030	0.879*	1.054†	0.921*	1.189	0.939*	0.977*	1.029
						Panel B: 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-CFJ	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-CIJ	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-CJ$^+$	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-CJ$^-$	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-CFJ$^-$	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

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

Table 7: Twenty Stock averages of Relative MSPEs – Standard vs. Noise-Robust Measures

	$h = 1$ (daily)			$h = 5$ (week)			$h = 22$ (month)			$h = 66$ (quarter)		
	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.	5 Sec.	60 Sec.	300 Sec.
					Panel A – Standard Jump Measures							
HAR-RV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000†	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-FJ	1.057	0.984	0.973	1.048†	0.916	0.934	1.064†	0.943	0.974	1.043	0.952	0.997
HAR-IJ	1.010	0.973	0.940	0.986	0.955	0.942	1.035	1.010	1.063	1.007†	1.014†	1.037
HAR-CJ$^+$	1.044	1.000	0.968	1.098†	0.939	0.945	1.203†	0.994	1.038	1.127†	1.004†	1.033
HAR-CJ$^-$	1.063	1.018	0.932	1.038	0.943	0.934	1.144†	0.970	1.026	1.078†	0.997†	1.018
HAR-CFJ$^+$	1.055†	0.999	0.969	1.153†	0.940	0.945	1.267†	0.994	1.038	1.173†	1.004	1.031
HAR-CFJ$^-$	1.103†	0.984	0.932	1.115†	0.938	0.937	1.228†	0.970	1.030	1.144†	0.997†	1.016
HAR-CIJ$^+$	1.044	0.979	0.939	1.090†	0.966	0.946	1.189†	1.010	1.080	1.129†	1.004†	1.042
HAR-CIJ$^-$	1.011	0.982	0.947	1.071†	0.960	0.945	1.213†	1.005	1.091	1.137†	1.006†	1.062
					Panel B – Noise-Robust Jump Measures							
HAR-RV	0.966	0.916	0.969	0.975	1.017	0.998	0.975	1.081	1.138	0.956†	1.050	1.032†
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-FJ	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-IJ	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-CJ$^+$	0.955	0.986	0.978	0.962	1.008	0.991	0.981	1.082	1.092	0.956	1.042	1.017†
HAR-CJ$^-$	0.973	0.943	0.961	0.950	0.984	0.994	0.938	1.043	1.126	0.936†	1.030	1.019
HAR-CFJ$^+$	0.947	0.987	0.980	0.952	1.010	0.993	0.967	1.086	1.091	0.924†	1.044	1.014
HAR-CFJ$^-$	0.963	0.938	0.962	0.962	0.984	0.994	0.948	1.047	1.107	0.945†	1.031	1.024
HAR-CIJ$^+$	0.972	0.926	0.950	0.957	1.022	0.994	0.966	1.073	1.091	0.952†	1.045	1.008
HAR-CIJ$^-$	0.964	0.935	0.948	0.948	1.025	0.986	0.969	1.061	1.116	0.943†	1.037	1.033
Memo:												
HAR-RV MSPE	373.1364	54.8865	22.7444	85.5808	16.9684	9.9258	27.0842	8.8123	6.3931	17.2674	7.9011	6.2917

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

Table 8: Model Averaging Results – Relative MSPEs at Different Horizons for SPY, BA, BFB, COST and KO

	$h=1$	$h=5$	$h=22$	$h=66$	$h=1$	$h=5$	$h=22$	$h=66$
	SPY – 300 seconds				SPY – 60 seconds			
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. – MSPE Weights	0.879*	**0.706****	**0.862****	**0.919****	0.875*	**0.914****	**0.850***	0.965*
Avg. – 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			
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. – MSPE Weights	**0.972***	**0.906***	**0.915****	0.959*	0.926*	**0.823***	**0.814***	**0.856****
Avg. – 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			
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. – MSPE Weights	0.962*	**0.871***	**0.920***	0.958*	0.817*	0.713*	0.888*	0.975*
Avg. – 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*

Note: The table reports the relative MSPE, the ratio of MSPE of the model indicated in the first column to the MSPE of the benchmark HAR-RV, in both cases using standard volatility measures as opposed to noise-robust measures. The best models refers to the min. MSPE model from the set of extended HAR models presented in Section 4. The bold entries are model averages with lower MSPEs than the MSPEs of both the HAR-RV and the best extended HAR models. The starred entries denote model averages with significantly lower MSPEs than the benchmark HAR-RV models, whereas doubled starred (superscript **) entries identify models whose MSPEs are significantly lower than the MSPEs of both the benchmark HAR-RV and the best extended HAR model.

