Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility*

Andrew J. Patton
Department of Economics
Duke University and
Oxford-Man Institute of Quantitative Finance
andrew.patton@duke.edu

Kevin Sheppard†
Department of Economics and
Oxford-Man Institute of Quantitative Finance
University of Oxford
kevin.sheppard@economics.ox.ac.uk

First Version: October 7, 2011
This Version: November 16, 2013

Abstract

Using recently proposed estimators of the variation of positive and negative returns ("realized semi-variances"), and high frequency data for the S&P 500 index and 105 individual stocks, this paper sheds new light on the predictability of equity price volatility. We show that future volatility is much more strongly related to the volatility of past negative returns than to that of positive returns, and this effect is stronger than that implied by standard asymmetric GARCH models. We also find that the impact of a jump on future volatility critically depends on the sign of the jump, with negative (positive) jumps in prices leading to significantly higher (lower) future volatility. We show that models exploiting these findings lead to significantly better out-of-sample forecast performance for forecast horizons ranging from 1 day to 3 months.

Keywords: Realized variance, semivariance, volatility forecasting, jumps, leverage effect

J.E.L. Codes: C58, C22, C53

*We thank Giampiero Gallo, Neil Shephard, and seminar participants at Cass Business School, CORE, Oxford, Pennsylvania, EC^2 in Aarhus, Society for Financial Econometrics in Melbourne, World Congress of the Econometric Society in Shanghai for helpful comments. Additional results and analyses are presented in an internet appendix to this paper, available at www.kevinsheppard.com.

†Contact author. Code used in this paper for computing realized quantities is available at www.kevinsheppard.com.

1 Introduction

The development of estimators of volatility based on high frequency (intra-daily) information has lead to great improvements in our ability to measure financial market volatility. Recent work in this area has yielded estimators that are robust to market microstructure effects, feasible in multivariate applications, and which can separate the volatility contributions of jumps from continuous changes in asset prices[1], see Andersen, Bollerslev, and Diebold (2009) for a recent survey of this growing literature. A key application of these new estimators of volatility is in forecasting: better measures of volatility enable us to better gauge the current level of volatility and to better understand its dynamics, both of which lead to better forecasts of future volatility. Volatility forecasting, while long useful in risk management, has recently become increasing important as volatility is now directly tradable using swaps and futures.[2]

This paper uses high frequency data to shed light on another key aspect of asset returns: the "leverage effect," and the impact of signed returns on future volatility more generally. The observation that negative equity returns lead to higher future volatility than positive returns is a well-established empirical regularity in the ARCH literature[3], see the review articles by Bollerslev, Engle, and Nelson (1994) and Andersen, Bollerselv, Christoffersen, and Diebold (2006) for example. Recent work in this literature has also found evidence of this relationship using high frequency returns, see Bollerslev, Litvinova, and Tauchen (2006), Barndorff-Nielsen, Kinnebrock, and Shephard (2010), Visser (2008) and Chen and Ghysels (2011). We build on these papers to exploit this relationship and obtain improved volatility forecasts.

We use a new estimator proposed by Barndorff-Nielsen, Kinnebrock, and Shephard (2010) called "realized semivariance," which decomposes the usual realized variance into a component that relates only to positive high frequency returns and a component that relates only to negative high frequency returns.[4]

[1] See Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2006), Zhang, Mykland, and Aït-Sahalia (2005), Aït-Sahalia, Mykland, and Zhang (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), amongst others.

[2] A partial list of papers on this topic includes Andersen, Bollerslev, Diebold, and Labys (2000),Andersen, Bollerslev, Diebold, and Labys (2003),Fleming, Kirby, and Ostdiek (2003), Corsi (2009), Liu and Maheu (2005), Lanne (2006), Lanne (2007), Chiriac and Voev (2007), Andersen, Bollerslev, and Diebold (2007), Visser (2008) and Chen and Ghysels (2011).

[3] Common ARCH models with a leverage effect include GJR-GARCH (Glosten, Jagannathan, and Runkle, 1993), TARCH (Zakoian, 1994), and EGARCH (Nelson, 1991).

[4] Semivariance, and the broader class of downside risk measures, has a long history in finance. Applications of semivariance in finance include Hogan and Warren (1974) who study semivariance in a general equilibrium framework, Lewis (1990) who examined its role in option performance, and Ang, Chen, and Xing (2006) who examined the role of semivariance and covariance in asset pricing. For more on semivariance and related measures, see Sortino and Satchell (2001).

Previous studies have almost exclusively employed even functions of high frequency returns (squares, absolute values, etc.) which of course eliminate any information that may be contained in the sign of these returns. High frequency returns are generally small, and it might reasonably be thought that there is little information to be gleaned from whether they happen to lie above or below zero. Using a simple autoregressive model, as in Corsi (2009) and Andersen, Bollerslev, and Diebold (2007), and high frequency data on the S&P 500 index and 105 of its constituent firms over the period 1997-2008, we show that this is far from true.

We present several novel findings about the volatility of equity returns. Firstly, we find that negative realized semivariance is much more important for future volatility than positive realized semivariance, and disentangling the effects of these two components significantly improves forecasts of future volatility. This is true whether the measure of future volatility is realized variance, bipower variation, negative realized semivariance or positive realized semivariance. Moreover, it is true for horizons ranging from one day to three months, both in-sample and (pseudo-)out-of-sample. Second, we use realized semivariances to obtain a measure of *signed* jump variation and we find that is important for predicting future volatility, with volatility attributable to negative jumps leading to significantly higher future volatility, and positive jumps leading to significantly *lower* volatility. Thus, while jumps of both signs are indicative of volatility, their impacts on current returns and on future volatility might lead one to label them "good volatility" and "bad volatility." Previous research, see Andersen, Bollerslev, and Diebold (2007), Forsberg and Ghysels (2007) and Busch, Christensen, and Nielsen (2011), reported that jumps were of only limited value for forecasting future volatility. Our finding that the impact of jumps depends critically upon the sign of the jump helps explain these results: averaging across both positive and negative jump variation the impact on future volatility is near zero.[5]

Bollerslev, Litvinova, and Tauchen (2006) were perhaps the first to note that the sign of high frequency returns contains useful information for future volatility, even several days into the future. They show that several standard stochastic volatility models are unable to match this feature. Chen and Ghysels (2011) propose a semiparametric model for aggregated volatility (e.g., daily or monthly) as a function of individ-

[5] Corsi, Pirino, and Renò (2010) find that jumps have a significant and positive impact on future volatility, when measured using a new threshold-type estimator for the integrated variance.

ual high frequency returns. The coefficient on lagged high frequency returns is the product of a parametric function of the lag (related to the MIDAS model of Ghysels, Santa-Clara, and Valkanov (2006)) and a nonparametric function of the return. With this model, the authors obtain nonparametric "news impact curves," and document evidence that these curves are asymmetric for returns on the S&P 500 and Dow Jones indices. A forecasting model based on realized semivariances avoids some of the difficulties of the semiparametric MIDAS model of Chen and Ghysels (2011), such as the fact that estimation of news impact curves requires either a local estimator of spot volatility (a difficult empirical problem) or a method for dealing with the persistence in large returns, which makes estimation of the curve for larger values difficult. Realized semivariances are simple daily statistics and require no choice of bandwidth or other smoothing parameters, and no nonlinear estimation.

We complement and extend existing work in a number of directions. First, we look at the leverage effect and forecasting for a large set of assets – 105 individual firms, and the S&P 500, the FTSE 100 and the EURO STOXX 50 indexes – and verify that the usefulness of realized semivariances relative to realized variances is not restricted only to broad stock indices. Second, we show that negative semivariances are useful for predicting a variety of different measures of volatility: realized volatility, bipower variation, and both realized semivariances. Third, we show the usefulness of simple autoregressive models that we use, all of which can be estimated using least squares, across horizons ranging from one day to three months. We also present results on the information in signed jump variation, a measure that does not fit into existing frameworks, and which helps us reconcile our findings with the existing literature.

The remainder of the paper is organized as follows. Section 2 describes the volatility estimators that we use in our empirical analysis. Section 3 discusses the high frequency data that we study, and introduces the models that we employ. Section 4 presents empirical results on the gains from using realized semivariances for forecasting, and Section 5 presents results from using signed jump variation for volatility forecasting. Section 6 presents results for a pseudo-out-of-sample forecasting application for the U.S. data, and results for two international stock indexes. Section 7 concludes. An internet appendix contains additional results and analyses.

2 Decomposing realized variance using signed returns

In this section we briefly describe the estimators that are used in our analysis, including the new estimators proposed by Barndorff-Nielsen, Kinnebrock, and Shephard (2010).

Consider a continuous-time stochastic process for log-prices, p_t, which consists of a continuous component and a pure jump component:

$$p_t = \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s + J_t, \tag{1}$$

where μ is a locally bounded predictable drift process, σ is a strictly positive cádlág process and J is a pure jump process. The quadratic variation of this process is:

$$[p,p] = \int_0^t \sigma_s^2 \mathrm{d}s + \sum_{0<s\leq t} (\Delta p_s)^2, \tag{2}$$

where $\Delta p_s = p_s - p_{s-}$ captures a jump, if present.

Andersen, Bollerslev, Diebold, and Labys (2001) introduced a natural estimator for the quadratic variation of a process as the sum of frequently sampled squared returns which is commonly known as realized variance (RV). For simplicity, suppose that prices $p_0, \ldots, p_n$ are observed at $n+1$ times, equally spaced on $[0, t]$. Using these returns, the n-sample realized variance, RV, is defined below, and can be shown to converge in probability to the quadratic variation as the time interval between observations becomes small (Andersen, Bollerslev, Diebold, and Labys, 2003).

$$RV = \sum_{i=1}^{n} r_i^2 \xrightarrow{p} [p,p], \ as \ n \to \infty, \tag{3}$$

where $r_i = p_i - p_{i-1}$. Barndorff-Nielsen and Shephard (2006) extended the study of estimating volatility from simple estimators of the quadratic variation to a broader class which includes bipower variation (BV). Unlike realized variance, the probability limit of BV only includes the component of quadratic variation due to the continuous part of the price process, the integrated variance.

$$BV = \mu_1^{-2} \sum_{i=2}^{n} |r_i|\,|r_{i-1}| \xrightarrow{p} \int_0^t \sigma_s^2 \mathrm{d}s, \; as \; n \to \infty, \tag{4}$$

where $\mu_1 = \sqrt{2/\pi}$. The difference of the above two estimators of price variability can be used to consistently estimate the variation due to jumps of quadratic variation:

$$RV - BV \xrightarrow{p} \sum_{0 \le s \le t} \Delta p_s^2. \tag{5}$$

Barndorff-Nielsen, Kinnebrock, and Shephard (2010) recently introduced new estimators which can capture the variation only due to negative or positive returns using an estimator named "realized semivariance." These estimators are defined as

$$\begin{aligned} RS^- &= \sum_{i=1}^{n} r_i^2 I_{[r_i<0]} \\ RS^+ &= \sum_{i=1}^{n} r_i^2 I_{[r_i>0]} \end{aligned} \tag{6}$$

These estimators provide a complete decomposition of RV, in that $RV = RS^+ + RS^-$. This decomposition holds exactly for any n, as well as in the limit. We use this decomposition of realized volatility extensively in our empirical analysis below.[6]

Barndorff-Nielsen, Kinnebrock, and Shephard (2010) show that, like realized variance, the limiting behavior of realized semivariance includes variation due to both the continuous part of the price process a well as the jump component. The use of the indicator function allows the *signed* jumps to be extracted, with each of the realized semivariances converging to one-half of the integrated variance plus the sum of

[6] Visser (2008) considers a similar estimator based on powers of absolute values of returns rather than squared returns. For one-step forecasts of the daily volatility of the S&P 500 index, he finds that using absolute returns (i.e., a power of 1) leads to the best in-sample fit. We leave the consideration of different powers for future research and focus on simple realized semivariances.

squared jumps with a negative/positive sign:

$$RS^{+} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 \mathrm{d}s + \sum_{0\le s\le t} \Delta p_s^2 I_{[\Delta p_s>0]} \tag{7}$$
$$RS^{-} \xrightarrow{p} \frac{1}{2}\int_0^t \sigma_s^2 \mathrm{d}s + \sum_{0\le s\le t} \Delta p_s^2 I_{[\Delta p_s<0]}$$

Note that the first term in the limit of both RS^+ and RS^- is one-half of the integrated variance. This has two implications. First, it reveals that a "complete" decomposition of realized variance into continuous and jump components, and positive and negative components, yields only three, not four, terms; the continuous component of volatility is not decomposable into positive and negative components. Second, it reveals that the variation due to the continuous component can be removed by simply subtracting one RS from the other, without the need to estimate it separately. The remaining part is what we define as the *signed jump variation:*

$$\begin{aligned} \Delta J^2 &\equiv RS^+ - RS^- \\ &\xrightarrow{p} \sum_{0\le s\le t} \Delta p_s^2 I_{[\Delta p_s>0]} - \sum_{0\le s\le t} \Delta p_s^2 I_{[\Delta p_s<0]}. \end{aligned} \tag{8}$$

In our analysis below we use RS^+, RS^- and ΔJ^2 to gain new insights into the empirical behavior of volatility as it relates to signed returns.

3 Data and Models

The data used in this paper consists of high-frequency transaction prices on all stocks that were ever a constituent of the S&P 100 index between June 23, 1997 and July 31, 2008. The start date corresponds to the first day that U.S. equities traded with a spread less than $\frac{1}{8}$ of a dollar.[7] We also study the S&P 500 index exchange traded fund (ETF), with ticker symbol SPDR, over this same period for comparison. Of the total of 154 distinct constituents of the S&P 100 index over this time period, we retain for our analysis the 105

[7]Trading volume and the magnitude of microstructure noise that affects realized-type estimators both changed around this date, see (Aït-Sahalia and Yu, 2009), and so we start our sample after this change took place.

that were continuously available for at least four years.

All prices are taken from the New York Stock Exchange's TAQ database. Data are filtered to include only those occurring between 9:30:00 and 16:00:00 (inclusive) and are cleaned according to the rules detailed in Appendix A. As we focus on price volatility over the trade day, overnight returns are excluded, and we avoid the need to adjust prices for splits or dividends.

3.1 Business Time Sampling and Sub-Sampling

All estimators were computed daily, using returns sampled in "business time" rather than the more familiar calendar time sampling. That is, rather than use prices that are evenly spaced in calendar time (say, every five minutes) we use prices that are evenly spaced in "event" time (say, every ten transactions). (This implies, of course, that we sample more often during periods with greater activity, and less often in quieter periods.) Under some conditions business-time sampling can be shown to produce realized measures with superior statistical properties, see Oomen (2005), and this sampling scheme is now common in this literature, see Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Bollerslev and Todorov (2011) for example.[8]

We sample prices 79 times per day, which corresponds to an average interval of 5 minutes. We use the first and last prices of the day as our first and last observations, and sample evenly across the intervening prices to obtain the remaining 77 observations. The choice to sample prices using an approximate 5-minute window is a standard one, and is motivated by the desire to avoid bid-ask bounce type microstructure noise.

Since price observations are available more often than our approximate 5-minute sampling period, there are many possible "grids" of approximate 5-minute prices that could be used, depending on which observation is used for the first sample. We use 10 different grids of 5-minute prices to obtain 10 different estimators, which are correlated but not identical, and then average these to obtain our final estimator. This approach is known as "sub-sampling" and was first proposed by Zhang, Mykland, and Aït-Sahalia

[8]Recent work by Li, Mykland, Renault, Zhang, and Zheng (2013) considers cases where trade arrivals are strongly related to volatility, and shows that bias in realized variance can arise in such cases. That paper does not consider realized semivariance, and we assume that our data fits into the usual framework where no such biases arise.

(2005). This procedure should produce a mild increase in precision relative to using a single estimator.

3.2 Volatility Estimator Implementation

Denote the observed log-prices on a given trade day as $p_0, p_1, \ldots, p_n$ where $n+1$ is the number of unique time stamps between 9:30:00 and 16:00:00 that have prices. Setting the number of price samples to 79 (which corresponds to sampling every 5 minutes on average), RV computed uniformly in business time starting from the j^{th} observation equals

$$RV^{(j)} = \sum_{i=1}^{78} \left(p_{\lfloor ik+j\delta \rfloor} - p_{\lfloor (i-1)k+j\delta \rfloor}\right)^2 \tag{9}$$

where $k = n/78$, $\delta = n/78 \times 1/10$ and $\lfloor \cdot \rfloor$ rounds down to the next integer. Prices outside of the trading day are set to the close price. The sub-sampled version is computed by averaging over 10 uniformly spaced windows,

$$RV = \frac{1}{10} \sum_{j=0}^{9} RV^{(j)} \tag{10}$$

Realized semivariances, RS^+ and RS^-, are constructed in an analogous manner.

