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

October 7, 2011

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. A simple model exploiting these findings leads to significantly better out-of-sample forecast performance, across 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

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 accommodate the presence of jumps 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.[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, Bollerslev, 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 to 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] 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.

---

[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 (2006b), Lanne (2006a), 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, have a long history in finance. Application 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).

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), Andersen, Bollerslev, and Diebold (2007) and Corsi, Pirino, and Renò (2010), 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. 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 around 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 individual 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",

---

[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.

3

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, nonparametric 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 as well as the S&P 500 index – and verify that the usefulness of realized semivariances relative to realized variances is not restricted only to broad stock indices. We introduce a fixed-effect panel HAR to aid inference on the importance of realized semivariance in the individual stocks. 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, and we relate realized variance to the two realized semivariance components in a novel Vector Heterogeneous Autoregression (VHAR). Third, we show the usefulness of simple HAR-type 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 earlier findings in the extant literature. In particular, we show that the impact of jumps on future volatility depends critically upon the sign of the jump; averaging across both positive and negative jumps the impact is around zero, while separating the jumps by sign leads to significant predictive gains.

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 for the 106 assets 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 individual stocks, and out-of-sample forecasting application, and a description of the time series properties of the new measure of jump variability. Section 7

4

concludes.

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:

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

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).

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

5

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

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

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

---

[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.

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 squared jumps with a negative/positive sign:

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

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

by Corollary 1 of Barndorff-Nielsen, Kinnebrock, and Shephard (2010). An interesting consequence of the limit of realized semivariances is that the variation due to the continuous component can be removed by simply subtracting one RS from the other, and the remaining part is what we define as the *signed jump variation:*

$$\Delta J^2 \quad \equiv \quad RS^+ - RS^- \tag{8}$$

$$\xrightarrow{p} \quad \sum_{0 \leq s \leq t} \Delta p_s^2 I_{[\Delta p_s > 0]} - \sum_{0 \leq s \leq t} \Delta p_s^2 I_{[\Delta p_s < 0]}.$$

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 consist of high-frequency 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

---

[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.

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 that were continuously available for at least four years.

All prices used were transactions taken from NYSE TAQ. Trades were filtered to include only those occurring between 9:30:00 and 16:00:00 (inclusive) and were 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 sample prices every, say, 5 minutes, we sample prices so that there are an even number of transactions between each observation. (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 (2010) for example.

We elect to 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 (2005). This procedure should produce a mild decrease in variance, and has been shown to be a reasonable choice for modeling the time-series properties of volatility in Andersen, Bollerslev, and Meddahi (2011).

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 jth 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).[8] Using a skip estimator was advocated in Huang and Tauchen (2005) as an important

---

[8]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.

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.[9]

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 Weizsacker (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,

$$y_{t+h} = \mu + \phi_1 y_t + \phi_5 \left(\frac{1}{4} \sum_{i=2}^{5} y_{t-i} \right) + \phi_{22} \left(\frac{1}{17} \sum_{i=6}^{22} y_{t-i} \right) + \epsilon_t \tag{12}$$

where y denotes the volatility measure (RV, BV, etc). This reparameterization allows for direct interpretation of the effect of y_t on y_{t+h} through ϕ_1. Throughout we will use $\bar{y}_5$ to indicate the average value over lags 2 to 5, and $\bar{y}_{22}$ 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, and we set the dependent variable to either the h-day average cumulative volatility measure, $\bar{y}_{h,t+h} = \frac{1}{h} \sum_{i=1}^{h} y_{t+i}$, or the h-day ahead volatility measure, y_{t+h}.

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

---

[9]We also conducted our empirical analysis using the MedRV estimator of Andersen, Dobrev, and Schaumburg (2010), which is an alternative jump-robust estimator of integrated variance. The resulting estimates and conclusions were almost identical to using BV and so we omit them in the interest of brevity.

then construct weights as the inverse of the fitted value from that model.[10]

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.