Table 9: Estimated Contribution of Jumps to QV – Comparison of Clock and Transaction Time Sampling Results

	Clock Time Sampling	Transaction Time 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

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

Table 10: SPY Volatility Forecasting Performance – Transaction-Based Sampling Results

	$h = 1$ (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-CFJ	0.973*	1.114	1.030	1.022
HAR-CIJ	0.981	0.999	1.061	1.017
HAR-CJ^{+}	1.037	1.119	0.956*	0.971*
HAR-CJ^{-}	0.990	1.003	1.036	1.012
HAR-CFJ^{+}	1.037	1.119	0.956*	0.971*
HAR-CFJ^{-}	0.990	1.003	1.036	1.012
HAR-CIJ^{+}	0.981*	0.996	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

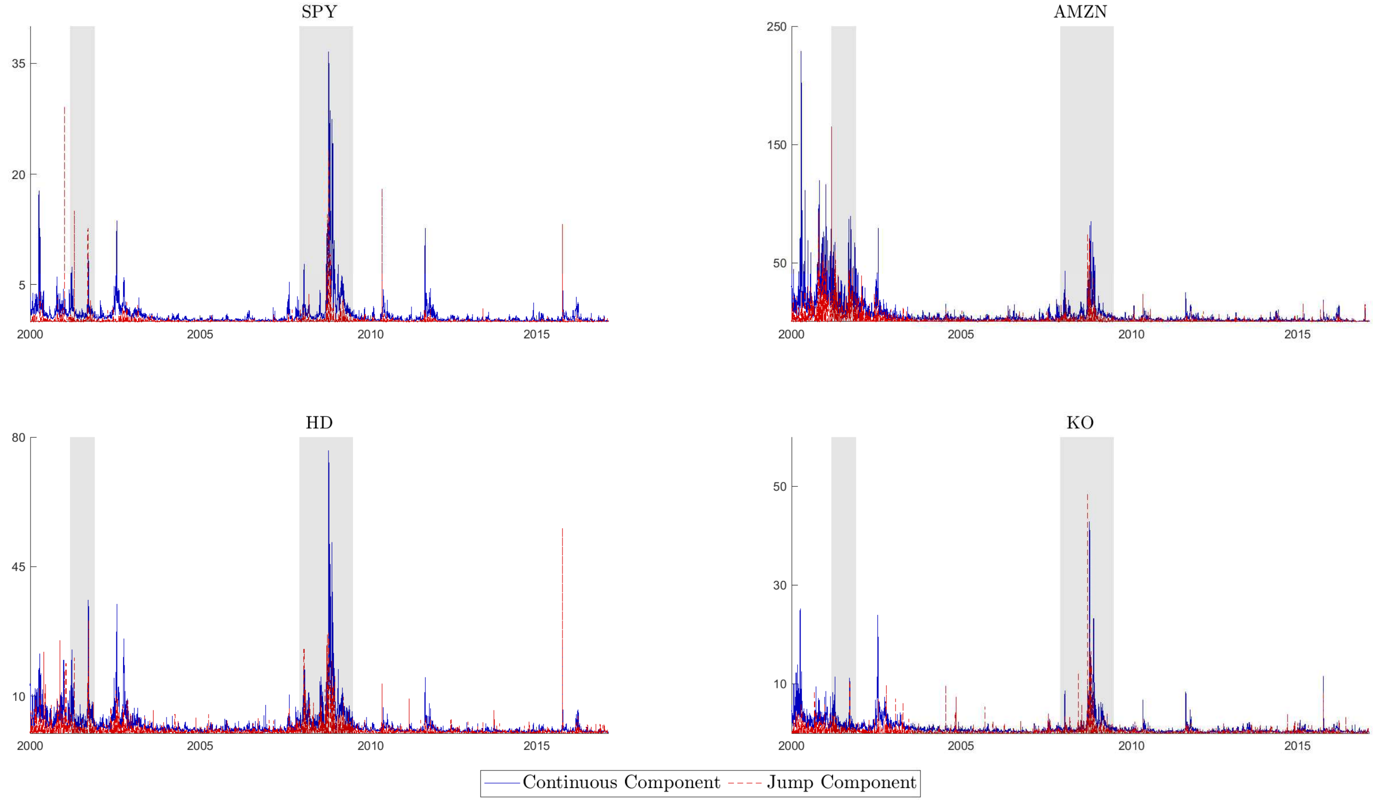
Note: The Table reports the relative MSPE of the extended HAR SPY volatility forecasting models at different horizons. The relative MSPEs are the ratio of the MSPEs of the extended HAR models relative to the benchmark HAR-RV model. The starred entries indicate statistically significant reductions in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

Table 11: SPY Model averaging Relative MSPEs – Comparison of Clock and Transaction-Based Sampling Results

	300 second, Clock-Based Sampling				Transaction-Based Sampling			
	$h = 1$	$h = 5$	$h = 22$	$h = 66$	$h = 1$	$h = 5$	$h = 22$	$h = 66$
HAR-RV benchmark	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
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. – MSPE Weights	0.879*	**0.706****	**0.862****	**0.919****	**0.926****	**0.950****	**0.889****	**0.961***
Avg. – 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