In addition to sub-sampling, the estimator for bipower variation was computed by averaging multiple "skip" versions. Skip versions of other estimators, particularly those of higher-order moments (such as fourth moments, or "integrated quarticity"), were found to possess superior statistical properties than returns computed using adjacent returns in Andersen, Bollerslev, and Diebold (2007). The "skip-q" bipower variation estimator is defined as

$$BV_q = \mu_1^{-2} \sum_{i=q+2}^{78} \left|p_{\lfloor ik \rfloor} - p_{\lfloor (i-1)k \rfloor}\right| \left|p_{\lfloor (i-1-q)k \rfloor} - p_{\lfloor (i-2-q)k \rfloor}\right|. \tag{11}$$

where $\mu_1 = \sqrt{2/\pi}$. The usual BV estimator is obtained when $q = 0$. We construct our estimator of bipower variation by averaging the skip-0 through skip-4 estimators, which represents a tradeoff between locality (skip-0) and robustness to both market microstructure noise and jumps that are not contained in a single sample (skip-4).[9] Using a skip estimator was advocated in Huang and Tauchen (2005) as an important correction to bipower which may be substantially biased in small samples, although to our knowledge the use of an average over multiple skip-q estimators is novel.[10]

Table 1 presents some summary statistics for the various volatility measures used in this paper. The upper panel presents average values for realized variance, bipower variation, positive and negative realized semivariances, and the signed jump variation measures. We see that the average value of daily RV for the SPDR was 1.154, implying 17.1% annualized volatility. The corresponding value for individual firms us 33.2%, indicating the higher average volatility of individual stock returns compared with the market. These figures reveal that variation due to jumps represent around 2% of total quadratic variation for the SPDR, while they represent around 13% for the average individual firm in our collection of 105 firms. (These proportions are ratios of averages of BV and RV across days; if we instead take the average of these ratios, we also get 2% and 13% as the proportion of quadratic variation due to jumps.) In the middle panel of this table we observe that the first-order autocorrelation of the SPDR volatility series (RV, BV, RS^+ and RS^-) ranges from 0.47 to 0.70. The autocorrelations of the signed jump variation series for the SPDR are lower, ranging from -0.11 to 0.06. The corresponding figures for the individual firms are similar. The lower panel presents correlations between the various volatility measures where the continuous component of volatility produces large correlation in RV, BV, $RS^+$and RS^-. The correlation between RS^+ and RS^-, at around 80%, is markedly lower than the correlation between these and either RV or BV indicating that there is novel information in this decomposition.

[INSERT TABLE 1 ABOUT HERE]

3.3 Model Estimation and Inference

We analyze the empirical features of these new measures of volatility using the popular Heterogeneous Autoregression (HAR) model, see Corsi (2009) and Müller, Dacorogna, Dav, Olsen, Pictet, and von Weiz-

[9] Events which are often identified as jumps in US equity data correspond to periods of rapid price movement although these jumps are usually characterized by multiple trades during the movement due to price continuity rules faced by market makers.

[10] We also conducted our empirical analysis using the *MedRV* estimator of Andersen, Dobrev, and Schaumburg (2012), which is an alternative jump-robust estimator of integrated variance. The resulting estimates and conclusions were almost identical to using BV and we omit them in the interest of brevity.

sacker (1997). HARs are parsimonious restricted versions of high-order autoregressions. The standard HAR in the realized variance literature regresses realized variance on three terms: the past 1-day, 5-day and 22-day average realized variances. To ease interpretation, we use a numerically identical reparameterization where the second term consists of only the realized variances between lags 2 and 5, and the third term consists of only the realized variances between lag 6 and 22,

$$\bar{y}_{h,t+h} = \mu + \phi_d y_t + \phi_w \left(\frac{1}{4} \sum_{i=1}^{4} y_{t-i} \right) + \phi_m \left(\frac{1}{17} \sum_{i=5}^{21} y_{t-i} \right) + \epsilon_{t+h} \tag{12}$$

where y denotes the volatility measure (RV, BV, etc.), and $\bar{y}_{h,t+h} = \frac{1}{h}\sum_{i=1}^{h} y_{t+i}$ is the h-day average cumulative volatility.[11] Throughout the paper, we will use $\bar{y}_{w,t}$ to indicate the average value over lags 2 to 5, and $\bar{y}_{m,t}$ to denote the average value between lags 6 and 22. We estimate the model above for forecast horizons ranging from h =1 to 66 days.

As the dependent variable in all of our regressions is a volatility measure, estimation by OLS has the unfortunate feature that the resulting estimates focus primarily on fitting periods of high variance, and place only little weight on more tranquil periods. This is an important drawback in our applications as the level of variance changes substantially across our sample period, and the level of the variance and the volatility in the error are known to have a positive relationship. To overcome this, we estimate our models using simple weighted least squares (WLS). To implement this we first estimate the model using OLS, and then construct weights as the inverse of the fitted value from that model.[12]

The left-hand-side variable includes leads of multiple days and so we use a Newey and West (1987) HAC to make inference on estimated parameters. The bandwidth used was $2(h-1)$ where h is the lead length of the left-hand-side variable.

[11] In the internet appendix we present results where the h-day ahead *daily* volatility measure, y_{t+h}, is used as the dependent variable rather than the cumulative volatility.

[12] This implementation of WLS is motivated by considering the residuals of the above regression to have heteroskedasticity related to the level of the process. This is related to standard asymptotic theory for realized measures, see Andersen, Bollerslev, Diebold, and Labys (2003). An alternative approach is to use OLS on log-volatility, however this leads, of course, to predictions of log-volatility rather than volatility in levels, and the latter are usually of primary interest in economic applications. For comparison, Tables 2, 3 and 4 in the web appendix present results from analyses based on log-volatility, and show that all of our conclusions hold using this alternative specification.

3.4 A Panel HAR for Volatility Modeling

Separate estimation of the models on the individual firms' realized variance is feasible, but does not provide a direct method to assess the significance of the average effect, and so we estimate a pooled unbalanced panel HAR with a fixed effects to facilitate inference on the average value of parameters. To illustrate, in the simplest specification the panel HAR is given by:

$$\overline{y}_{h,i,t+h} = \mu_i + \phi_d y_{i,t} + \phi_w \overline{y}_{w,i,t} + \phi_m \overline{y}_{m,i,t} + \epsilon_{i,t+h}, \;\; i = 1,\ldots,n_t, \;\; t = 1,\ldots,T,$$

where μ_i is a fixed effect which allows each firm to have different levels of long-run volatility. Let $Y_{i,t} = [y_{i,t}, \bar{y}_{w,i,t}, \bar{y}_{m,i,t}]'$, then the model for each firm's realized variance can be compactly expressed as:

$$\overline{y}_{h,i,t+h} = \mu_i + \boldsymbol{\phi}' Y_{i,t} + \epsilon_{i,t+h}, \;\; i = 1,\ldots,n_t, \;\; t = 1,\ldots,T.$$

Next, define $\tilde{y}_{h,i,t+h} = \overline{y}_{h,i,t+h} - \hat{\upsilon}_{h,i}$ and $\tilde{Y}_{i,t} = Y_{i,t} - \hat{\Upsilon}_i$ where $\hat{\upsilon}_{h,i}$ and $\hat{\Upsilon}_i$ are the WLS estimates of the mean of $\overline{y}_{h,i}$ and Y_i, respectively. The pooled parameters are then estimated by:

$$\hat{\boldsymbol{\phi}} = \left(T^{-1} \sum_{t=1}^{T} \left(n_t^{-1} \sum_{i=1}^{n_t} w_{i,t} \tilde{Y}_{i,t} \tilde{Y}_{i,t}' \right) \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \left(n_t^{-1} \sum_{i=1}^{n_t} w_{i,t} \tilde{Y}_{i,t} \tilde{y}_{h,i,t+h} \right) \right). \tag{13}$$

where $w_{i,t}$ are the weights and n_t are the number of firms in the cross section at date t.[13]

Inference can be conducted using the asymptotic distribution

$$\begin{aligned} \sqrt{T}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right) &\overset{d}{\to} N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right) \;\; \text{as } T \to \infty \qquad (14) \\ \text{where } \boldsymbol{\Sigma} &= \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left(n_t^{-1} \sum_{i=1}^{n_t} w_{i,t} \tilde{Y}_{i,t} \tilde{Y}_{i,t}' \right) \\ \boldsymbol{\Omega} &= \text{avar}\left(T^{-1/2} \sum_{t=1}^{T} \mathbf{z}_{t+h} \right) \end{aligned}$$

[13]Our analysis takes the cross-section size, n_t, as finite while the time series length diverges. In our application we have $n_t \in [71, 100]$ and T=2,795. If an approximate factor structure holds in the returns we study, which is empirically plausible, then the same inference approach could be applied even if $n_t \to \infty$, as in that case we would find $\text{plim}_{n_t\to\infty} \text{V}\left[n_t^{-1}\sum_{i=1}^{n_t} w_{i,t}\tilde{Y}_{i,t}\epsilon_{i,t}\right] \to \tau^2 > 0$. A similar result was found in the context of composite likelihood estimation, and this asymptotic distribution can be seem as a special case of Engle, Shephard, and Sheppard (2008).

$$\mathbf{z}_{t+h} = n_t^{-1} \sum_{i=1}^{n_t} w_{i,t} \tilde{Y}_{i,t} \epsilon_{i,t+h}.$$

In addition to the results from the panel estimation, we also fit the models to each series individually and summarize the results as aggregates in the tables below.

4 Predicting Volatility using Realized Semivariances

Before moving into models that decompose realized volatility into signed components, it is useful to establish a set of reference results. We fit a reference specification, the standard HAR model:

$$\overline{RV}_{h,t+h} = \mu + \phi_d RV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h} \tag{15}$$

to both the S&P 500 ETF and the panel where $\overline{RV}_{w,t}$ is the average between lags 2 and 5 and $\overline{RV}_{m,t}$ is the average value using lags 6 through 22. This model is identical to the specification studied in Andersen, Bollerslev, and Diebold (2007). The panel version of the model is identical to eq. (15) except for the inclusion fixed effects to permit different long-run variances for each asset. Tables 2a and 2b each contain four panels, one for each horizon 1, 5, 22 and 66. The first line of each panel contains the estimated parameters and t-statistics for this specification. These results are in line with those previously documented in the literature: substantial persistence, with $\phi_d + \phi_w + \phi_m$ close to 1, and the role of recent information, captured by ϕ_d, diminishing as the horizon increases.[14] The results for both the SPDR and the panel are similar, although the SPDR has somewhat larger coefficients on recent information. The final column reports the R^2, which is computed using the WLS parameter estimates and the original, unmodified data.

[INSERT TABLES 2a AND 2b ABOUT HERE]

[14] Tables A.6a and A.6b in the internet appendix contain corresponding results when the dependent variable is the h-day ahead daily volatility. These tables reveal that, as expected, much, but not all, of the predictive power in the model for cumulative realized variance occurs at short horizons.

4.1 Decomposing Recent Quadratic Variation

Given the exact decomposition of RV into RS^+ and RS^-, we extend eq. (15) to obtain a direct test of whether signed realized variance is informative for future volatility. Here, we only decompose the most recent volatility (RV_t), and in the web appendix we present results and analysis when all three volatility terms are decomposed. Applying this decomposition produces the specification:

$$\overline{RV}_{h,t+h} = \mu + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}. \tag{16}$$

The panel specification of the above model includes fixed effects but is otherwise identical. Note that if the decomposition of RV into $RS^+$and RS^- added no information we would expect to find $\phi_d^+ = \phi_d^- = \phi_d$. Our first new empirical results using realized semivariances are presented in the second row of each panel of Tables 2a and 2b. In the models for the SPDR (Table 2a), we find that the coefficient on negative semivariance is larger and more significant than that on positive semivariance for all horizons. In fact, the coefficient on positive semivariance is not significantly different from zero for $h = 1$, 5 and 22, while it is small and significantly *negative* for $h = 66$. The semivariance model explains 10-20% more of the variation in future volatility than the model which contains only realized variance. The effect of lagged RV implied by this specification is $(\phi_d^+ + \phi_d^-)/2$, and we see that it is similar in magnitude to the coefficient found in the reference specification where we only include lagged RV, which indicates that models which only use RV are essentially averaging the vastly different effects of positive and negative returns. The results for the panel of individual volatility series also reveal that negative semivariance has a larger and more significant impact on future volatility, although in these results we also find that positive semivariance has significant coefficients. The difference in the results for the index and for the panel points to differences in the impact of idiosyncratic jumps in the individual firms' volatility, which we explore in the next section.[15]

Figure 1 contains the point estimates of ϕ_d^+ and ϕ_d^- from eq. (16) for all horizons between 1 and 66

[15] While the coefficients on negative semivariances are positive for both the SPDR and the panel of individual stocks, one difference between the two sets of results is that the coefficients on positive semivariances are generally insignificant or negative for the SPDR, and positive for the panel of individual stocks. This may be due to the presence of idiosyncratic jumps in individual stocks, while these are "averaged out" in the SPDR market index and only "systematic" jump behavior is captured. We leave detailed analysis of idiosyncratic and systematic jumps for future research.

along with pointwise confidence intervals. For the SPDR, positive semivariance plays essentially no role at any horizon. The effect of negative semivariance is significant and positive, and declines as the horizon increases. In the panel both positive and negative semivariances are significant although the coefficients differ substantially in magnitude for all horizons. The effect of positive semivariance is economically small from horizon 15. The smoothness indicated in both curves is a feature of the estimated parameters – no additional smoothing was used to produce these figures.

[INSERT FIGURE 1 ABOUT HERE]

As noted above, if the decomposition of RV into $RS^{+}$and RS^{-} added no new information, then we would expect to see $\phi_d^{+} = \phi_d^{-} = \phi_d$. We reject this restriction at the 0.05 level for all but 3 out of 66 horizons (h=36,43,48) for the SPDR, and in the panel this null is rejected for all horizons.[16] We interpret these findings as strong evidence that decomposing RV into its signed components significantly improves the explanatory power of this model.

Realized variance can be decomposed not only at the first lag but at higher lags as well. A "full" decomposition allows for a refined view of the sources of persistence of these two components of realized variance, and leads to a natural Vector HAR (VHAR) specification for RS^{+} and RS^{-}. We estimated this model using both RV as well as the two semivariances as dependent variables, and present the results in Appendix B. We find that negative realized semivariance is much more important for *both* negative and positive realized semivariance, and disentangling the effects of these two components significantly improves forecasts of both measures of future volatility. This holds for horizons ranging from one day to three months.

4.2 Comparison with a Simple Leverage Effect Variable

The classic leverage effect, whether due to varying firm leverage as in Black (1976) and Christie (1982) or volatility-feedback in Campbell and Hentschel (1992), is usually modeled using a lagged squared return interacted with an indicator for negative returns, as in Glosten, Jagannathan, and Runkle (1993). In this section we determine whether our approach using information from realized semivariances adds anything

[16] Detailed test results for each horizon are omitted in the interests of brevity, but are available from the authors upon request.

beyond this simple approach. To do so, we augment the regressions from the previous section with a term that interacts the lagged realized variance with an indicator for negative lagged daily returns, $RV_t I_{[r_t<0]}$.[17]

$$\overline{RV}_{h,t+h} = \mu + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \gamma RV_t I_{[r_t<0]} + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h} \tag{17}$$

If realized semivariance added no new information beyond the interaction variable then we would expect $\phi_d^+ = \phi_d^-$ and γ to be significant.

The final row in each panel of Tables 2a and 2b contain the parameter estimates from this model. In all cases the magnitude of the coefficient on the interaction term is small, and we again find that the coefficient on negative realized semivariance is much larger than that on positive semivariance. In models based on the SPDR, the interaction term has the opposite sign to what is commonly found at $h = 22$ and 66, and is insignificant at the 1-day horizon. This coefficient in the panel model is significantly positive but small, generally only 10% of the magnitude the coefficient on negative realized semivariance, and in all cases the gain in R^2 from including this interaction variable is just 0.001.[18]

The results in this section show that negative semivariance captures the asymmetric impact of negative and positive past returns on future volatility better than the usual method of using an indicator for the sign of the lagged daily return. This is true across all horizons considered (1, 5, 22 and 66 days). Thus there is more information about future volatility in the high frequency negative variation of returns than in the direction of the price over a whole day.

5 Signed Jump Information

All of the models estimated thus far examine the role that decomposing realized variances into positive and negative realized semivariance can play in explaining future volatility. These results consistently sug-

[17] We interact the indicator variable with the lagged realized variance rather than the lagged squared return as the latter is a noisier measure of volatility than the former. The results using the usual version of this interaction variable, $r_t^2 I_{[r_t<0]}$ are even weaker than those discussed here.

[18] We also considered a specification that includes the indicator variable $I_{[r_t<0]}$ in addition to the interaction variable, see Tables A.2a and A.2b in the internet appendix. This table shows that our main results continue to hold in this more general specification: the coefficients on positive and negative semivariances are each qualitatively similar in size, and are strongly significantly different from each other, regardless of the inclusion of the indicator and/or interaction variable.

gest that the information content of negative realized semivariance is substantially larger than that of positive realized semivariance. While the theory of BNKS shows that the difference in these two can be attributed to differences in jump variation, the direct effect of jumps is diluted since realized semivariances also contain one-half of the integrated variance, see eq. (7).

In this section we use *signed jump variation*, $\Delta J_t^2 \equiv RS_t^+ - RS_t^-$, as a simple method to isolate the information from signed jumps. This difference eliminates the common integrated variance term and produces a measure that is positive when a day is dominated by an upward jump and negative when a day is dominated by a downward jump. This measure has the advantage that a jump-robust estimator of integrated variance, such as BV or *MedRV*, is not needed; we obtain the measure simply as the difference between RS_t^+ and RS_t^-. If jumps are rare, as often found in the stochastic volatility literature, then this measure should broadly correspond to the jump variation when a jump occurs, and to mean zero noise otherwise.