3.4 A Panel HAR for Volatility Modeling

Separate estimation of the models on the individual firms' realized variance is feasible (results for individual stocks are presented in Section 6.1), but does not provide a direct method to assess the significance of the average effect, and so we fit 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

$$\overline{y}_{h,i,t+h} = \mu_i + \phi_1 y_{i,t} + \phi_5 \overline{y}_{5,i,t} + \phi_{22} \overline{y}_{22,i,t} + \epsilon_{i,t}, \ \ i = 1,\ldots,n_t, \ \ t = 1,\ldots,T,$$

where μ_i is a fixed effect which allows each firm to have different levels of long-run volatility. Let $Y_{i,t} = [y_{i,t}, \bar{y}_{5,i,t}, \bar{y}_{22,i,t}]'$, then the model for each firm's realized variance can be compactly expressed as:

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

Define $\tilde{y}_{h,i,t+h} = \overline{y}_{h,i,t+h} - \bar{\bar{y}}_{h,i}$ and $\tilde{Y}_{i,t} = Y_{i,t} - \bar{Y}_i$ where $\bar{\bar{y}}_{h,i}$ and $\bar{Y}_i$ are the WLS estimates of the mean of $\overline{y}$ and Y, respectively. The pooled parameters are then estimated by

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

where $w_{i,t}$ are the weights and n_t are the number of firms in the cross section at date t.[11]

Inference can be conducted using the asymptotic distribution

---

[10]All models were estimated using an alternative weighting scheme constructed using an EWMA with a smoothing parameter of 0.94 in place of the HAR and were found to be robust to the choice of weighting.

[11]Our analysis takes the cross-section size, n_t, as finite. If a form of 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).

$$\sqrt{T}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right) \text{ as } T \to \infty \tag{14}$$

$$where \ \boldsymbol{\Sigma} = \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left(n_t^{-1} \sum_{i=1}^{n_t} w_{i,t} \tilde{Y}_{i,t} \tilde{Y}_{i,t}' \right)$$

$$\boldsymbol{\Omega} = \text{avar}\left(T^{-1/2} \sum_{t=1}^{T} \mathbf{z}_t \right)$$

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

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 which decompose realized volatility into signed components, it is useful to establish a set of reference results. We fit a reference specification

$$\overline{RV}_{t+h} = \mu + \phi_1 RV_t + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t \tag{15}$$

to both the S&P 500 ETF and the panel where $\overline{RV}_{5,t}$ is the average between lags 2 and 5 and $\overline{RV}_{22,t}$ 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 this model is identical except for the inclusion fixed effects to permit different long-run variances for each asset. Tables 1a and 1b 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-statistic for this specification. These results are in line with that has been previously documented in the literature: substantial persistence with $\phi_1 + \phi_5 + \phi_{22}$ close to 1 and where the role of recent information, captured by ϕ_1, diminishes as the horizon increases. 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 WLS parameter estimates and the original data.

Tables 2a and 2b contain the parameters estimated using a model which fits the day-h realized variance instead of the cumulative realized variance. The specification only differs in the left-hand side variables, and is given by

$$RV_{t+h} = \mu + \phi_1 RV_t + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t. \tag{16}$$

(The $h = 1$ results are identical to those in Tables 1a and 1b and so are omitted.) The results in Table 2 reveal, as expected, that much, but not all, of the predictive power in the model for cumulative realized variance occurs at short horizons.

[INSERT TABLES 1a, 1b , 2a AND 2b ABOUT HERE]

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. We initially only decompose the most recent volatility (RV_t), and in the next section we decompose all three volatility terms. Applying this decomposition produces the specification

$$\overline{RV}_{t+h} = \mu + \phi_1^+ RS_t^+ + \phi_1^- RS_t^- + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t. \tag{17}$$

The panel specification of the above model includes fixed effects but is otherwise identical. To make ϕ^+ and ϕ^- directly comparable to ϕ we multiply RS^+ and RS^- by a factor of two. If the decomposition of RV into RS^+ and RS^- added no information we would expect to find $\phi_1^+ = \phi_1^- = \phi_1$.

Our first new empirical results using realized semivariances are presented in the second row of each panel of Tables 1a and 1a. In the models for the SPDR (left panels), 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 $\phi_1^+ + \phi_1^-$, 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 good and bad news. The results for the panel of individual volatility series also find 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 persistence of idiosyncratic jumps in the individual firms' volatility, which we explore in the next section.

Figure 1 contains the point estimates of ϕ_1^+ and ϕ_1^- from eq. 17 for all horizons between 1 and 66 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, but 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.

As noted above, if the decomposition of RV into $RS+$ and $RS-$ added no new information, then we would expect to see $\phi_1^+ = \phi_1^- = \phi_1$. 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.[12] We interpret these findings as strong evidence that decomposing RV into its signed components significantly improves the explanatory power of this model.