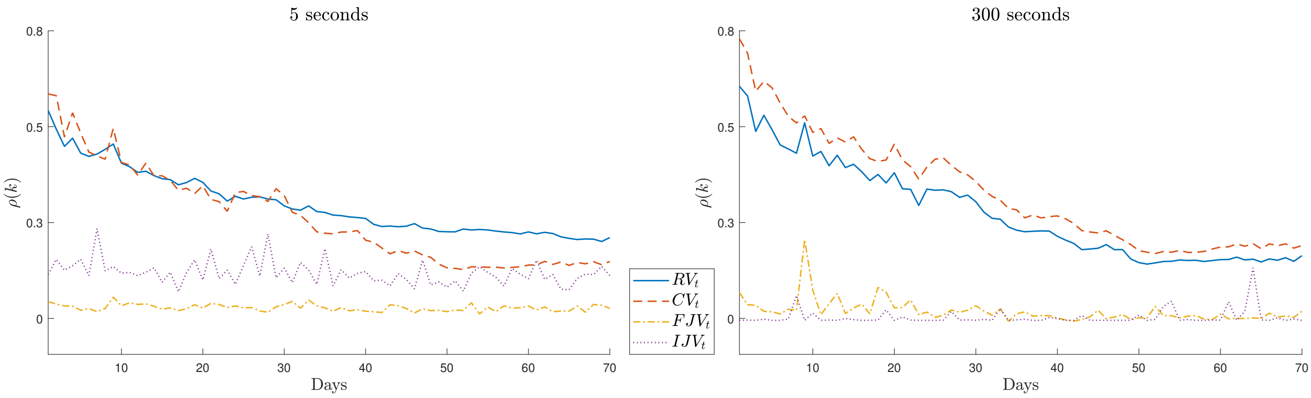
Note: The table compares the forecasting performance of the extended HAR SPY volatility forecasting models at different horizons h using clock and transaction based realized measures. The clock-based results use 300 second returns. The relative MSPEs are the ratio of the MSPEs of the models indicated in the first column to the MSPE of the benchmark HAR-RV model. The bold entries are models averages with lower MSPEs than the MSPEs of both the HAR-RV and the best extended model. The starred entries denote model averages with significantly lower MSPEs than the benchmark HAR-RV models, whereas doubles starred (**) entries identify models whose MSPEs are significantly lower then the MSPEs of both the benchmark HAR-RV and the best extended HAR model.

Figure 1: Time Series of Realized Volatility – Jump and Continuous Components



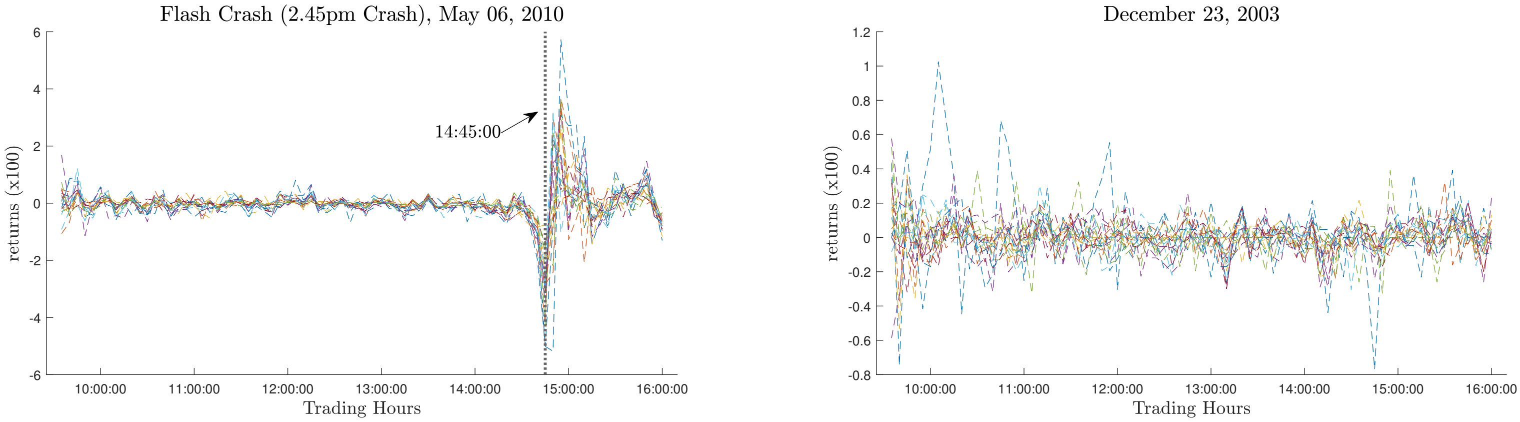
Note: This figure depicts the elements of the realized volatility for SPY and three individual stocks estimated at the 300 second frequency. The three individual stocks have the largest, smalles and average RV. NBER dated U.S. recession are shaded grey.

Figure 2: Autocorrelation Function of SPY Realized Measures



Note: The figure graphs the autocorrelation of the realized variance and its elements. The autocorrelations at the 5 and 300 second frequencies were estimated using noise-robust and raw estimators, respectively.

Figure 3: Systematic versus Idiosyncratic Jumps



Note: The figure depicts in two plots the intraday returns of the 20 individual stocks across two different trading days. The left plot displays the behavior of the stocks during the Flash Crash of May 06, 2010, where all the stocks jump together, whereas the right panel show a normal day on December 23, 2003, where all jumps are idiosyncratic.