To explore the role that signed jumps play in future variance we formulate a model which contains signed jump variation and an estimator of the variation due to the continuous part (bipower variation):

$$\overline{RV}_{h,t+h} = \mu + \phi_J \Delta J_t^2 + \phi_C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}. \tag{18}$$

The panel specification includes fixed effects but is otherwise identical.[19]

Results from the model with signed jumps are presented in the second row of each of the four panels in Tables 3a and 3b. Signed jump variation, ΔJ_t^2, has a uniformly negative sign and is significant for all forecast horizons. This reveals that days dominated by negative jumps lead to higher future volatility, while days with positive jumps lead to *lower* future volatility. This result is quite different from that of Andersen, Bollerslev, and Diebold (2007), who found that (unsigned) jumps lead to only a slight decrease in future variance in the S&P 500.[20] By including information about the sign of the jump, we find that the jump variable does indeed help predict future volatility.

[19]It is worth noting that while this specification is similar to our baseline model (eq. 16) it is not nested by it, as it is not possible to construct a measure of the continuous component of variation from the two realized semivariances alone.

[20]It should be noted, however, that Andersen, Bollerslev, and Diebold (2007) pretest for jumps and so on days where no jump component is detected their jump measure is exactly zero. Since we do not pretest, we may have a noisier jump measure, although it remains consistent for the object of interest.

We next modify this model to use BV as the dependent variable, in order to see whether signed jump variation is useful for predicting future *continuous* variation. The results from this model are presented in the bottom row of Tables 3a and 3b, and reveal that using BV as the dependent variable resulted in virtually identical estimates to those obtained using RV. Thus signed jump variation is indeed useful for predicting the continuous part of volatility. This is a novel finding, and one that cannot be detected without drawing on information about the *sign* of the high frequency returns.

[INSERT TABLES 3a AND 3b ABOUT HERE]

To determine whether the coefficient on positive jump variation differs from that of negative jump variation, and thus whether the impact of jumps is driven more by positive or negative jump variation, we extend this model to include $\sum \Delta p_s^2 I_{[\Delta p_s > 0]}$ and $\sum \Delta p_s^2 I_{[\Delta p_s < 0]}$ separately. One option would be to subtract one half of a consistent estimator of the IV, for example to use $RS_t^+ - \frac{1}{2} BV_t$. We opt instead for a simpler specification which uses an indicator for which realized semivariance was larger. This model is:

$$\overline{RV}_{h,t+h} = \mu + \phi^{J+} \Delta J_t^{2+} + \phi^{J-} \Delta J_t^{2-} + \phi^C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h} \quad (19)$$

$$\text{where } \Delta J_t^{2+} = \left(RS_t^+ - RS_t^-\right) \mathrm{I}_{\left[(RS_t^+ - RS_t^-) > 0\right]} \text{ and } \Delta J_t^{2-} = \left(RS_t^+ - RS_t^-\right) \mathrm{I}_{\left[(RS_t^+ - RS_t^-) < 0\right]}$$

If the two signed jump components have equal predictive power then we expect to find $\phi^{J+} = \phi^{J-} = \phi^J$. The penultimate row of each panel in Table 3a contains estimates for this extended jump specification. For the SPDR we find that both signed jump components have a negative sign, and for the longest two horizons ($h = 22$ and $h = 66$) the coefficients are almost equal. For the shorter two horizons ($h = 1$ and $h = 5$), the coefficient on the negative jump component is larger, in magnitude, than on the positive jump component, indicating that the increase in future volatility is larger in magnitude following a negative jump than the decrease in future volatility following a positive jump. We test the null $H_0 : \phi^{J+} = \phi^{J-}$ we reject only at the one-step-ahead horizon. In the panel, both types of jumps lead to higher future volatility for the $h = 1$ horizon, although the magnitude of the coefficient differs by a factor of 10, and negative jumps have a larger effect. At longer horizons, "good" jumps lead to lower volatility while "bad" jumps lead to higher volatility.[21] Figure 2 contains a plot of the coefficients for all 66 leads for both the SPDR and the panel. Aside from some mixed evidence for very short term effects, both sets of coefficients are negative and significant. The significance of the variation due to positive jumps contrasts with the weaker evidence of significance for positive realized semivariance. These results may be reconciled by noting that positive realized semivariance contains, in the limit, both the variation due to positive jumps and one-half of integrated variance, the latter also appearing in negative semivariance. By "stripping out" the integrated variance component and focusing only on the jump component, we find that positive jumps do have an important (negative) impact on future volatility. This model was also fit to the individual firm series, and Figure 3 shows the magnitude and statistical significant of the coefficient, which was significantly negative in 83 series at the 1-day horizon and 89 at the 5-day horizon, indicating a strong directional effect of jumps on future volatility.

Finally, in the top row of each panel of Table 3a, we consider a model with no jump variation measures; we include just BV at the one-day lag, along with $\overline{RV}_{w,t}$ and $\overline{RV}_{m,t}$. Consistent with the significance of the signed jump variation measures in the specifications discussed above, we observe a substantial drop in R^2, particularly at short horizons. For the SPDR, R^2 falls from 0.61 to 0.56 for $h = 1$, and from 0.62 to 0.58 for $h = 5$.

[INSERT FIGURE 2 ABOUT HERE]

6 Out-of-sample evidence

This section presents two "out-of-sample" checks on the conclusions from the previous section on the importance of signed measures of variation. The first is an analysis of two international stock indices, the FTSE 100 index of United Kingdom stocks and the EURO STOXX 50 index of stocks from 12 European countries. The second is an analysis of the pseudo-out-of-sample forecasting performance of models

[21] We note that this sign change in the reaction of future volatility to current price moves is consistent with the original "leverage" explanation offered by Black (1976), which focuses on the degree of financial leverage of a firm, although that explanation does not distinguish between small and large price moves (corresponding to continuous and jump variation, in our framework). It may also be related to differences between the economic sources, e.g. news announcements, of positive and negative jumps, see Bajgrowicz, Scaillet, and Treccani (2012) for related work. We do not attempt to identify individual jumps, and so the effects we report may be interpreted as average effects for positive and negative news.

based on those above.

6.1 International evidence

The previous sections presented results for the SPDR, an exchange traded fund tracking the S&P 500 index of U.S. firms, and for 105 individual U.S. firms. In this section we present results for two international equity indices, the FTSE 100 index of United Kingdom stocks and the EURO STOXX 50 index of stocks from 12 European countries. Data on both indices is taken from Thomson Reuters Tick History, and covers the same period as the main results.[22] Both indices are computed from the underlying basket of 100 and 50 stocks, respectively, and prices were cleaned using the rules 1, 2, 5 and 6 from Appendix A using the local market trading times in place of U.S. open hours and daily Tick History verified high-low range data. This data is much cleaner than TAQ data and a total of 6 (high-frequency) observation were removed.

Table 4 presents results for the FTSE and the STOXX. The top row of each panel presents results for a standard HAR, corresponding to the top row of each panel in Table 2a for the SPDR. The second row presents results for a HAR model with the one-day lag of realized variance decomposed into positive and negative semivariance, (eq. 16) and can be compared with the second row of each panel in Table 2a. In common with the results for the SPDR, we find that negative realized semivariance is much more important for predicting future volatility than positive semivariance. For the FTSE index, positive semivariance is significant for only the two shorter horizons. For the STOXX index, it is significant at all four horizons, but has coefficients that are less than one-half of those on negative semivariance in all cases. For both indices and all four forecast horizons we can reject, at the 0.05 level, the null that $\phi_d^+ = \phi_d^-$, and thus we conclude that realized semivariances yield significant explanatory gains for both of these indices.

The third row of each panel in Table 4 presents results for a model that includes a measure of continuous and signed jump variation (eq. 18) and can be compared with the results presented in the top row of each panel of Table 3a. For both indices and all four forecast horizons we find that ϕ_J is negative, consistent with the results for the SPDR. This parameter is significantly negative for all four horizons for the FTSE, and for all but the longest horizon for the STOXX. This suggests that negative jumps lead to higher

[22]The first date available for the EURO STOXX 50 is February 26, 1998, and so we use that as the start date for that series; for the FTSE 100 index we start on June 23, 1997.

future volatility, while positive jumps lead to *lower* future volatility, further motivating the monikers "bad volatility" and "good volatility."

[INSERT TABLE 4 ABOUT HERE]

6.2 Pseudo-out-of-sample forecast performance

We now consider a pseudo out-of-sample forecasting application to see whether the in-sample gains documented in Sections 4 and 5 lead to better forecasts out-of-sample. We consider three classes of models, each with two or three variations. All models include $\overline{RV}_{w,t}$ and $\overline{RV}_{m,t}$, and they differ in what previous-day information is used. The first model, denoted $\widehat{RV}^{HAR}$, is the standard RV-HAR containing lags 1, 5 and 22 of RV (eq. 15). The second, denoted $\widehat{RV}^{GJR}$, augments the standard HAR with an interaction term which allows for asymmetry in persistence when the previous daily return was negative, $RV_t \mathrm{I}_{[r_t<0]}$, (eq. 17). The second class of models uses information in positive and negative semivariance: $\widehat{RV}^{RS}$ is a specification that decomposes recent realized variance into positive and negative semivariance (eq. 16), and $\widehat{RV}^{RS-}$, is a restricted version of $\widehat{RV}^{RS}$ where positive realized semivariance is excluded from the model, motivated by the relative magnitude of the coefficient and limited significance of this variable in Tables 2a and 2b. The third class of models considers the information in jump variation: $\widehat{RV}^{BV}$ is a model that excludes jump information and only includes bipower variation; $\widehat{RV}^{\Delta J^2}$ is a specification that includes BV_t and ΔJ_t^2 (eq. 18), and $\widehat{RV}^{\Delta J^{2\pm}}$ is a specification that breaks ΔJ_t^2 into its positive and negative components (eq. 19).

All forecasts are generated using rolling WLS regressions based on 1,004 observations (4 years), and parameter estimates are updated daily. Only series that contain at least 500 out-of-sample data points are included, reducing the number of individual firms from 105 to 95. No restrictions on the parameters are imposed and forecasts are occasionally negative (approximately .004%), and so an "insanity filter" is used to ensure that the forecasts were no smaller than the smallest realization observed in the estimation window.

Forecast performance is evaluated using unconditional Diebold and Mariano (1995) - Giacomini and

White (2006) tests, using the negative of the Gaussian quasi-likelihood as the loss function,

$$L\left(\widehat{RV}_{h,t+h|t}, \overline{RV}_{h,t+h}\right) = \ln\left(\widehat{RV}_{h,t+h|t}\right) + \frac{\overline{RV}_{h,t+h}}{\widehat{RV}_{h,t+h|t}}.$$

This "QLIKE" loss function has been shown to be robust to noise in the proxy for volatility in Patton (2011), and to have good power properties in Patton and Sheppard (2009).

Table 5 contains results from the forecasting analysis. Each of the three panels contains results from comparing one pair of forecasting models. Within each panel, the left-most column contains the value of the DM test statistic for the S&P 500 ETF, and the two right columns contain the percentage of the 95 individual series which favor the each of the competing models using a 2-sided 5% test.

The top-left panel compares the standard HAR with a semivariance-based model which decomposes the first lag. The DM test statistic is positive across all forecast horizons, indicating the superior out-of-sample performance of the semivariance model for the S&P 500 ETF, and rejects the null of equal performance in favor of the semivariance-based model in 22% to 30% of individual series. The middle panel of the top row compares the standard HAR to a model which includes only negative semivariance at the first lag. This is our preferred realized semivariance specification in light of the weak evidence of significant of positive semivariance, and this model has the same number of parameters as the standard HAR. The restricted semivariance model outperforms the standard HAR at all horizons for the S&P 500, and provides better performance for individual stocks than the less parsimonious specification. The top-right panel compares the parsimonious realized semivariance specification to the realized variance HAR which includes the interaction variable using the sign of the lagged return. The interaction variable appears to help at short horizons, with the performance of that model being not significant different from our preferred semivariance specification, however the asymmetry-augmented HAR is significantly outperformed at longer horizons by the semivariance-based forecast.

The lower-left panel compares models that differ in the one-day lag information: the first model uses BV while the second model uses negative semivariance (RS^{-}). We observe that negative semivariance outperforms BV for the SPDR at all four horizons. For individual series the outperformance is significant for all but the shortest horizon, where the two models perform comparably well. The middle panel of the bottom row of Table 5 compares a forecasting model that excludes jump information (from the one-day lag) with a model that includes it through the variable ΔJ_t^2. We see that the model that incorporates signed jump information significantly outperforms, for the SPDR, the one that does not for two out of the four horizons ($h = 5$ and $h = 22$). For individual stocks ΔJ_t^2 significantly outperforms BV for between 17% and 25% of series. Finally, the lower-right panel compares the model based on BV with one that breaks jump variation into its positive and negative components. We again find that information from signed jumps significantly improves out-of-sample forecast performance, although generally less than a simple model with only negative semivariance.

[INSERT TABLE 5 ABOUT HERE]

Table 6 reports the out-of-sample R^2 values for the seven forecasting models considered in Table 5. Relative to a baseline HAR specification, the best semivariance-based alternative generates gains in out-of-sample R2 of between 1.1% ($h = 66$) and 3.0% for the SPDR ($h = 22$), and between 0.5% ($h = 66$) and 13.5% ($h = 1$) for the individual stocks. Thus statistically significant gains documented in Table 6 also correspond to economically meaningful improvements.

[INSERT TABLE 6 ABOUT HERE]

7 Conclusion

This paper shows the sizable and significant gains for predicting equity volatility by incorporating signed high frequency volatility information. Our analysis is based on the "realized semivariance" estimators recently proposed by Barndorff-Nielsen, Kinnebrock, and Shephard (2010). These simple estimators allow us to decompose realized volatility into a part coming from positive high frequency returns and a part coming from negative high frequency returns. For three equity market indices and a set of 105 individual stocks, we find that negative realized semivariance is much more important for future volatility than positive realized semivariance, and disentangling the effects of these two components significantly improves forecasts of future volatility. This is true whether the measure of future volatility is realized variance, bipower variation, negative realized semivariance or positive realized semivariance, and holds for horizons ranging from one day to three months. We also find that jump variation is important for predicting future volatility, with volatility attributable to negative jumps leading to significantly higher future volatility, while positive jumps lead to significantly *lower* volatility. This may explain earlier results in this literature, see Andersen, Bollerslev, and Diebold (2007) and Busch, Christensen, and Nielsen (2011) for example, who found that jumps are of limited use for forecasting future volatility; only by including the jump size *and* sign are the gains from jumps realized. Assessing the usefulness of realized semivariances and signed jump variation in concrete financial applications, such as portfolio management, density forecasting and derivatives pricing, as in Fleming, Kirby, and Ostdiek (2003), Maheu and McCurdy (2011) and Christoffersen and Jacobs (2004) for example, represents an interesting area for future research.

References

AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005): "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, 18(2), 351.

AÏT-SAHALIA, Y., AND J. YU (2009): "High Frequency Market Microstructure Noise Estimates and Liquidity Measures," *Annals of Applied Statistics*, 3(1), 422–457.

ANDERSEN, T. G., T. BOLLERSELV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2006): "Volatility and Correlation Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann. North Holland Press, Amsterdam.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *The Review of Economics and Statistics*, 89(4), 701–720.

——— (2009): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics*, ed. by Y. Aït-Sahalia, and L. P. Hansen. North Hollan, Amsterdam.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000): "Great Realizations," *Risk*, 13(99-061), 105–108.

——— (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96(453), 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(1), 3–29.

ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2012): "Jump-robust volatility estimation using nearest neighbor truncation," *Journal of Econometrics*, 169(1), 75–93.

ANG, A., J. CHEN, AND Y. XING (2006): "Downside Risk," *Review of Financial Studies*, 19(4), 1191–1239.

BAJGROWICZ, P., O. SCAILLET, AND A. TRECCANI (2012): "Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News," Discussion paper, Swiss Finance Institute Research Paper No. 11-36.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2004): "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 73(4), 885–925.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76(6), 1481–1536.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), C1–C32.

BARNDORFF-NIELSEN, O. E., S. KINNEBROCK, AND N. SHEPHARD (2010): "Measuring downside risk – realised semivariance," in *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, ed. by T. Bollerslev, J. Russell, and M. Watson. Oxford University Press.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4(1), 1–30.

BLACK, F. (1976): "Studies of Stock Prices Volatility Changes," in *Proceeding from the American Statistical Association, Business and Economics Statistics Section*, pp. 177–181. American Statistical Association.

BOLLERSLEV, T., R. F. ENGLE, AND D. B. NELSON (1994): "ARCH Models," in *Handbook of Econometrics*, vol. 4. Elsevier North Holland.

BOLLERSLEV, T., J. LITVINOVA, AND G. TAUCHEN (2006): "Leverage and volatility feedback effects in high-frequency data," *Journal of Financial Econometrics*, 4(3), 353.

BOLLERSLEV, T., AND V. TODOROV (2011): "Tails, Fears, and Risk Premia," *The Journal of Finance*, 66, 2165–2211.

BUSCH, T., B. CHRISTENSEN, AND M. NIELSEN (2011): "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," *Journal of Econometrics*, 160, 48–57.

CAMPBELL, J. Y., AND L. HENTSCHEL (1992): "No News Is Good News: An Asymetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31(3), 281–318.

CHEN, X., AND E. GHYSELS (2011): "News–Good or Bad–and Its Impact on Volatility Predictions over Multiple Horizons," *Review of Financial Studies*, 24(1), 46–81.