[INSERT FIGURE 1 ABOUT HERE]

Tables 2a and 2b contain results for the same model using the day-h realized variance as the left-hand-side variable. These results show the effect on the variance on a particular day ahead and so the magnitude of the coefficients is predictably smaller. The coefficient on negative semivariance remains positive and strongly statistically significant at all horizons for both the SPDR and the panel. The coefficient on positive semivariance for the SPDR is not significant at short horizons, and significantly negative for horizons $h = 22$ and 66, while for the panel it is significantly positive at short horizons and not significantly different from

---

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

zero at the longer horizons. The models with semivariance also appear to improve predictability, in terms of R^2, at all horizons, and are particularly important in the model for the market variance.

4.2 Comparison with a Simple Leverage Effect Variable

The classic leverage effect, whether due to varying firm leverage as in 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. In this section we determine whether our approach using information from realized semivariances adds anything 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, $RV_t I_{[r_t<0]}$.[13]

$$\overline{RV}_{t+h} = \mu + \phi_1^+ RS_t^+ + \phi_1^- RS_t^- + \gamma RV_t I_{[r_t<0]} + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t \tag{18}$$

Like realized semivariances, we multiply the interaction variance by a factor of two so that its coefficient is of a similar magnitude to that on RV. If realized semivariance adds no new information beyond the interaction variable then we expect $\phi_1^+ = \phi_1^-$ and γ to be significant.

The final row in each panel of Tables 1a and 1b contain the parameter estimates from this model. In all cases the magnitude of the coefficient on the interaction term is small. In models based on the SPDR, the interaction term has the opposite sign at $h = 22$ and 66 from what is commonly found, 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. Similar results are reported in Tables 2a and 2b where the day-h realized variance as the left-hand-side variable.

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

---

[13]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, $r_t^2 I_{[r_t<0]}$ were even weaker than those discussed here.

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.

4.3   Completely Decomposing Quadratic Variation

The specification in eq. (17) 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. With this modification we obtain:

$$\overline{RV}_{t+h} = \mu + \phi_1^+ RS_t^+ + \phi_1^- RS_t^- + \phi_5^+ \overline{RS}_t^{5+} + \phi_5^- \overline{RS}_t^{5-} + \phi_{22}^+ \overline{RS}_t^{22+} + \phi_{22}^- \overline{RS}_t^{22-} + \epsilon_t \qquad (19)$$

where $\overline{RS}^{j-}$ is the j-day average of negative semivariance and $\overline{RS}^{j+}$ is the j-day average of positive semi-variance.[14]

Results from this extended specification are presented in the first row of each column of Table 3a. 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 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.

[INSERT TABLES 3a AND 3b HERE]

Decomposing realized variance at all lags allows us to consider a "Vector HAR" (VHAR) for the two semi-variances. 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.

---

[14]We continue to use the rotated parametrization where the lag-5 term includes the average between lags 2 and 5, and the 22-lag term contains the average between lags 6 and 22. Chen and Ghysels (2011), when a jump variable is also included, call this model the "HAR-S-RV-J" model, and use it as one of the benchmarks in their study.

$$
\begin{bmatrix} \overline{RS}_{t+h}^{+} \\ \overline{RS}_{t+h}^{-} \end{bmatrix} = \begin{bmatrix} \mu^{+} \\ \mu^{-} \end{bmatrix} + \boldsymbol{\phi}_1 \begin{bmatrix} RS_t^{+} \\ RS_t^{-} \end{bmatrix} + \boldsymbol{\phi}_5 \begin{bmatrix} \overline{RS}_t^{5+} \\ \overline{RS}_t^{5-} \end{bmatrix} + \boldsymbol{\phi}_{22} \begin{bmatrix} RS_t^{22+} \\ RS_t^{22-} \end{bmatrix} + \begin{bmatrix} \epsilon_t^{+} \\ \epsilon_t^{-} \end{bmatrix} \tag{20}
$$

where

$$
\boldsymbol{\phi}_j = \begin{bmatrix} \phi_{j+}^{+} & \phi_{j+}^{-} \\ \phi_{j-}^{+} & \phi_{j-}^{-} \end{bmatrix}.
$$

Results of the VHAR are presented in the two lower rows of each panel of Tables 3a and 3b. The estimates for both semivariances are virtually identical, with small or negative coefficients on lagged 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 left-most sub-panel of Table 4 reports the rejection frequency for two nulls on the complete VHAR and on each semivariance.[15] The first null is that positive semivariance can be excluded, $H_0 : \phi_1^{+} = \phi_5^{+} = \phi_{22}^{+} = 0$, and the other tests whether negative semivariance can be excluded. We find that positive semivariance can be excluded from 19.1.% of the joint models, 19.1% of the models for positive semivariance, and 25.7% of the models which just contained 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.

[INSERT TABLE 4 HERE]

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, $H_0 : \phi_1^{+} + \phi_5^{+} + \phi_{22}^{+} = \phi_1^{-} + \phi_5^{-} + \phi_{22}^{-}$ in each of the two semivariance models. The second panel of Table 4 reveals that the null can be rejected in almost all cases,

---

[15]All hypothesis tests were implemented using a Wald test where the parameter covariance matrix was estimated using a HAC as described in section 3.3.

and when rejected, indicated that the sum of the coefficients on the negative semivariance was larger than the sum of the coefficient on the positive semivariance. We next test for equality only at lag 1, $H_0 : \phi_1^+ = \phi_1^-$. As shown in Table 4, 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 negative semivariances get greater weight in these predictive models in almost all cases.

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. 20 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 computed. 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. The final column of Table 4 shows that equality 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.

5 Signed Jump Information

All of the models estimated thus far examined the role that decomposing realized variances into positive and negative realized semivariance can play in explaining future volatility. These results consistently demonstrated that the information content of negative realized semivariance was 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 half of the integrated variance.

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 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 added advantage that a jump-robust estimator of integrated variance, such as MedRV or BV, is not needed; we obtain the measure simply as the difference between RS_t^+ and RS_t^-. If jumps are rare, as is often found in the stochastic volatility literature, then this measure should broadly correspond to the jump variation when there a jump occurs and be mean zero noise otherwise.

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

$$\overline{RV}_{t+h} = \mu + \phi_J \Delta J_t^2 + \phi_C BV_t + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t. \tag{21}$$

The panel specification includes fixed effects but is otherwise identical.[16]

Results from the model with signed jumps are presented in the top row of each of the four panels in Tables 5a and 5b. 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.[17] By including information about the sign of the jump, we find that the jump variable does indeed help predict future volatility.

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 second row of Tables 5a and 5b, 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

---

[16]It is worth noting that while this specification is similar to our baseline model (eq. 17) 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 semivariance alone.

[17]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.

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 5a AND 5b ABOUT HERE]

The final specification we consider allows for the coefficient on positive jump variation to differ from that of negative jump variation, which enables us to test whether the impact of jumps is driven more by positive or negative jump variation. To do this, we need a method to estimate $\sum \Delta p_s^2 I_{[\Delta p_s > 0]}$ 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:

$$\overline{RV}_{t+h} = \mu + \phi^{J+} \Delta J_t^{2+} + \phi^{J-} \Delta J_t^{2-} + \phi^C BV_t + \phi_5 \overline{RV}_{5,t} + \phi_{22} \overline{RV}_{22,t} + \epsilon_t \tag{22}$$

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

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

If the signed jump components have equal predictive power then we expect to find $\phi^{J+} = \phi^{J-} = \phi^J$.

The bottom row of each panel in Table 5a contains estimates for the 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 the coefficient on the negative jump component is larger, in magnitude, than on the positive jump component, indicating that negative jumps increase future volatility by more than the amount by which positive jumps lower future volatility, and we test the null $H_0 : \phi^{J+} = \phi^{J-}$ we only reject at the one-step-ahead horizon ($h = 1$). 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 the "good" jumps lead to lower volatility while "bad" jumps lead to higher volatility, consistent with the restriction that $\phi^{J+} = \phi^{J-}$. 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.

[INSERT FIGURE 2 ABOUT HERE]

6   Further Analysis

This section presents additional analysis of our results. First, we present results from estimating our main empirical specifications on individual stock returns, rather than on the index or via panel regressions. Second, we look at an out-of-sample forecasting application to see whether the in-sample gains documented above lead to better forecasts out-of-sample. Finally, we take a detailed look at the new time series studied in this paper, the *signed jump variation* series.

6.1   Results for Individual Stocks

In this section we examine whether the significant findings from the previous sections that realized semi-variances contain useful information for predicting future volatility also hold at the individual stock level. We do so by estimating our three main models, the baseline specifications (eq. (17)) and the jump specifications (eqs. (21) and (22)) for each of the 105 individual firm variances.

To facilitate presentation of these results, we report only the proportion of significant coefficients and their sign, using 2-sided tests with sizes of 5%. In the baseline specification, the left panel of Table 6 shows that the negative realized semivariance is significant and positive for all but one of the series. Positive realized semivariance is only broadly significantly positive for the shortest horizons, and the percentage of series which reject the null declines from 84% for $h = 1$ to 26% for $h = 66$ days. These results are in line with the results from the panel specification.