CHIRIAC, R., AND V. VOEV (2007): "Long Memory Modelling of Realized Covariance Matrices," Discussion paper, University of Konstanz.

CHRISTIE, A. (1982): "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10(4), 407–432.

CHRISTOFFERSEN, P. F., AND K. JACOBS (2004): "Which GARCH Model for Option Valuation?," *Management Science*, 50, 1204–1221.

CORSI, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

CORSI, F., D. PIRINO, AND R. RENÒ (2010): "Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting," *Journal of Econometrics*, 159, 276–288.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–263.

ENGLE, R. F., N. SHEPHARD, AND K. SHEPPARD (2008): "Fitting vast dimensional time-varying covariance models," Discussion paper, Oxford-Man Institute of Quantitative Finance.

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2003): "The economic value of volatility timing using "realized" volatility," *Journal of Financial Economics*, 67(3), 473–509.

FORSBERG, L., AND E. GHYSELS (2007): "Why do absolute returns predict volatility so well?," *Journal of Financial Econometrics*, 5(1), 31–67.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131(1-2), 59–95.

GIACOMINI, R., AND H. WHITE (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): "On the Relationship between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48(5), 1779–1801.

HOGAN, W. W., AND J. M. WARREN (1974): "Toward the Development of an Equilibrium Capital-Market Model Based on Semivariance," *The Journal of Financial and Quantitative Analysis*, 9(1), 1–11.

HUANG, X., AND G. TAUCHEN (2005): "The Relative Contribution of Jumps to Total Price Variance," *Journal of Financial Econometrics*, 3(4), 456–499.

LANNE, M. (2006): "A Mixture Multiplicative Error Model for Realized Volatility," *Journal of Financial Econometrics*, 4(4), 594–616.

——— (2007): "Forecasting Realized Exchange Rate Volatility by Decomposition," *International Journal of Forecasting*, 23, 307–320, EUI Working Paper ECO No. 2006/20.

LEWIS, A. L. (1990): "Semivariance and the performance of portfolios with options," *Financial Analysts Journal*, 46(4), 67–76.

LI, Y., P. A. MYKLAND, E. RENAULT, L. ZHANG, AND X. ZHENG (2013): "Realized Volatility When Sampling Times are Possibly Endogenous," *Econometric Theory*, 30, 1–26.

LIU, C., AND J. M. MAHEU (2005): "Modeling and Forecasting Realized Volatility: the Role of Power Variation," Discussion paper, Univeristy of Toronto.

MAHEU, J., AND T. MCCURDY (2011): "Do high frequency measures of volatility improve forecasts of return distributions?," *Journal of Econometrics*, 160(1), 69–76.

MÜLLER, U., M. DACOROGNA, R. DAV, R. OLSEN, O. PICTET, AND J. VON WEIZSACKER (1997): "Volatilities of different time resolutions - analysing the dynamics of market components," *Journal of Empirical Finance*, 4, 213–39.

NELSON, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A new approach," *Econometrica*, 59(2), 347–370.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

OOMEN, R. C. A. (2005): "Properties of Bias-Corrected Realized Variance Under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3(4), 555–577.

PATTON, A. J. (2011): "Volatility Forecast Comparison using Imperfect Volatility Proxies," *Journal of Econometrics*, 160(1), 246–256.

PATTON, A. J., AND K. SHEPPARD (2009): "Optimal combinations of realised volatility estimators," *International Journal of Forecasting*, 25(2), 218–238.

SORTINO, F. A., AND S. SATCHELL (2001): *Managing Downside Risk in Financial Markets*. Butterworth-Heinemann.

VISSER, M. P. (2008): "Forecasting S&P 500 Daily Volatility using a Proxy for Downward Price Pressure," Discussion paper, Korteweg-de Vries Institute for Mathematics, University of Amsterdam.

ZAKOIAN, J. M. (1994): "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18(5), 931–955.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

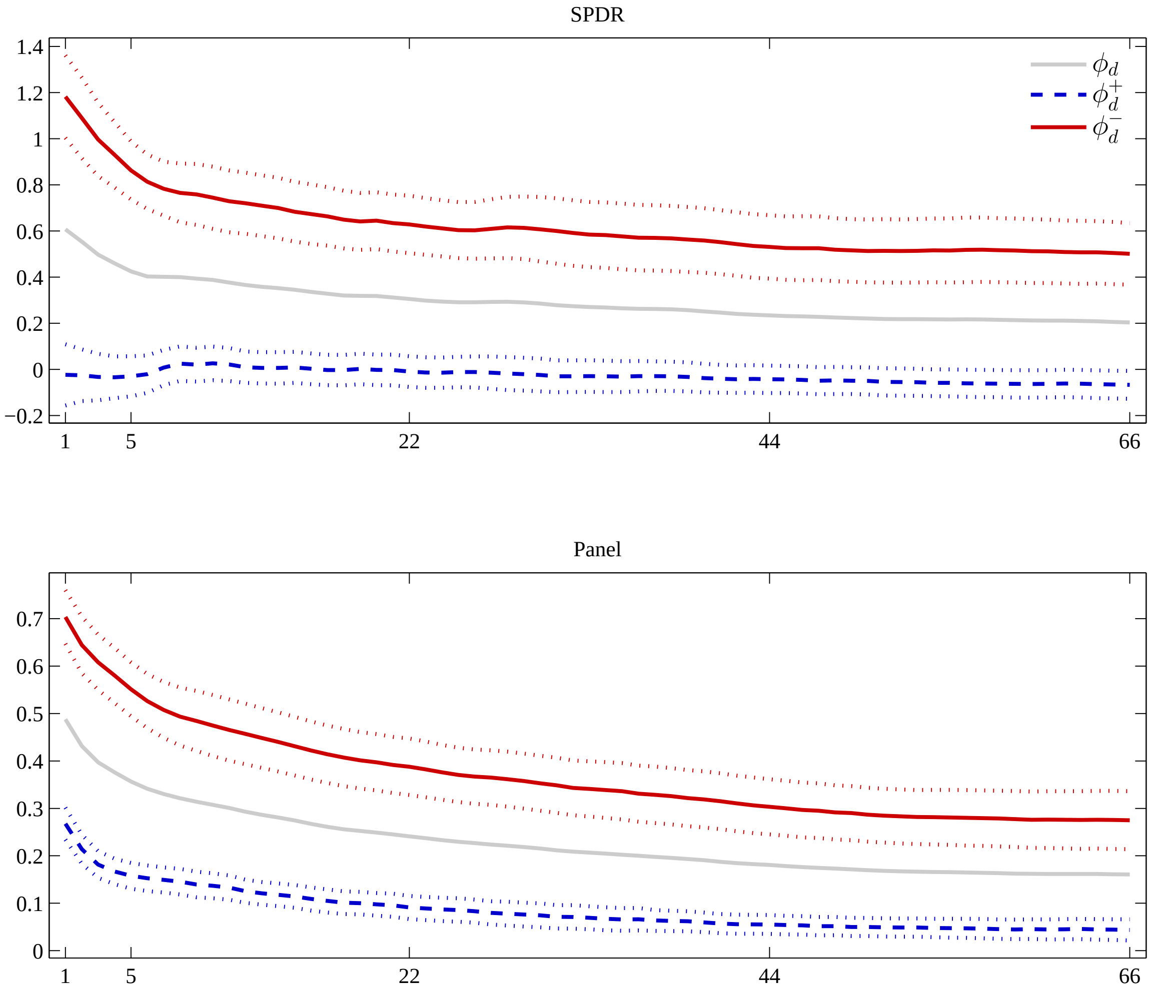

Figure 1: Estimated coefficients from a model that decomposes realized variance into its signed components, $\overline{RV}_{h,i,t+h} = \mu_i + \phi_d RV_{i,t} + \phi_d^+ RS_{i,t}^+ + \phi_d^- RS_{i,t}^- + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{t+h}$. 95% confidence intervals are indicated using dotted lines, and the estimated coefficient from a standard HAR model, ϕ_d, is presented in a light solid line. The top panel contains results for the S&P 500 SPDR and the bottom panel contains results for the panel of individual firm realized variances.

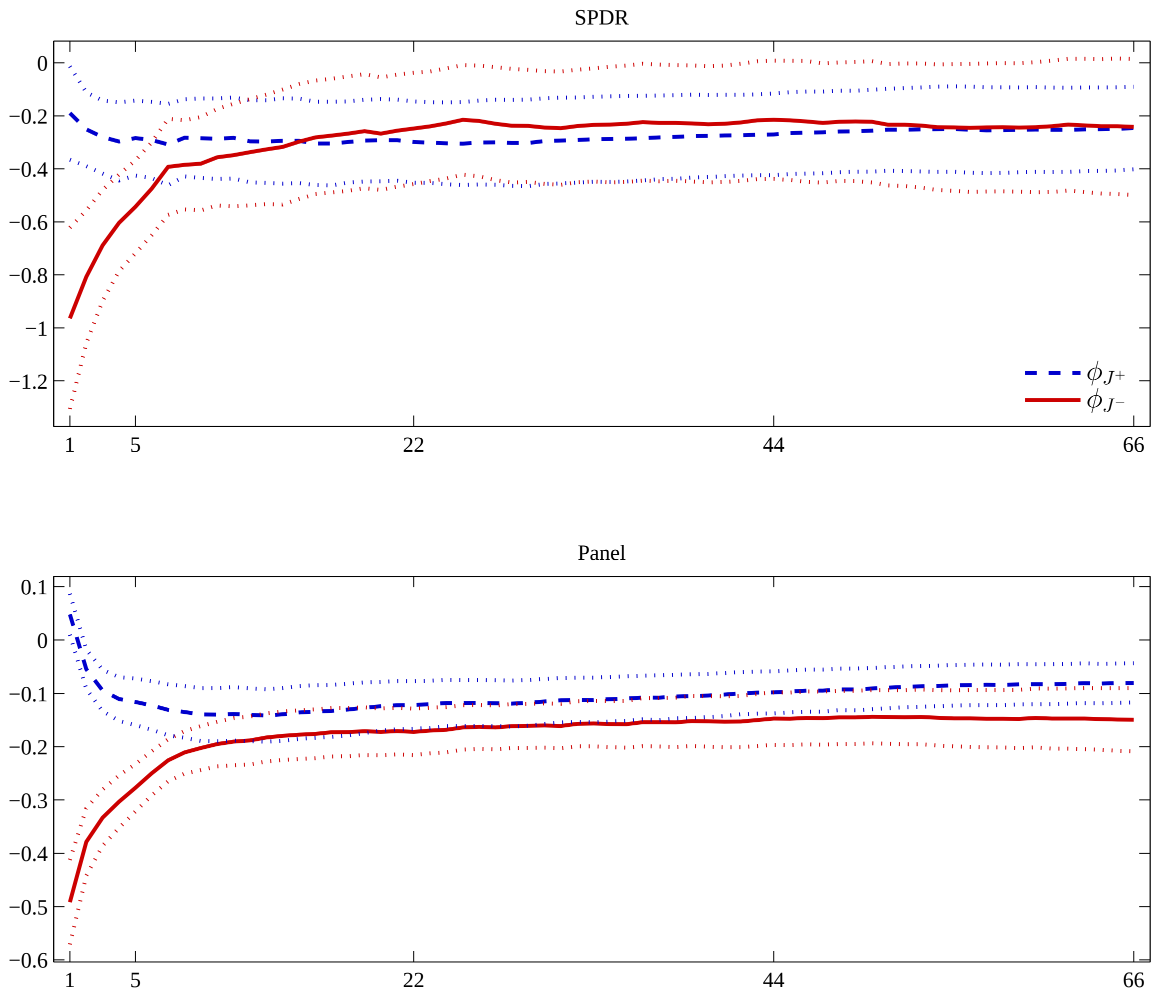

Figure 2: Estimated coefficients from a model that includes both signed jump variation and bipower variation, $\overline{RV}_{h,i,t+h} = \mu_i + \phi^{J+}\Delta J^{2+} + \phi^{J-}\Delta J^{2-} + \phi^{C}BV^{-}_{i,t} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$.. The top panel contains the estimated parameters for the S&P 500 SPDR and the bottom panel contains the estimated parameters in the panel of individual firms. 95% confidence intervals are indicated using dashed lines.

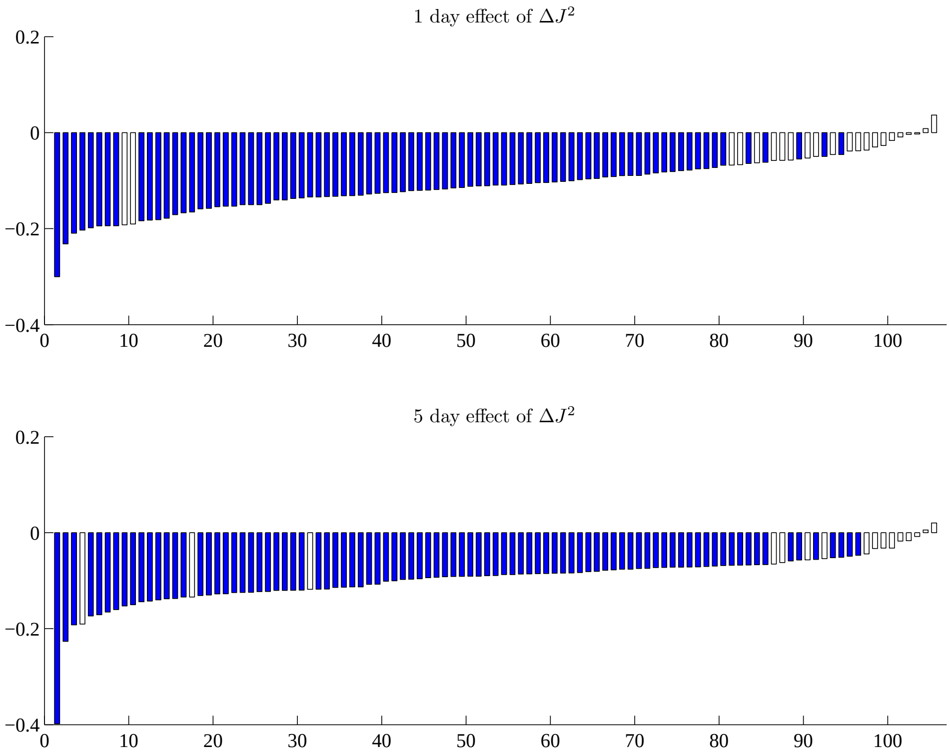

Figure 3: Effects of signed jump variation on individual firm volatilities, sorted by size. The magnitude of the coefficient on ΔJ^2 is indicated as distance from the horizontal axis. Solid bars indicate significance at the 5% level.

Data Summary Statistics

Averages

	SPDR	Mean	$Q_{.05}$	Median	$Q_{.95}$
RV	1.154	4.391	1.758	3.542	10.675
BV	1.131	3.821	1.540	2.999	9.521
RS^{+}	0.583	2.192	0.887	1.777	5.286
RS^{-}	0.571	2.199	0.874	1.806	5.388
ΔJ^{2}	0.012	-0.008	-0.210	0.011	0.107
ΔJ^{2+}	0.098	0.403	0.168	0.337	0.904
ΔJ^{2-}	-0.086	-0.411	-1.032	-0.334	-0.150

Autocorrelations

	SPDR	Mean	$Q_{.05}$	Median	$Q_{.95}$
RV	0.633	0.629	0.397	0.667	0.765
BV	0.682	0.637	0.420	0.658	0.788
RS^{+}	0.469	0.550	0.341	0.578	0.690
RS^{-}	0.704	0.592	0.340	0.624	0.757
ΔJ^{2}	-0.112	-0.013	-0.148	-0.003	0.092
ΔJ^{2+}	0.029	0.112	0.015	0.115	0.206
ΔJ^{2-}	0.062	0.133	0.055	0.127	0.276

Correlations

	RV	BV	RS^{+}	RS^{-}	ΔJ^{2}	ΔJ^{2+}	ΔJ^{2-}
RV	–	0.981	0.945	0.942	0.071	0.512	-0.472
BV	0.988	–	0.923	0.934	0.050	0.468	-0.452
RS^{+}	0.965	0.931	–	0.787	0.373	0.720	-0.217
RS^{-}	0.943	0.962	0.824	–	-0.252	0.238	-0.685
ΔJ^{2}	0.391	0.304	0.618	0.063	–	0.777	0.696
ΔJ^{2+}	0.613	0.520	0.782	0.338	0.909	–	0.122
ΔJ^{2-}	-0.340	-0.353	-0.148	-0.549	0.501	0.094	–

Table 1: The top panel contains the average values for realized variance (RV), bi-power variation (BV), positive and negative semivariance (RS^{+} and RS^{-}), jump variation (ΔJ^{2}) and signed jump-variation (ΔJ^{2+} and ΔJ^{2-}), scaled by 100. The left column contains values for the S&P 500 ETF (SPDR). The right four columns contain the average, 5% and 95% quantiles and the median from the panel of 105 stocks. The second panel contains the 1st autocorrelation for each of the series, and the right four columns report the average, 5% and 95% quantiles and median from the 105 individual stocks. The bottom panel contains the correlations for the seven variables; entries below the diagonal are computed using the SPDR data, and entries above the diagonal are average correlations for the 105 stocks.