[INSERT TABLE 6 HERE]

The jump and the extended jump specifications, in the middle and right panels of Table 6, also confirm that the findings in the pooled panel model are pervasive in the individual variance series. For example, no series have a significant and positive coefficient on the signed jump variable, ΔJ_t^2, while between 62% and 85% have significant and negative coefficients on this variable.

Figure 3 contains a plot of the estimated coefficients on the 105 individual firms from the jump specifi-

21

cation. The y-value indicated the magnitude of the coefficient where solid bars are statistically significant. The most striking feature of this plot is the pervasiveness of the negative sign on signed jump variation – even in cases where it is insignificant – coupled with the presence of no significantly positive coefficients in either plot.

[INSERT FIGURE 3 ABOUT HERE]

6.2 Out-of-Sample Forecast Performance

We now present a forecasting application to assess the out-of-sample predictive power of realized semi-variance. Two models using realized variance are specified. The first, denoted $\widehat{RV}_1$, is the standard RV-HAR containing lags 1, 5 and 22 (eq. (15)). The second, denoted $\widehat{RV}'_1$, augments the standard HAR with an interaction term which allows for asymmetry in persistence when the previous daily return was negative, $RV_t \mathrm{I}_{[r_t < 0]}$. The performance of these models is compared to that of two models based on realized semivariances. The first, denoted $\widehat{RV}_2$, is the specification which decomposes recent realized variance into positive and negative semivariance (e.g. 17). The second, denoted $\widehat{RV}'_2$, is a restricted version of the first where positive realized semivariance is excluded from the model, motivated by the limited significance of this variable in Table 1a. All forecasts are generated using rolling WLS regressions based on 1,004 observations (4 years), and parameter estimates are updated daily. Only series which contained at least 500 out-of-sample data points were included, reducing the number of assets from 105 to 83 (in addition to the market index, SPDR). No restrictions on the parameters were imposed and forecasts were occasionally negative (approximately .004%), and so an "insanity filter" was used which ensured that the forecast was no smaller than the smallest value observed in the estimation window.

Forecasting performance was evaluated using unconditional Diebold-Mariano-Giacomini-White tests (Diebold and Mariano (1995) and Giacomini and White (2006)) using the negative of the Gaussian quasi-likelihood as the loss function,

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

22

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 7 contains results from the forecasting exercise. Each of the three panels contains results from comparing one set of forecasting models. Within each panel, the left-most column contains the value of the DM test statistic and the two right columns contain the percentage of the 83 series which favor the realized semivariance based model or realized variance model using a 2-sided 5% test. The left-most panel compares the standard HAR with a semivariance-based model which decomposes the first lag. The DM test statistic is uniformly positive indicating superior 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 20 to 30% of series. The middle panel compares the standard HAR to a model which includes only negative semivariance at the first lag. This is our preferred 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 right-most 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.

[INSERT TABLE 7 ABOUT HERE]

6.3 Signed Jump Variation

The time series of percentage signed jump variation, $\%\Delta J_t^2 = (RS_t^+ - RS_t^-)/RV_t$, is also of direct interest. If jumps are rare (one or none a day) then this series contains either jumps or noise. Figure 4 contains the time series plot of the SPDR signed jump variation series. This series appears consistent with a heteroskedastic white noise process. The first 5 autocorrelations are -0.111, -0.022, -0.037, 0.015 and 0.0212 and each is individually insignificantly different from zero using heteroskedasticity robust inference. Signed jump variation

23

comprises 15% of total variance, although the absolute percentage due to signed jump variation has some persistence.

While it is tempting to interpret the heteroskedasticity as evidence of increased jump activity, it cannot be directly interpreted since the asymptotic variance of signed jump variation is proportional to the integrated quarticity which will be high in periods of high volatility. Figure 5 also highlights six days, the three largest and three smallest on the previous graph. The first two positive and first two negative both have large, obvious jumps. The third panel of each have period of large moves although the obviousness off the jumps is less apparent. The two largest jumps, the positive one on September 18, 2007 and the negative jump on December 11, 2007, both correspond to Federal Reserve rate cuts. On September 18, 2007 the Fed reduced the federal funds rate 50 basis points which triggered a rally. The December 11, 2007 drop also corresponded to an cut by the Fed – 25 basis points – which was interpreted as an ominous sign of an impending recession.