HAR estimation results for the SPDR, cumulative volatility

$$\overline{RV}_{h,t+h} = \mu + \phi_d RV_t + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \gamma RV_t I_{[r_t<0]} + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$	0.607				0.268	0.120	0.532
	(17.0)				(8.1)	(4.9)	
		−0.024	1.182		0.291	0.120	0.611
		(−0.3)	(13.0)		(9.3)	(4.9)	
		0.037	1.064	0.050	0.293	0.121	0.611
		(0.4)	(7.4)	(1.4)	(9.3)	(5.0)	
$h=5$	0.425				0.409	0.158	0.563
	(14.7)				(8.6)	(4.0)	
		−0.030	0.862		0.421	0.155	0.620
		(−0.7)	(13.4)		(8.8)	(4.0)	
		0.073	0.650	0.092	0.424	0.157	0.619
		(1.1)	(6.6)	(2.5)	(8.8)	(4.0)	
$h=22$	0.305				0.357	0.265	0.468
	(11.8)				(7.7)	(4.8)	
		−0.009	0.628		0.359	0.261	0.508
		(−0.3)	(9.9)		(7.5)	(4.8)	
		−0.012	0.635	−0.003	0.359	0.261	0.508
		(−0.2)	(5.8)	(−0.1)	(7.6)	(4.8)	
$h=66$	0.203				0.256	0.299	0.282
	(8.4)				(7.4)	(5.3)	
		−0.067	0.501		0.253	0.294	0.313
		(−2.2)	(7.3)		(6.8)	(5.3)	
		−0.121	0.622	−0.054	0.251	0.292	0.315
		(−1.8)	(4.1)	(−1.3)	(6.8)	(5.2)	

Table 2a: Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used. The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data. Robust t-statistics are reported in parentheses.

HAR estimation results for the panel of 105 individual stocks, cumulative volatility

$$\overline{RV}_{h,i,t+h} = \mu_i + \phi_d RV_{i,t} + \phi_d^+ RS_{i,t}^+ + \phi_d^- RS_{i,t}^- + \gamma RV_{i,t} I_{[r_{i,t}<0]} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$	0.488				0.315	0.172	0.395
	(39.7)				(28.1)	(16.2)	
		0.268	0.704		0.317	0.172	0.398
		(15.6)	(24.5)		(28.7)	(16.4)	
		0.316	0.607	0.046	0.317	0.172	0.398
		(17.1)	(18.9)	(6.8)	(28.7)	(16.5)	
$h=5$	0.357				0.357	0.247	0.525
	(23.2)				(16.3)	(10.2)	
		0.158	0.551		0.359	0.247	0.529
		(11.4)	(19.0)		(16.4)	(10.3)	
		0.210	0.444	0.052	0.359	0.248	0.529
		(14.5)	(19.6)	(7.7)	(16.5)	(10.3)	
$h=22$	0.241				0.312	0.360	0.511
	(14.8)				(11.9)	(10.3)	
		0.091	0.388		0.314	0.360	0.513
		(7.3)	(12.8)		(11.9)	(10.4)	
		0.126	0.314	0.036	0.314	0.360	0.514
		(9.4)	(13.8)	(5.1)	(11.9)	(10.4)	
$h=66$	0.161				0.236	0.431	0.450
	(12.2)				(10.7)	(12.0)	
		0.044	0.275		0.238	0.431	0.452
		(3.9)	(8.8)		(10.6)	(12.1)	
		0.062	0.235	0.020	0.238	0.432	0.452
		(5.9)	(8.5)	(3.8)	(10.6)	(12.1)	

Table 2b: Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used. The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data. Robust t-statistics are reported in parentheses.

The impact of signed jump variation on future volatility, results for the SPDR

$$\overline{RM}_{h,t+h} = \mu + \phi_J \Delta J_t^2 + \phi_{J^+} \Delta J_t^{2+} + \phi_{J^-} \Delta J_t^{2-} + \phi_C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	RV				0.645 (17.5)	0.255 (7.8)	0.119 (4.9)	0.561
	RV	−0.572 (−7.7)			0.610 (18.4)	0.282 (9.0)	0.120 (5.0)	0.613
	RV		−0.190 (−2.1)	−0.964 (−5.5)	0.545 (16.8)	0.289 (9.4)	0.120 (5.0)	0.621
	BV	−0.549 (−9.5)			0.596 (20.4)	0.278 (10.2)	0.098 (5.0)	0.663
$h=5$	RV				0.466 (14.0)	0.389 (8.2)	0.156 (3.9)	0.584
	RV	−0.408 (−9.0)			0.449 (13.6)	0.406 (8.5)	0.154 (4.0)	0.622
	RV		−0.284 (−3.9)	−0.544 (−6.1)	0.426 (11.9)	0.409 (8.6)	0.154 (4.0)	0.622
	BV	−0.392 (−9.2)			0.440 (15.0)	0.389 (8.6)	0.137 (3.9)	0.633
$h=22$	RV				0.346 (11.7)	0.334 (7.1)	0.262 (4.8)	0.485
	RV	−0.276 (−7.3)			0.342 (11.2)	0.342 (7.1)	0.260 (4.8)	0.512
	RV		−0.299 (−3.8)	−0.248 (−2.3)	0.346 (10.0)	0.341 (7.0)	0.260 (4.8)	0.513
	BV	−0.264 (−7.0)			0.333 (10.9)	0.330 (6.8)	0.243 (4.7)	0.505
$h=66$	RV				0.237 (9.3)	0.237 (6.6)	0.296 (5.4)	0.290
	RV	−0.244 (−5.4)			0.240 (9.6)	0.240 (6.3)	0.293 (5.3)	0.312
	RV		−0.246 (−3.1)	−0.242 (−1.9)	0.240 (8.3)	0.240 (6.0)	0.293 (5.3)	0.312
	BV	−0.233 (−5.1)			0.232 (9.2)	0.231 (6.1)	0.279 (5.3)	0.304

Table 3a: Models that include signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). ΔJ_t^{2+} and ΔJ_t^{2-} decompose ΔJ_t^2 using an indicator variable for the sign of the difference where $\Delta J_t^{2+} = \Delta J_t^2 I_{[RS_t^+ - RS_t^- > 0]}$. The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

The impact of signed jump variation on future volatility, results for the panel of 105 individual stocks

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_J \Delta J^2_{i,t} + \phi_{J^+} \Delta J^{2+}_{i,t} + \phi_{J^-} \Delta J^{2-}_{i,t} + \phi_C BV_{i,t} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	RV				0.566 (38.6)	0.325 (28.6)	0.181 (17.0)	0.394
	RV	−0.215 (−10.5)			0.563 (40.0)	0.327 (29.1)	0.182 (17.2)	0.397
	RV		0.048 (2.5)	−0.492 (−12.2)	0.502 (38.1)	0.330 (29.5)	0.182 (17.3)	0.399
	BV	−0.178 (−10.9)			0.488 (41.9)	0.258 (28.3)	0.136 (16.2)	0.427
$h=5$	RV				0.414 (22.8)	0.364 (16.4)	0.255 (10.4)	0.524
	RV	−0.195 (−11.5)			0.411 (23.0)	0.366 (16.5)	0.255 (10.5)	0.527
	RV		−0.116 (−5.2)	−0.277 (−12.2)	0.392 (20.9)	0.367 (16.6)	0.255 (10.5)	0.527
	BV	−0.161 (−11.6)			0.358 (23.6)	0.293 (16.0)	0.196 (10.1)	0.538
$h=22$	RV				0.281 (13.7)	0.316 (12.1)	0.366 (10.3)	0.510
	RV	−0.146 (−8.7)			0.279 (13.9)	0.318 (12.1)	0.366 (10.4)	0.512
	RV		−0.122 (−5.3)	−0.172 (−7.8)	0.273 (13.0)	0.319 (12.1)	0.366 (10.4)	0.512
	BV	−0.120 (−8.6)			0.244 (13.8)	0.256 (11.4)	0.287 (10.0)	0.505
$h=66$	RV				0.182 (11.2)	0.242 (10.8)	0.437 (12.1)	0.448
	RV	−0.114 (−5.6)			0.180 (11.5)	0.244 (10.7)	0.437 (12.1)	0.450
	RV		−0.080 (−4.3)	−0.149 (−4.9)	0.171 (11.5)	0.245 (10.6)	0.437 (12.1)	0.450
	BV	−0.092 (−5.4)			0.158 (11.4)	0.194 (10.0)	0.348 (12.4)	0.439

Table 3b: Models that include signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). $\Delta J^{2+}_{i,t}$ and $\Delta J^{2-}_{i,t}$ decompose $\Delta J^2_{i,t}$ using an indicator variable for the sign of the difference where $\Delta J^{2+}_{i,t} = \Delta J^2_{i,t} I_{[RS^+_{i,t} - RS^-_{i,t} > 0]}$. The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

International Evidence

$$\overline{RV}_{h,t+h} = \mu + \phi_d RV_t + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \phi_J \Delta J_t^2 + \phi_C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

FTSE 100

	ϕ_d	ϕ_d^+	ϕ_d^-	ϕ_J	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	0.489					0.341	0.169	0.362
	(8.4)					(6.9)	(4.6)	
		0.183	0.708			0.361	0.180	0.386
		(2.2)	(8.0)			(7.2)	(5.0)	
				−0.248	0.579	0.344	0.184	0.412
				(−3.7)	(7.8)	(6.9)	(5.0)	
$h=5$	0.373					0.450	0.175	0.449
	(7.7)					(10.2)	(3.8)	
		0.138	0.541			0.466	0.184	0.472
		(2.9)	(7.2)			(10.3)	(4.0)	
				−0.186	0.466	0.441	0.183	0.496
				(−4.4)	(6.8)	(10.2)	(3.9)	
$h=22$	0.250					0.366	0.333	0.369
	(8.0)					(5.6)	(4.5)	
		0.034	0.406			0.380	0.340	0.392
		(0.8)	(6.3)			(5.7)	(4.7)	
				−0.180	0.299	0.367	0.341	0.401
				(−3.6)	(7.1)	(5.4)	(4.7)	
$h=66$	0.171					0.308	0.292	0.245
	(6.5)					(5.7)	(3.9)	
		0.029	0.277			0.317	0.296	0.260
		(0.7)	(4.4)			(5.6)	(4.0)	
				−0.124	0.198	0.313	0.298	0.268
				(−2.7)	(4.4)	(5.6)	(3.9)	

EURO STOXX 50

	ϕ_d	ϕ_d^+	ϕ_d^-	ϕ_J	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	0.515					0.334	0.110	0.625
	(12.9)					(10.4)	(4.3)	
		0.209	0.727			0.361	0.119	0.640
		(3.7)	(12.6)			(11.1)	(4.7)	
				−0.227	0.633	0.332	0.105	0.639
				(−5.5)	(12.6)	(10.2)	(4.2)	
$h=5$	0.448					0.392	0.109	0.637
	(6.7)					(8.5)	(2.8)	
		0.243	0.582			0.412	0.116	0.645
		(3.6)	(7.0)			(8.8)	(3.0)	
				−0.148	0.558	0.383	0.105	0.638
				(−3.9)	(6.4)	(7.6)	(2.8)	
$h=22$	0.299					0.365	0.232	0.508
	(6.5)					(5.0)	(3.7)	
		0.165	0.386			0.379	0.236	0.513
		(3.2)	(5.5)			(5.1)	(3.8)	
				−0.101	0.366	0.362	0.230	0.505
				(−2.2)	(7.8)	(4.7)	(3.7)	
$h=66$	0.221					0.304	0.210	0.383
	(4.6)					(3.9)	(3.3)	
		0.122	0.284			0.315	0.213	0.388
		(2.4)	(3.5)			(3.8)	(3.3)	
				−0.071	0.278	0.298	0.207	0.379
				(−1.2)	(5.5)	(3.5)	(3.2)	

Table 4: Models fit on international equity market realized variance and semivariance. The top panel contains models estimated on the realized variance of the FTSE 100 and the bottom contains results from the EURO STOXX 50. Models were fit to 1-day, 1-week, 1-month and 1-quarter ahead cumulative realized variance. Each subpanel contains 3 models: the first is a reference HAR which only includes realized variance, the second decomposes realized variance into positive and negative realized semivariance and the third splits recent volatility in to a signed-jump measure, ΔJ^2, and bi-power variation (BV), a estimate of the continuous component of variance.

Out-of-sample forecast comparison of models for volatility forecasting

	$\widehat{RV}^{HAR} - \widehat{RV}^{RS}$			$\widehat{RV}^{HAR} - \widehat{RV}^{RS^-}$			$\widehat{RV}^{GJR} - \widehat{RV}^{RS^-}$		
	SPDR	**Other Assets**		**SPDR**	**Other Assets**		**SPDR**	**Other Assets**	
	DM	Favor HAR	Favor RS	DM	Favor HAR	Favor RS^-	DM	Favor GJR	Favor RS^-
$h=1$	1.48	2.1	22.1	2.69	3.2	12.6	0.02	9.5	2.1
$h=5$	3.16	2.1	29.5	4.29	5.3	33.7	1.25	3.2	10.5
$h=22$	3.96	2.1	26.3	4.41	1.1	33.7	3.07	2.1	16.8
$h=66$	3.64	1.1	23.2	5.23	1.1	36.8	4.27	3.2	16.8

	$\widehat{RV}^{BV} - \widehat{RV}^{RS^-}$			$\widehat{RV}^{BV} - \widehat{RV}^{\Delta J^2}$			$\widehat{RV}^{BV} - \widehat{RV}^{\Delta J^{2\pm}}$		
	SPDR	**Other Assets**		**SPDR**	**Other Assets**		**SPDR**	**Other Assets**	
	DM	Favor BV	Favor RS^-	DM	Favor BV	Favor ΔJ^2	DM	Favor BV	Favor $\Delta J^{2\pm}$
$h=1$	2.21	7.4	8.4	-0.08	2.1	16.8	1.53	1.1	15.8
$h=5$	3.37	7.4	25.3	2.15	2.1	25.3	3.27	1.1	25.3
$h=22$	3.33	3.2	30.5	3.10	3.2	18.9	3.22	1.1	12.6
$h=66$	4.13	1.1	45.3	1.41	1.1	17.9	3.59	2.1	16.8

Table 5: Each of the panels contains results for tests of equal predictive accuracy. The left most column contains the Diebold-Mariano-Giacomini-White test statistic for the S&P 500 SPDR where a positive test statistic indicates the realized semivariance model (or model using jump variation) outperformed the realized variance (or bipower variation) model. The remaining two columns report the percentage of the 95 individual loss differentials which reject the null, and the direction of the rejection using a 5% 2-sided test.

Out-of-sample R^2

SPDR

	$\widehat{RV}^{HAR}$	$\widehat{RV}^{GJR}$	$\widehat{RV}^{RS}$	$\widehat{RV}^{RS^-}$	$\widehat{RV}^{BV}$	$\widehat{RV}^{\Delta J^2}$	$\widehat{RV}^{\Delta J^{2\pm}}$
$h=1$	66.7	69.0	67.8	68.8	68.0	68.9	**69.3**
$h=5$	64.8	67.8	67.8	67.6	65.6	67.7	**67.8**
$h=22$	52.4	53.0	**54.1**	53.9	53.0	54.1	54.1
$h=66$	42.8	43.0	43.8	43.7	43.1	43.7	**43.9**

Individual

	$\widehat{RV}^{HAR}$	$\widehat{RV}^{GJR}$	$\widehat{RV}^{RS}$	$\widehat{RV}^{RS^-}$	$\widehat{RV}^{BV}$	$\widehat{RV}^{\Delta J^2}$	$\widehat{RV}^{\Delta J^{2\pm}}$
$h=1$	40.4	51.5	50.2	50.5	46.8	**53.9**	53.0
$h=5$	59.1	61.1	61.6	61.8	60.7	**62.6**	61.3
$h=22$	55.1	55.4	55.8	**56.0**	55.5	56.0	55.8
$h=66$	51.7	51.0	51.8	**52.2**	51.7	51.8	51.3

Table 6: Out-of-sample R^2 for the alternative models used in the forecast evaluation. The OOS R^2 is computed as one minus the ratio of out-of-sample model-based MSE to the out-of-sample MSE from a forecast that only includes a constant. The largest value in each row is in bold.

Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility Internet Appendix

November 16, 2013

A Data Cleaning

Only transaction data were taken from the NYSE TAQ. All series were automatically cleaned according to a set of six rules:

1. Transactions outside of 9:30:00 and 16:00:00 were discarded.
2. Transactions with zero price or volume were discarded.
3. Each day the most active exchange was determined. Only transactions from this exchange were retained.
4. Only trades with conditions E, F or blank were retained.
5. Transaction prices outside of the CRSP high or low for the day were discarded.
6. Trade with immediate reversals more than 5 times a 50-sample moving window - excluding the transaction being tested - were discarded.

These rules are similar to those of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009), and prices were not manually cleaned for problems not addressed by these rules.

B Completely Decomposing Quadratic Variation

The specification in eq. (16) of the paper restricts the coefficients on the weekly and monthly realized semivariance to be identical. This restriction can be relaxed by decomposing the RV terms at all lags.[A.1] With this modification we obtain:

$$\overline{RV}_{h,t+h} = \mu + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \phi_w^+ \overline{RS}_{w,t}^+ + \phi_w^- \overline{RS}_{w,t}^- + \phi_m^+ \overline{RS}_{m,t}^+ + \phi_m^- \overline{RS}_{m,t}^- + \epsilon_{t+h}. \quad \text{(A.1)}$$

Results from this extended specification are presented in the first row of each column of Table A.1a. In both sets of results, those using the SPDR and those based on the panel, the negative semivariance dominates the positive semivariance. In the models using the SPDR, the coefficients on positive semivariance are always either significantly negative or insignificantly different from zero. The coefficient on the terms involving negative semivariance are uniformly positive and significant. In the panel estimation the same general pattern appears, although some of the coefficients on positive semivariance, especially at short horizons and lag 1, are significantly positive. Interestingly, as the horizon increases, the persistence of the volatility in the panel shifts to the negative semivariance, particularly at the longer lags.

Decomposing realized variance at all lags allows us to consider a "Vector HAR" (VHAR) for the two semivariances. Such a model allows us to determine whether lagged realized semivariances of the same sign as the dependent variable are more useful than lagged semivariances of the opposite sign.

$$\begin{bmatrix} \overline{RS}_{h,t+h}^+ \\ \overline{RS}_{h,t+h}^- \end{bmatrix} = \begin{bmatrix} \mu^+ \\ \mu^- \end{bmatrix} + \boldsymbol{\phi}_d \begin{bmatrix} RS_t^+ \\ RS_t^- \end{bmatrix} + \boldsymbol{\phi}_w \begin{bmatrix} \overline{RS}_{w,t}^+ \\ \overline{RS}_{w,t}^- \end{bmatrix} + \boldsymbol{\phi}_m \begin{bmatrix} \overline{RS}_{m,t}^+ \\ \overline{RS}_{m,t}^- \end{bmatrix} + \begin{bmatrix} \epsilon_{t+h}^+ \\ \epsilon_{t+h}^- \end{bmatrix} \quad \text{(A.2)}$$

where

$$\boldsymbol{\phi}_j = \begin{bmatrix} \phi_{j+}^+ & \phi_{j+}^- \\ \phi_{j-}^+ & \phi_{j-}^- \end{bmatrix}, \quad j \in \{d, w, m\}$$

Results of the VHAR are presented in the two lower rows of each panel of Tables A.1a and A.1b. The estimates for each equation of the VHAR are virtually identical, with small or negative coefficients on lagged

[A.1] When a jump variable is also included, Chen and Ghysels (2011) call this model the "HAR-S-RV-J" model, and use it as one of the benchmarks in their study.

positive semivariance and large, significant coefficients on lagged negative semivariance. The results in the panel are similar with parameter estimates on negative semivariance uniformly large and highly significant. Thus negative semivariance is useful for predicting *both* positive and negative future semivariance. This is a novel and somewhat surprising result.

This leads us to test whether positive semivariance is actually needed in the VHAR models. We perform these tests on the individual models for the SPDR and the 105 constituent volatility series. The first null hypothesis is that positive semivariance can be excluded, $H_0: \phi_d^+ = \phi_w^+ = \phi_m^+ = 0$, and the other tests whether negative semivariance can be excluded. We find that positive semivariance can be excluded from 19.1.% of the joint (bivariate VHAR) models, 19.1% of the models for positive semivariance, and 25.7% of the models for negative semivariance. Negative semivariance can only be excluded in 1 of the 105 volatility series. Thus, while most of the predictability for future semivariance appears to come from lagged negative semivariance, the lagged positive semivariance also carries some information.

We next test whether the sum of the coefficients on the positive semivariance is equal to the sum of the coefficients on the negative semivariance, $H_0 : \phi_d^+ + \phi_w^+ + \phi_m^+ = \phi_d^- + \phi_w^- + \phi_m^-$ in each of the two semivariance models. This null can be rejected for all but 3 (25) of the 105 positive (negative) semivariance models, and, when rejected, the sum of the coefficients on the negative semivariance is larger than the sum of the coefficient on the positive semivariance in all but two cases. We next test for equality only at the first lag, $H_0 : \phi_d^+ = \phi_d^-$. This null is rejected in 61% of the positive semivariance models and 78% of the negative semivariance models, and when rejected typically indicates that the coefficient on negative semivariance is larger than the coefficient on positive semivariance. Thus both of these sets of hypothesis tests reveal that negative semivariances have greater weight in these predictive models for almost all assets considered here.

Finally, we test whether the persistence of each series, as measured by the maximum eigenvalue of the companion form of a HAR, is equal for the two semivariances. This is done by restricting the off-diagonal elements in eq. A.2 to be zero, and estimating the remaining parameters and the (joint) asymptotic covariance matrix. The asymptotic distribution is used to simulate 1,000 draws of the parameters, each one is then transformed into companion form, and the maximum eigenvalue of the companion matrix is com-

Forecasting measures of future volatility using realized semivariances, results for the SPDR

$$\overline{RM}_{h,t+h} = \mu + \phi_d^+ RS_t^+ + \phi_w^+ \overline{RS}^+_{w,t} + \phi_m^+ \overline{RS}^+_{m,t} + \phi_d^- RS_t^- + \phi_w^- \overline{RS}^-_{w,t} + \phi_m^- \overline{RS}^-_{m,t} + \epsilon_{t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=1$	RV	−0.087	−0.179	−0.233	1.162	0.788	0.536	0.621
		(−1.3)	(−2.5)	(−1.9)	(12.7)	(9.1)	(3.4)	
	RS^+	−0.084	−0.110	−0.069	0.644	0.404	0.219	0.539
		(−1.9)	(−2.9)	(−1.0)	(10.0)	(8.1)	(2.3)	
	RS^-	−0.003	−0.069	−0.163	0.518	0.384	0.317	0.621
		(−0.1)	(−1.7)	(−2.3)	(14.9)	(8.0)	(3.6)	
$h=5$	RV	−0.070	0.011	−0.474	0.836	0.815	0.880	0.642
		(−1.5)	(0.1)	(−2.4)	(13.4)	(5.9)	(3.7)	
	RS^+	−0.049	0.015	−0.207	0.447	0.405	0.403	0.635
		(−1.9)	(0.2)	(−2.0)	(12.6)	(5.8)	(3.2)	
	RS^-	−0.022	−0.004	−0.267	0.389	0.410	0.477	0.623
		(−0.9)	(−0.1)	(−2.6)	(13.6)	(5.9)	(3.9)	
$h=22$	RV	−0.058	−0.217	−0.920	0.583	0.907	1.607	0.565
		(−1.3)	(−1.3)	(−1.8)	(10.3)	(5.9)	(2.6)	
	RS^+	−0.031	−0.107	−0.494	0.301	0.451	0.851	0.583
		(−1.3)	(−1.2)	(−1.9)	(10.2)	(5.5)	(2.6)	
	RS^-	−0.027	−0.110	−0.425	0.282	0.457	0.756	0.538
		(−1.3)	(−1.4)	(−1.7)	(10.3)	(6.1)	(2.5)	
$h=66$	RV	−0.119	−0.411	−0.835	0.447	0.932	1.580	0.364
		(−2.7)	(−2.5)	(−1.3)	(7.9)	(5.0)	(2.3)	
	RS^+	−0.064	−0.218	−0.444	0.235	0.488	0.819	0.378
		(−2.8)	(−2.5)	(−1.3)	(8.0)	(5.0)	(2.3)	
	RS^-	−0.055	−0.193	−0.392	0.212	0.444	0.761	0.347
		(−2.7)	(−2.4)	(−1.3)	(7.8)	(5.0)	(2.4)	

Table A.1a: Extended model where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-). The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

puted. The null is tested using the percentage of times where $\lambda^- > \lambda^+$, were λ^+ is the largest eigenvalue from the companion matrix for HAR for positive semivariance. The null is rejected if λ^- is greater than λ^+ in more than 97.5% or in less than 2.5% of the simulations. quality is rejected in 89.6% of the series, and negative semivariance is found to be more persistent in 91.6% of the rejections.

These results indicate that negative semivariance is more useful for predicting realized variance and *both* realized semivariances, and that negative semivariance is more persistent than positive semivariance.

Forecasting measures of future volatility using realized semivariances, results for the panel of 105 individual stocks.

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_d^+ RS_{i,t}^+ + \phi_w^+ \overline{RS}_{w,i,t}^+ + \phi_m^+ \overline{RS}_{m,i,t}^+ + \phi_d^- RS_{i,t}^- + \phi_w^- \overline{RS}_{w,i,t}^- + \phi_m^- \overline{RS}_{m,i,t}^- + \epsilon_{i,t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=1$	RV	0.262	0.105	0.065	0.696	0.529	0.290	0.399
		(15.3)	(5.5)	(2.5)	(24.3)	(25.1)	(11.0)	
	RS^+	0.144	0.061	0.058	0.339	0.253	0.118	0.380
		(14.5)	(6.0)	(4.1)	(18.3)	(22.6)	(8.3)	
	RS^-	0.118	0.045	0.007	0.357	0.277	0.172	0.309
		(13.5)	(4.1)	(0.5)	(28.7)	(23.4)	(11.1)	
$h=5$	RV	0.152	0.136	0.033	0.542	0.577	0.479	0.531
		(10.7)	(3.7)	(0.5)	(19.2)	(16.5)	(7.9)	
	RS^+	0.079	0.079	0.038	0.272	0.279	0.214	0.536
		(11.0)	(4.3)	(1.3)	(18.0)	(16.4)	(7.6)	
	RS^-	0.073	0.058	−0.005	0.270	0.298	0.265	0.472
		(10.0)	(3.1)	(−0.2)	(20.0)	(15.9)	(7.8)	
$h=22$	RV	0.085	0.057	0.022	0.377	0.561	0.724	0.517
		(6.4)	(1.3)	(0.2)	(13.2)	(11.9)	(5.2)	
	RS^+	0.045	0.035	0.015	0.189	0.275	0.358	0.524
		(6.8)	(1.6)	(0.3)	(12.8)	(11.9)	(5.2)	
	RS^-	0.039	0.022	0.008	0.188	0.286	0.366	0.494
		(5.9)	(1.0)	(0.1)	(13.5)	(11.8)	(5.2)	
$h=66$	RV	0.038	−0.001	0.119	0.266	0.467	0.770	0.455
		(2.9)	(−0.0)	(0.7)	(9.4)	(7.1)	(4.7)	
	RS^+	0.020	0.004	0.066	0.134	0.233	0.378	0.460
		(3.1)	(0.1)	(0.8)	(9.2)	(7.0)	(4.5)	
	RS^-	0.017	−0.005	0.053	0.132	0.235	0.392	0.444
		(2.6)	(−0.2)	(0.7)	(9.5)	(7.2)	(4.9)	

Table A.1b: Extended model where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-). The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

C Sign of Past Returns

We consider an extended model similar to the model presented in table 2a and 2b in the main paper where an indicator variable for a negative return on the previous day was included as a level-shift in addition to the interaction with lagged realized variance. Tables A.2a and A.2b contain the results from this additional specification applied to the SPDR and the panel of firm volatilities, respectively. The indicator variable typically has a negative sign indicating that level of volatility is lower subsequent to a negative shock when compared to a model without this term, although the response to the interaction term with lagged realized variance increases slightly and so the net effect of a negative return is still likely positive. The coefficients on the realized semivariance, RS^{+} and RS^{-}, are virtually unchanged by this modification.

Extended HAR estimation results for the SPDR, cumulative volatility

$$\overline{RV}_{h,t+h} = \mu + \delta I_{[r_{t-1}<0]} + \phi_d RV_t + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \gamma RV_t I_{[r_t<0]} + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

	δ	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$		0.607				0.268	0.120	0.532
		(17.0)				(8.1)	(4.9)	
			−0.024	1.182		0.291	0.120	0.611
			(−0.3)	(13.0)		(9.3)	(4.9)	
			0.037	1.064	0.050	0.293	0.121	0.611
			(0.4)	(7.4)	(1.4)	(9.3)	(5.0)	
	−4.185		0.004	1.052	0.100	0.291	0.121	0.617
	(−2.0)		(0.0)	(7.4)	(2.3)	(9.2)	(5.0)	
$h=5$		0.425				0.409	0.158	0.563
		(14.7)				(8.6)	(4.0)	
			−0.030	0.862		0.421	0.155	0.620
			(−0.7)	(13.4)		(8.8)	(4.0)	
			0.073	0.650	0.092	0.424	0.157	0.619
			(1.1)	(6.6)	(2.5)	(8.8)	(4.0)	
	−4.034		0.049	0.639	0.132	0.422	0.157	0.621
	(−2.1)		(0.8)	(6.5)	(2.8)	(8.7)	(4.1)	
$h=22$		0.305				0.357	0.265	0.468
		(11.8)				(7.7)	(4.8)	
			−0.009	0.628		0.359	0.261	0.508
			(−0.3)	(9.9)		(7.5)	(4.8)	
			−0.012	0.635	−0.003	0.359	0.261	0.508
			(−0.2)	(5.8)	(−0.1)	(7.6)	(4.8)	
	−0.656		−0.015	0.633	0.003	0.359	0.261	0.508
	(−0.4)		(−0.3)	(5.7)	(0.1)	(7.6)	(4.8)	
$h=66$		0.203				0.256	0.299	0.282
		(8.4)				(7.4)	(5.3)	
			−0.067	0.501		0.253	0.294	0.313
			(−2.2)	(7.3)		(6.8)	(5.3)	
			−0.121	0.622	−0.054	0.251	0.292	0.315
			(−1.8)	(4.1)	(−1.3)	(6.8)	(5.2)	
	1.666		−0.116	0.627	−0.065	0.252	0.292	0.315
	(0.6)		(−1.9)	(4.0)	(−1.2)	(6.8)	(5.2)	

Table A.2a: All models use the h-day cumulative variance as the dependent variable (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 4 models: the first model is a standard RV HAR, the second decomposes realized variance into positive and negative realized semivariance at the first lag, the third specification adds an asymmetric term where the sign of the most recent daily return is used (these are repeated from Table 2a of the main paper) and the final model augments the third with an indicator variable as a level-shift. The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

Extended HAR estimation results for the panel of 105 individual stocks, cumulative volatility

$$\overline{RV}_{h,i,t+h} = \mu_i + \delta I_{[r_{i,t-1}<0]} + \phi_d RV_{t,i} + \phi_d^+ RS_{i,t}^+ + \phi_d^- RS_{i,t}^- + \gamma RV_{i,t} I_{[r_{i,t}<0]} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	δ	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$		0.488 (39.7)				0.315 (28.1)	0.172 (16.2)	0.395
			0.268 (15.6)	0.704 (24.5)		0.317 (28.7)	0.172 (16.4)	0.398
			0.316 (17.1)	0.607 (18.9)	0.046 (6.8)	0.317 (28.7)	0.172 (16.5)	0.398
	−3.306 (−2.8)		0.311 (16.4)	0.597 (19.3)	0.061 (6.8)	0.317 (28.8)	0.173 (16.5)	0.399
$h=5$		0.357 (23.2)				0.357 (16.3)	0.247 (10.2)	0.525
			0.158 (11.4)	0.551 (19.0)		0.359 (16.4)	0.247 (10.3)	0.529
			0.210 (14.5)	0.444 (19.6)	0.052 (7.7)	0.359 (16.5)	0.248 (10.3)	0.529
	−3.382 (−2.5)		0.206 (14.4)	0.435 (20.1)	0.065 (6.3)	0.359 (16.4)	0.248 (10.4)	0.529
$h=22$		0.241 (14.8)				0.312 (11.9)	0.360 (10.3)	0.511
			0.091 (7.3)	0.388 (12.8)		0.314 (11.9)	0.360 (10.4)	0.513
			0.126 (9.4)	0.314 (13.8)	0.036 (5.1)	0.314 (11.9)	0.360 (10.4)	0.514
	−0.780 (−0.6)		0.125 (9.4)	0.313 (13.8)	0.039 (4.5)	0.314 (11.9)	0.360 (10.4)	0.514
$h=66$		0.161 (12.2)				0.236 (10.7)	0.431 (12.0)	0.450
			0.044 (3.9)	0.275 (8.8)		0.238 (10.6)	0.431 (12.1)	0.452
			0.062 (5.9)	0.235 (8.5)	0.020 (3.8)	0.238 (10.6)	0.432 (12.1)	0.452
	0.731 (0.5)		0.063 (6.0)	0.237 (8.3)	0.018 (3.2)	0.238 (10.6)	0.431 (12.1)	0.452

Table A.2b: All models use the h-day cumulative variance as the dependent variable (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 4 models: the first model is a standard RV HAR, the second decomposes realized variance into positive and negative realized semivariance at the first lag, the third specification adds an asymmetric term where the sign of the most recent daily return is used (these are repeated from Table 2b of the main paper) and the final model augments the third with an indicator variable as a level-shift. The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

D Model in Logs

The models in the paper, as well as the VHAR in this appendix, are all estimated in levels using weighted least squares. As an alternative, we estimate the same models using the log of realized variance and realized semivariance with ordinary least squares. The log transformation naturally eliminates the primary source of heteroskedasticity in realized variance and related estimators and so there is no need for weighting. The signed jump variation measures are not always positive, and so these were replaced with percentage jump variation measures defined below.

Level Parameterization	Log Parameterization
ΔJ_t^2	$\%\Delta J_t^2 = \ln\left(1 + \Delta J_t^2 / RV_t\right)$
ΔJ_t^{2+}	$\%\Delta J_t^{2+} = \ln\left(1 + \Delta J_t^2 / RV_t\right) I_{[\Delta J_t^2 > 0]}$
ΔJ_t^{2-}	$\%\Delta J_t^{2-} = \ln\left(1 + \Delta J_t^2 / RV_t\right) I_{[\Delta J_t^2 < 0]}$

Results for SPDR

Tables A.3a, A.4a and A.5a correspond to Tables 2a and 3a from the main paper, and Table A.1a from this internet appendix. The results in terms of the sign, relative magnitude and statistical significance are remarkably similar. The magnitudes of the coefficients differ since the exact decomposition of RV into RS^+ and RS^- no longer holds, and the log eliminates scale differences between the realized variance and the realized semivariance. Table A.4a contains the models built using the modified ΔJ^2 measures, which continue to have uniformly negative signs indicating that negative jumps increase volatility and positive jumps decrease volatility. The main conclusions, that negative semivariance is extremely informative about future volatility, are strengthened with these results.

Results for the Panel

Tables A.3b, A.4b and A.5b corresponds to Tables 2b and 3b from the main paper, and Table A.1b from this internet appendix. Like the results for the SPDR, these are virtually identical in terms of the relative magnitude of the coefficients and statistical significance, and serve to reinforce the importance of decomposed volatility estimators in forecasting volatility. In particular, Table A.4b confirms that the sign on $\%\Delta J^2$ is negative and statistically significant at all horizons.

HAR estimation results for the SPDR, log cumulative volatility

$$\ln \overline{RV}_{h,t+h} = \mu + \phi_d \ln RV_t + \phi_d^+ \ln RS_t^+ + \phi_d^- \ln RS_t^- + \gamma \ln RV_t I_{[r_t<0]} + \phi_w \ln \overline{RV}_{w,t} + \phi_m \ln \overline{RV}_{m,t} + \epsilon_{t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$	0.520				0.309	0.123	0.778
	(24.0)				(11.9)	(5.9)	
		0.044	0.446		0.335	0.124	0.786
		(2.0)	(20.9)		(13.4)	(6.0)	
		0.116	0.361	0.021	0.336	0.125	0.787
		(4.0)	(11.8)	(4.0)	(13.4)	(6.1)	
$h=5$	0.390				0.356	0.176	0.798
	(19.6)				(10.2)	(5.3)	
		0.033	0.336		0.374	0.176	0.803
		(1.9)	(17.1)		(10.8)	(5.4)	
		0.078	0.283	0.013	0.375	0.177	0.804
		(3.0)	(10.1)	(2.7)	(10.7)	(5.4)	
$h=22$	0.291				0.311	0.241	0.717
	(13.6)				(8.8)	(5.3)	
		0.036	0.243		0.322	0.240	0.720
		(2.2)	(11.3)		(9.1)	(5.4)	
		0.039	0.238	0.001	0.322	0.240	0.720
		(1.5)	(7.3)	(0.2)	(9.1)	(5.4)	
$h=66$	0.212				0.232	0.299	0.618
	(11.0)				(8.7)	(5.5)	
		0.012	0.190		0.241	0.299	0.621
		(0.6)	(7.1)		(8.5)	(5.5)	
		−0.007	0.213	−0.006	0.241	0.298	0.621
		(−0.2)	(4.5)	(−0.9)	(8.5)	(5.5)	

Table A.3a: Reference, base and asymmetric model parameter estimates using log h-day cumulative variance as the dependent variable (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used.

HAR estimation results for the panel of 105 individual stocks, log cumulative volatility

$$\ln \overline{RV}_{h,i,t+h} = \mu_i + \phi_d \ln RV_{i,t} + \phi_d^+ \ln RS_{i,t}^+ + \phi_d^- \ln RS_{i,t}^- + \gamma \ln RV_{i,t} I_{[r_{i,t}<0]} + \phi_w \ln \overline{RV}_{w,i,t} + \phi_m \ln \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=1$	0.438				0.320	0.195	0.724
	(59.1)				(34.9)	(19.5)	
		0.140	0.296		0.321	0.195	0.725
		(23.1)	(37.3)		(35.4)	(19.6)	
		0.174	0.257	0.008	0.321	0.195	0.726
		(28.7)	(33.4)	(10.7)	(35.5)	(19.8)	
$h=5$	0.323				0.346	0.257	0.779
	(32.4)				(20.8)	(12.6)	
		0.085	0.238		0.346	0.256	0.780
		(16.4)	(25.9)		(21.0)	(12.7)	
		0.116	0.203	0.008	0.346	0.257	0.781
		(19.3)	(26.2)	(9.4)	(21.1)	(12.7)	
$h=22$	0.232				0.300	0.329	0.735
	(18.8)				(16.3)	(11.1)	
		0.053	0.179		0.300	0.329	0.737
		(8.4)	(16.1)		(16.4)	(11.1)	
		0.075	0.154	0.005	0.300	0.329	0.737
		(10.2)	(16.8)	(6.0)	(16.4)	(11.2)	
$h=66$	0.166				0.237	0.387	0.676
	(18.4)				(16.7)	(12.5)	
		0.031	0.134		0.237	0.387	0.677
		(4.8)	(11.1)		(16.6)	(12.6)	
		0.045	0.117	0.004	0.238	0.387	0.677
		(7.0)	(11.8)	(3.7)	(16.6)	(12.6)	

Table A.3b: Reference, base and asymmetric model parameter estimates using log h-day cumulative variance as the dependent variance (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used. The reported R^2 measure is the average of the 105 R^2s for the individual series.

The impact of signed jump variation on future log volatility, results for the SPDR

$$\ln \overline{RM}_{h,t+h} = \mu + \phi_J \%\Delta J_t^2 + \phi_{J^+} \%\Delta J_t^{2+} + \phi_{J^-} \%\Delta J_t^{2-} + \phi_C \ln BV_t + \phi_w \ln \overline{RV}_{w,t} + \phi_m \ln \overline{RV}_{m,t} + \epsilon_{t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	RV				0.523 (25.2)	0.297 (11.6)	0.124 (6.0)	0.780
	RV	−0.412 (−10.5)			0.489 (24.5)	0.332 (13.3)	0.124 (6.1)	0.788
	RV		−0.215 (−2.6)	−0.539 (−7.9)	0.488 (24.5)	0.335 (13.5)	0.124 (6.1)	0.789
	BV	−0.419 (−10.7)			0.505 (25.4)	0.335 (13.5)	0.120 (6.0)	0.794
$h=5$	RV				0.393 (19.4)	0.346 (9.7)	0.177 (5.4)	0.799
	RV	−0.303 (−9.7)			0.368 (18.3)	0.372 (10.6)	0.176 (5.4)	0.804
	RV		−0.234 (−3.4)	−0.347 (−6.7)	0.368 (18.2)	0.373 (10.6)	0.177 (5.4)	0.804
	BV	−0.306 (−9.6)			0.381 (19.2)	0.377 (10.8)	0.171 (5.3)	0.807
$h=22$	RV				0.296 (12.6)	0.301 (8.6)	0.241 (5.3)	0.718
	RV	−0.197 (−6.3)			0.279 (11.8)	0.318 (9.0)	0.241 (5.4)	0.721
	RV		−0.234 (−3.2)	−0.174 (−2.7)	0.280 (11.8)	0.318 (8.9)	0.240 (5.4)	0.721
	BV	−0.199 (−6.2)			0.287 (11.8)	0.322 (8.8)	0.239 (5.2)	0.718
$h=66$	RV				0.218 (10.0)	0.223 (8.2)	0.299 (5.4)	0.619
	RV	−0.167 (−3.7)			0.204 (9.7)	0.237 (8.3)	0.299 (5.5)	0.621
	RV		−0.252 (−2.8)	−0.111 (−1.1)	0.205 (9.6)	0.236 (8.1)	0.298 (5.5)	0.621
	BV	−0.169 (−3.6)			0.208 (9.7)	0.239 (8.2)	0.301 (5.5)	0.619

Table A.4a: Models estimated in logs that includes signed jump information where quadratic variation has been decomposed into signed jump variation, $\%\Delta J^2 = \ln\left(1 + \Delta J_t^2 / RV_t\right)$, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). $\%\Delta J_t^{2+}$ and $\%\Delta J_t^{2-}$ decompose $\%\Delta J_t^2$ using an indicator variable for the sign of the difference where $\Delta J_{i,t}^{2+} = \%\Delta J_{i,t}^2 I_{[RS_t^+ - RS_t^- > 0]}$.

The impact of signed jump variation on future log volatility, results for the panel of 105 individual stocks

$$\ln \overline{RM}_{h,i,t+h} = \mu_i + \phi_J \%\Delta J^2_{i,t} + \phi_{J^+} \%\Delta J^{2+}_{i,t} + \phi_{J^-} \%\Delta J^{2-}_{i,t} + \phi_C \ln BV_{i,t} + \phi_w \ln \overline{RV}_{w,i,t} + \phi_m \ln \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=1$	RV				0.421 (58.0)	0.330 (35.6)	0.206 (20.5)	0.723
	RV	−0.177 (−14.6)			0.418 (59.3)	0.332 (36.4)	0.206 (20.8)	0.725
	RV		0.036 (2.0)	−0.308 (−18.0)	0.415 (59.0)	0.334 (36.7)	0.206 (20.9)	0.726
	BV	−0.171 (−13.8)			0.437 (60.8)	0.322 (34.8)	0.186 (18.7)	0.712
$h=5$	RV				0.305 (31.3)	0.357 (21.1)	0.266 (12.9)	0.777
	RV	−0.162 (−14.9)			0.303 (31.5)	0.359 (21.4)	0.266 (13.0)	0.779
	RV		−0.104 (−5.3)	−0.197 (−15.3)	0.302 (31.3)	0.360 (21.4)	0.266 (13.0)	0.779
	BV	−0.159 (−14.5)			0.317 (32.8)	0.351 (21.1)	0.249 (12.4)	0.772
$h=22$	RV				0.211 (17.3)	0.313 (16.8)	0.338 (11.2)	0.733
	RV	−0.133 (−10.2)			0.209 (17.4)	0.315 (16.9)	0.338 (11.3)	0.734
	RV		−0.088 (−4.0)	−0.161 (−10.1)	0.208 (17.3)	0.315 (16.9)	0.338 (11.3)	0.734
	BV	−0.130 (−10.1)			0.220 (17.9)	0.308 (16.2)	0.321 (10.8)	0.728
$h=66$	RV				0.140 (14.9)	0.254 (16.8)	0.395 (12.6)	0.673
	RV	−0.110 (−6.4)			0.138 (15.1)	0.255 (16.8)	0.395 (12.6)	0.674
	RV		−0.058 (−2.3)	−0.143 (−6.5)	0.138 (15.1)	0.256 (16.7)	0.395 (12.6)	0.674
	BV	−0.107 (−6.2)			0.146 (15.3)	0.247 (15.9)	0.382 (12.6)	0.668

Table A.4b: Models estimated in logs that includes signed jump information where quadratic variation has been decomposed into signed jump variation, $\%\Delta J^2 = \ln\left(1 + \Delta J_t^2/RV_t\right)$, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). $\%\Delta J^{2+}_{i,t}$ and $\%\Delta J^{2-}_{i,t}$ decompose $\%\Delta J^2_{i,t}$ using an indicator variable for the sign of the difference where $\%\Delta J^{2+}_{i,t} = \%\Delta J^2_{i,t} I_{[RS^+_{i,t} - RS^-_{i,t} > 0]}$. R^2 values in the final column are the average across the R^2s of the 105 individual series.

Forecasting measures of future volatility using log realized semivariances, results for the SPDR

$$\ln \overline{RM}_{h,t+h} = \mu + \phi_d^+ \ln RS_t^+ + \phi_w^+ \ln \overline{RS}_{w,t}^+ + \phi_m^+ \ln \overline{RS}_{m,t}^+ + \phi_d^- \ln RS_t^- + \phi_w^- \ln \overline{RS}_{w,t}^- + \phi_m^- \ln \overline{RS}_{m,t}^- + \epsilon_{t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=1$	RV	0.014	−0.070	−0.067	0.442	0.410	0.220	0.789
		(0.6)	(−1.8)	(−0.9)	(20.8)	(10.2)	(2.8)	
	RS^+	−0.015	−0.114	0.014	0.476	0.468	0.122	0.786
		(−0.6)	(−2.9)	(0.2)	(21.4)	(11.4)	(1.5)	
	RS^-	0.050	−0.003	−0.132	0.409	0.332	0.295	0.722
		(1.9)	(−0.1)	(−1.5)	(17.1)	(6.9)	(3.1)	
$h=5$	RV	0.011	0.010	−0.163	0.332	0.358	0.371	0.806
		(0.5)	(0.2)	(−1.3)	(17.4)	(5.7)	(2.7)	
	RS^+	−0.005	−0.008	−0.122	0.355	0.382	0.324	0.817
		(−0.3)	(−0.1)	(−1.0)	(18.5)	(6.3)	(2.5)	
	RS^-	0.029	0.030	−0.197	0.310	0.333	0.406	0.779
		(1.3)	(0.5)	(−1.4)	(15.5)	(4.8)	(2.7)	
$h=22$	RV	0.010	−0.054	−0.270	0.236	0.362	0.555	0.726
		(0.4)	(−0.7)	(−0.9)	(11.8)	(5.2)	(1.7)	
	RS^+	0.008	−0.055	−0.280	0.241	0.365	0.571	0.736
		(0.3)	(−0.7)	(−1.0)	(12.1)	(5.3)	(1.8)	
	RS^-	0.012	−0.051	−0.255	0.232	0.359	0.531	0.710
		(0.5)	(−0.7)	(−0.9)	(11.2)	(5.2)	(1.6)	
$h=66$	RV	−0.016	−0.116	−0.192	0.184	0.347	0.534	0.627
		(−0.5)	(−1.3)	(−0.6)	(8.0)	(3.9)	(1.6)	
	RS^+	−0.019	−0.124	−0.216	0.190	0.361	0.555	0.631
		(−0.6)	(−1.3)	(−0.7)	(8.1)	(4.0)	(1.7)	
	RS^-	−0.013	−0.109	−0.167	0.178	0.334	0.512	0.622
		(−0.4)	(−1.2)	(−0.5)	(7.9)	(3.9)	(1.6)	

Table A.5a: Extended model estimated in logs where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-).

Forecasting measures of future volatility using log realized semivariances, results for the panel of 105 individual stocks.

$$\ln \overline{RM}_{h,i,t+h} = \mu_i + \phi_d^+ \ln RS_{i,t}^+ + \phi_w^+ \ln \overline{RS}_{w,i,t}^+ + \phi_m^+ \ln \overline{RS}_{m,i,t}^+ + \phi_d^- \ln RS_{i,t}^- + \phi_w^- \ln \overline{RS}_{w,i,t}^- + \phi_m^- \ln \overline{RS}_{m,i,t}^- + \epsilon_{i,t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=1$	RV	0.137 (22.7)	0.051 (5.6)	0.038 (2.6)	0.291 (37.0)	0.271 (30.4)	0.163 (11.7)	0.726
	RS^+	0.146 (22.6)	0.053 (5.5)	0.067 (4.4)	0.285 (33.6)	0.270 (28.8)	0.128 (8.9)	0.689
	RS^-	0.130 (19.7)	0.051 (4.8)	0.009 (0.5)	0.296 (36.9)	0.270 (26.7)	0.196 (12.2)	0.681
$h=5$	RV	0.082 (15.4)	0.062 (4.1)	0.036 (1.2)	0.233 (26.2)	0.284 (19.7)	0.227 (8.3)	0.782
	RS^+	0.085 (16.1)	0.065 (4.5)	0.062 (2.1)	0.234 (25.6)	0.280 (19.9)	0.199 (7.7)	0.778
	RS^-	0.080 (14.1)	0.059 (3.7)	0.009 (0.3)	0.232 (26.5)	0.287 (18.9)	0.255 (8.5)	0.760
$h=22$	RV	0.050 (7.5)	0.033 (1.7)	0.051 (0.9)	0.174 (16.6)	0.265 (13.6)	0.286 (4.6)	0.739
	RS^+	0.052 (8.1)	0.039 (2.0)	0.061 (1.1)	0.173 (16.3)	0.260 (13.5)	0.276 (4.5)	0.744
	RS^-	0.048 (7.0)	0.028 (1.4)	0.041 (0.7)	0.175 (16.9)	0.270 (13.7)	0.295 (4.7)	0.725
$h=66$	RV	0.029 (4.0)	0.015 (0.6)	0.116 (1.6)	0.129 (12.1)	0.221 (8.3)	0.278 (4.5)	0.679
	RS^+	0.030 (4.3)	0.019 (0.8)	0.123 (1.7)	0.129 (11.8)	0.218 (8.2)	0.270 (4.3)	0.683
	RS^-	0.027 (3.8)	0.011 (0.5)	0.109 (1.5)	0.130 (12.3)	0.223 (8.5)	0.286 (4.6)	0.670

Table A.5b: Extended model estimated in logs where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-). The reported R^2 are the average across the R^2s of the 105 individual series.

E Model using Spot RV

The main results in the paper use the h-day cumulative RV, and show that negative semivariance is useful for horizons out to 66 days. However, some of the longer term predictability may be simply due to short term predictability, and so we also estimate the models on "spot" volatility, that is, the volatility on day-$t+h$, rather than the cumulative volatility. Aside from this simple change in the dependent variable, the models are unmodified and all estimation uses weighted least squares. We do not report results for $h = 1$ case since this is identical to the results reported in the paper.

Results for SPDR

Tables A.6a, A.7a and A.8a correspond to tables 2a and 3a in the main paper, and Table A.1a of this web appendix. The results in terms of the sign, relative magnitude and statistical significance are preserved although the t-stats are noticeably smaller. Table A.6a shows that negative semivariance is useful for all prediction horizons even when only predicting the variance on a single day. Table A.8a shows that positive semivariance typically produces a reduction in both positive and negative semivariance (and total realized variance), and that the coefficients on negative semivariance are large and statistically significant even at long horizons. Finally, Table A.7a shows that the effect of jumps can be detected even at the longest horizons.

Results for Panel

Tables A.6b, A.7b and A.8b correspond to tables 2b and 3b in the main paper, and Table A.1b in this web appendix. The results on the panel are broadly similar to those in the paper and show that negative semivariance plays a larger role than positive semivariance, especially at longer horizons.

HAR estimation results for the SPDR, spot volatility

$$RV_{h,t+h} = \mu + \phi_d RV_t + \phi_d^+ RS_t^+ + \phi_d^- RS_t^- + \gamma RV_t I_{[r_{t-1}<0]} + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h = 5$	0.270				0.526	0.177	0.327
	(6.5)				(6.0)	(3.3)	
		−0.014	0.559		0.528	0.173	0.341
		(−0.3)	(6.1)		(6.1)	(3.3)	
		0.072	0.370	0.084	0.532	0.175	0.340
		(1.0)	(2.7)	(1.9)	(6.1)	(3.3)	
$h = 22$	0.148				0.410	0.250	0.128
	(3.2)				(4.1)	(2.5)	
		−0.105	0.439		0.401	0.245	0.138
		(−1.5)	(4.0)		(4.0)	(2.5)	
		−0.109	0.450	−0.005	0.401	0.245	0.138
		(−1.1)	(2.5)	(−0.1)	(4.0)	(2.5)	
$h = 66$	0.038				0.159	0.257	0.021
	(1.9)				(2.7)	(3.2)	
		−0.105	0.220		0.150	0.253	0.022
		(−1.7)	(2.6)		(2.6)	(3.2)	
		−0.133	0.287	−0.031	0.149	0.252	0.022
		(−1.8)	(2.5)	(−0.8)	(2.6)	(3.1)	

Table A.6a: Reference, base and asymmetric model parameter estimates using day$-t+h$ realized variance as the dependent variance (t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used. The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

HAR estimation results for the panel of 105 individual stocks, spot volatility

$$RV_{h,i,t+h} = \mu_i + \phi_d RV_{i,t} + \phi_d^+ RS_{i,t}^+ + \phi_d^- RS_{i,t}^- + \gamma RV_{i,t} I_{[r_{i,t-1}<0]} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	ϕ_d	ϕ_d^+	ϕ_d^-	γ	ϕ_w	ϕ_m	R^2
$h=5$	0.273				0.376	0.299	0.258
	(14.3)				(11.4)	(8.9)	
		0.119	0.424		0.377	0.299	0.260
		(5.2)	(13.1)		(11.4)	(8.9)	
		0.162	0.334	0.044	0.378	0.299	0.260
		(7.1)	(13.1)	(4.6)	(11.4)	(9.0)	
$h=22$	0.161				0.242	0.453	0.165
	(7.2)				(5.6)	(8.2)	
		0.019	0.300		0.243	0.453	0.165
		(0.8)	(7.6)		(5.7)	(8.2)	
		0.048	0.237	0.032	0.244	0.454	0.166
		(2.2)	(6.3)	(3.7)	(5.7)	(8.2)	
$h=66$	0.109				0.173	0.424	0.084
	(7.7)				(5.0)	(6.7)	
		0.006	0.210		0.174	0.423	0.084
		(0.2)	(5.5)		(5.0)	(6.7)	
		0.022	0.173	0.019	0.174	0.424	0.084
		(1.1)	(6.3)	(1.5)	(5.0)	(6.7)	

Table A.6b: Reference, base and asymmetric model parameter estimates using day$-t+h$ realized variance as the dependent variance (robust t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated in the left most column. Each panel contains 3 models: the first model corresponds to the reference model using only realized variance, the second decomposes realized variance into positive and negative realized semivariance at the first lag, and the third specification adds an asymmetric term where the sign of the most recent daily return is used. The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

The impact of signed jump variation on future spot volatility, results for the SPDR

$$RM_{h,t+h} = \mu + \phi_J \Delta J_t^2 + \phi_{J^+} \Delta J_t^{2+} + \phi_{J^-} \Delta J_t^{2-} + \phi_C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \epsilon_{t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=5$	RV				0.307 (7.2)	0.504 (5.8)	0.175 (3.3)	0.334
	RV	−0.248 (−4.0)			0.302 (7.2)	0.512 (5.9)	0.173 (3.3)	0.343
	RV		−0.252 (−2.7)	−0.243 (−1.7)	0.302 (6.6)	0.511 (5.9)	0.173 (3.2)	0.343
	BV	−0.262 (−4.5)			0.303 (8.0)	0.479 (6.3)	0.164 (3.2)	0.370
$h=22$	RV				0.194 (3.7)	0.378 (3.9)	0.246 (2.5)	0.130
	RV	−0.236 (−3.2)			0.200 (3.8)	0.379 (3.8)	0.244 (2.5)	0.138
	RV		−0.348 (−3.3)	−0.057 (−0.3)	0.230 (3.8)	0.371 (3.6)	0.244 (2.5)	0.138
	BV	−0.239 (−3.4)			0.186 (3.7)	0.367 (3.7)	0.235 (2.4)	0.148
$h=66$	RV				0.042 (1.8)	0.156 (2.6)	0.257 (3.2)	0.021
	RV	−0.144 (−2.1)			0.053 (2.0)	0.153 (2.6)	0.255 (3.2)	0.022
	RV		−0.029 (−0.4)	−0.367 (−2.4)	0.018 (0.6)	0.163 (2.6)	0.255 (3.2)	0.023
	BV	−0.150 (−2.5)			0.061 (2.4)	0.140 (2.5)	0.242 (3.2)	0.023

Table A.7a: Models that include signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). $\Delta J_{i,t}^{2+}$ and $\Delta J_{i,t}^{2-}$ decompose $\Delta J_{i,t}^2$ using an indicator variable for the sign of the difference where $\Delta J_{i,t}^{2+} = \Delta J_{i,t}^2 I_{[RS^+ - RS^- > 0]}$. The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

The impact of signed jump variation on future spot volatility, results for the panel of 105 individual stocks

$$RM_{h,i,t+h} = \mu_i + \phi_J \Delta J_{i,t}^2 + \phi_{J^+} \Delta J_{i,t}^{2+} + \phi_{J^-} \Delta J_{i,t}^{2-} + \phi_C BV_{i,t} + \phi_w \overline{RV}_{w,i,t} + \phi_m \overline{RV}_{m,i,t} + \epsilon_{i,t+h}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_w	ϕ_m	R^2
$h=5$	RV				0.320 (13.7)	0.380 (11.5)	0.305 (9.0)	0.258
	RV	−0.151 (−7.3)			0.317 (13.8)	0.382 (11.6)	0.305 (9.1)	0.259
	RV		−0.138 (−4.2)	−0.165 (−6.1)	0.314 (12.4)	0.382 (11.6)	0.305 (9.1)	0.259
	BV	−0.129 (−7.5)			0.278 (14.5)	0.308 (11.1)	0.236 (8.6)	0.281
$h=22$	RV				0.185 (6.6)	0.246 (5.8)	0.458 (8.2)	0.164
	RV	−0.139 (−5.8)			0.183 (6.7)	0.248 (5.8)	0.458 (8.3)	0.165
	RV		−0.111 (−3.6)	−0.168 (−4.4)	0.175 (6.1)	0.249 (5.8)	0.458 (8.3)	0.165
	BV	−0.109 (−5.5)			0.160 (6.7)	0.201 (5.5)	0.363 (7.8)	0.169
$h=66$	RV				0.115 (5.9)	0.180 (5.2)	0.429 (6.7)	0.083
	RV	−0.101 (−3.5)			0.114 (6.0)	0.183 (5.2)	0.429 (6.8)	0.084
	RV		−0.049 (−1.6)	−0.156 (−3.4)	0.099 (4.6)	0.184 (5.2)	0.429 (6.8)	0.084
	BV	−0.081 (−3.2)			0.098 (5.8)	0.144 (4.9)	0.345 (6.6)	0.087

Table A.7b: Models that include signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (robust t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV) or bipower variation (BV). $\Delta J_{i,t}^{2+}$ and $\Delta J_{i,t}^{2-}$ decompose $\Delta J_{i,t}^2$ using an indicator variable for the sign of the difference where $\Delta J_{i,t}^{2+} = \Delta J_{i,t}^2 I_{[RS^+ - RS^- > 0]}$. The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

Forecasting measures of future volatility using spot realized semivariances, results for the SPDR

$$RM_{h,t+h} = \mu + \phi_d^+ RS_t^+ + \phi_w^+ \overline{RS}_{w,t}^+ + \phi_m^+ \overline{RS}_{m,t}^+ + \phi_d^- RS_t^- + \phi_w^- \overline{RS}_{w,t}^- + \phi_m^- \overline{RS}_{m,t}^- + \epsilon_{t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=5$	RV	−0.048 (−0.9)	0.068 (0.3)	−0.694 (−2.4)	0.526 (5.8)	0.948 (3.6)	1.171 (3.3)	0.363
	RS^+	−0.039 (−1.2)	0.032 (0.3)	−0.322 (−2.1)	0.263 (4.9)	0.519 (3.3)	0.546 (3.1)	0.302
	RS^-	−0.009 (−0.3)	0.035 (0.4)	−0.372 (−2.5)	0.263 (6.2)	0.429 (3.6)	0.624 (3.4)	0.377
$h=22$	RV	−0.145 (−1.7)	−0.264 (−0.9)	−0.721 (−1.1)	0.391 (3.5)	1.053 (3.4)	1.366 (1.8)	0.165
	RS^+	−0.071 (−1.5)	−0.155 (−1.0)	−0.437 (−1.3)	0.203 (3.2)	0.544 (3.3)	0.781 (1.9)	0.137
	RS^-	−0.074 (−2.0)	−0.109 (−0.8)	−0.284 (−0.9)	0.188 (3.7)	0.509 (3.2)	0.585 (1.6)	0.169
$h=66$	RV	−0.127 (−1.9)	−0.228 (−1.4)	−0.596 (−0.9)	0.183 (2.4)	0.536 (2.8)	1.216 (1.7)	0.026
	RS^+	−0.061 (−1.5)	−0.115 (−1.4)	−0.221 (−0.6)	0.074 (1.5)	0.286 (2.7)	0.523 (1.3)	0.018
	RS^-	−0.066 (−2.2)	−0.114 (−1.5)	−0.376 (−1.2)	0.109 (3.3)	0.251 (2.7)	0.693 (2.1)	0.033

Table A.8a: Extended model where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-). The R^2 measure is constructed using the WLS parameter estimates and the original, unmodified data.

Forecasting measures of future volatility using spot realized semivariances, results for the panel of 105 individual stocks.

$$RM_{h,i,t+h} = \mu_i + \phi_d^+ RS_{i,t}^+ + \phi_w^+ \overline{RS}_{w,i,t}^+ + \phi_m^+ \overline{RS}_{m,i,t}^+ + \phi_d^- RS_{i,t}^- + \phi_w^- \overline{RS}_{w,i,t}^- + \phi_m^- \overline{RS}_{m,i,t}^- + \epsilon_{i,t+h}$$

	RM	ϕ_d^+	ϕ_w^+	ϕ_m^+	ϕ_d^-	ϕ_w^-	ϕ_m^-	R^2
$h=5$	RV	0.113 (5.0)	0.155 (2.7)	−0.000 (−0.0)	0.415 (13.0)	0.591 (11.9)	0.621 (6.9)	0.261
	RS^+	0.059 (4.4)	0.093 (2.9)	0.012 (0.3)	0.207 (12.4)	0.287 (11.3)	0.293 (6.8)	0.241
	RS^-	0.055 (5.2)	0.063 (2.4)	−0.012 (−0.3)	0.208 (12.5)	0.304 (12.1)	0.328 (6.8)	0.209
$h=22$	RV	0.013 (0.5)	0.013 (0.2)	0.101 (0.5)	0.291 (7.7)	0.463 (7.0)	0.834 (4.2)	0.167
	RS^+	0.008 (0.6)	0.002 (0.1)	0.051 (0.5)	0.150 (7.5)	0.233 (7.3)	0.419 (4.0)	0.150
	RS^-	0.005 (0.5)	0.010 (0.3)	0.050 (0.5)	0.141 (7.8)	0.230 (6.6)	0.414 (4.2)	0.144
$h=66$	RV	0.001 (0.0)	−0.007 (−0.1)	0.257 (1.7)	0.204 (5.7)	0.350 (5.3)	0.608 (3.6)	0.085
	RS^+	0.001 (0.1)	−0.001 (−0.0)	0.140 (1.8)	0.100 (4.9)	0.175 (5.5)	0.293 (3.3)	0.075
	RS^-	−0.000 (−0.0)	−0.007 (−0.2)	0.117 (1.6)	0.103 (6.5)	0.176 (5.1)	0.315 (4.0)	0.073

Table A.8b: Extended model where RV at all lags is decomposed into positive and negative semivariance (robust t-statistics in parentheses). Each of the three panels contains results for the forecast horizon indicated at the left. RM indicates the dependent variable, realized variance (RV), positive realized semivariance (RS^+) or negative realized semivariance (RS^-). The final column reports the average of the 105 R^2s for the individual assets constructed using the WLS parameter estimates and the original, unmodified data.

F Autocorrelations of RV, RS^+ and RS^-

Figure A.1 contains a plot of the first 66 autocorrelations of realized variance and positive and negative realized semivariance for the S&P 500 ETF. While the long-run behavior of the three autocorrelation series is similar, negative semivariance has uniformly larger autocorrelations than positive semivariance. The difference at the first lag is 0.19.

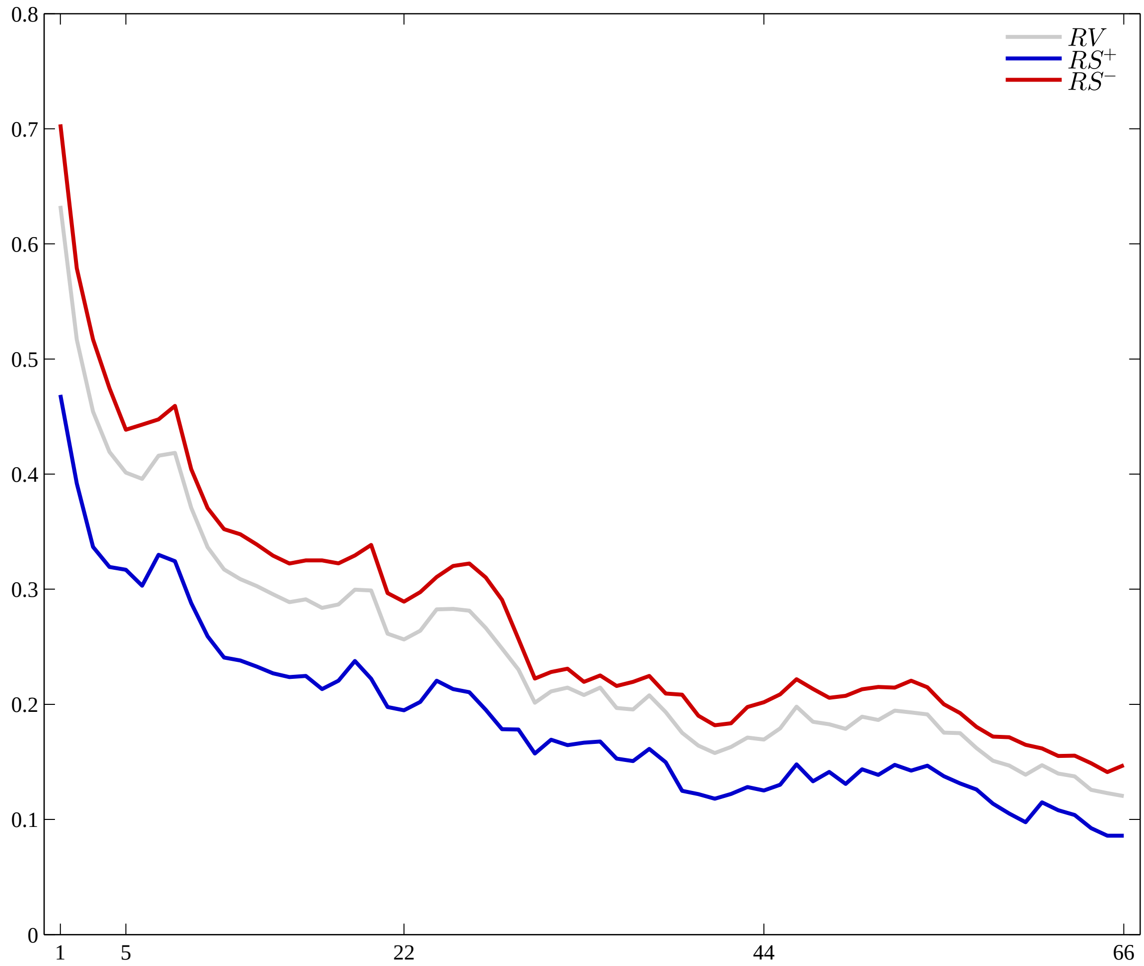

Figure A.1: The first 66 autocorrelations of realized variance and positive and negative realized semi variance for the S&P 500 ETF (SPDR).