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 both an equity market index and 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, Boller-

24

slev, 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.

A Data Cleaning

Only transaction data were taken from the NYSE TAQ. All series were automatically cleaned according to a set of 6 rules:

1. Transactions outside of 9:30:00 AM and 16:00:00 were discarded

   (a) Transactions with a 0 price or volume were discarded

   (b) Each day the most active exchange was determined. Only transactions from this exchange were retained.

   (c) Only trades with conditions E, F or blank were retained.

   (d) Transaction prices outside of the CRSP high or low were discarded.

   (e) 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.

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. BOLLERSLEV, 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., T. BOLLERSLEV, AND N. MEDDAHI (2011): "Realized volatility forecasting and market microstructure noise," *Journal of Econometrics*, 160(1), 220 – 234.

ANDERSEN, T. G., D. DOBREV, AND E. SCHAUMBURG (2010): "Jump-robust volatility estimation using nearest neighbor truncation," Staff Reports 465, Federal Reserve Bank of New York.

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

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.

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 (2010): "Tails, fears and risk premia," *Journal of Finance, Forthcoming, Economic Research Initiatives at Duke (ERID) Working Paper Paper No. 34.*

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.

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, Forthcoming*.

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. (2006a): "Forecasting Realized Volatility by Decomposition," Discussion paper, European Univeristy Institute, EUI Working Paper ECO No. 2006/20.

——— (2006b): "A Mixture Multiplicative Error Model for Realized Volatility," Discussion paper, European Univeristy Institute, ECO No. 2006/3.

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

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

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. (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., AND K. SHEPPARD (2009): "Evaluating Volatility Forecasts," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch, p. 750. Springer.

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

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.
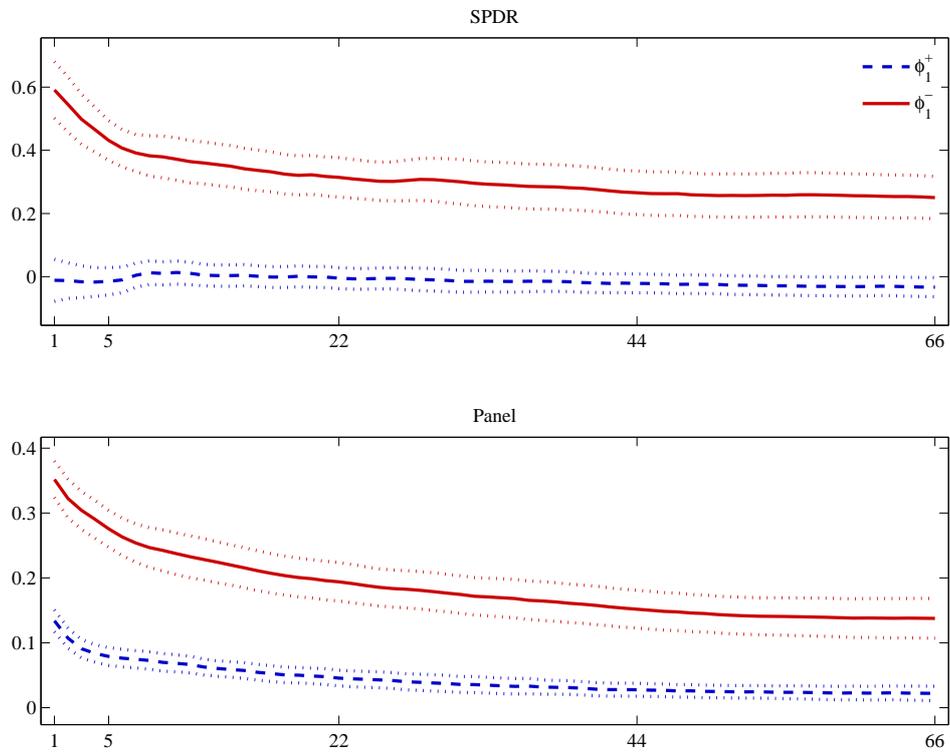
Figure 1: Estimated coefficients from the levels model that decomposed realized variance into its signed components, $\overline{RV}_{h,i,t+h} = \mu_i + \phi_1^+ RS_{i,t}^+ + \phi_1^- RS_{i,t}^- + \phi_5 \overline{RV}_{5,i,t} + \phi_{22} \overline{RV}_{22,i,t} + \epsilon_t$. 95% confidence intervals indicated using dashed lines. 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.
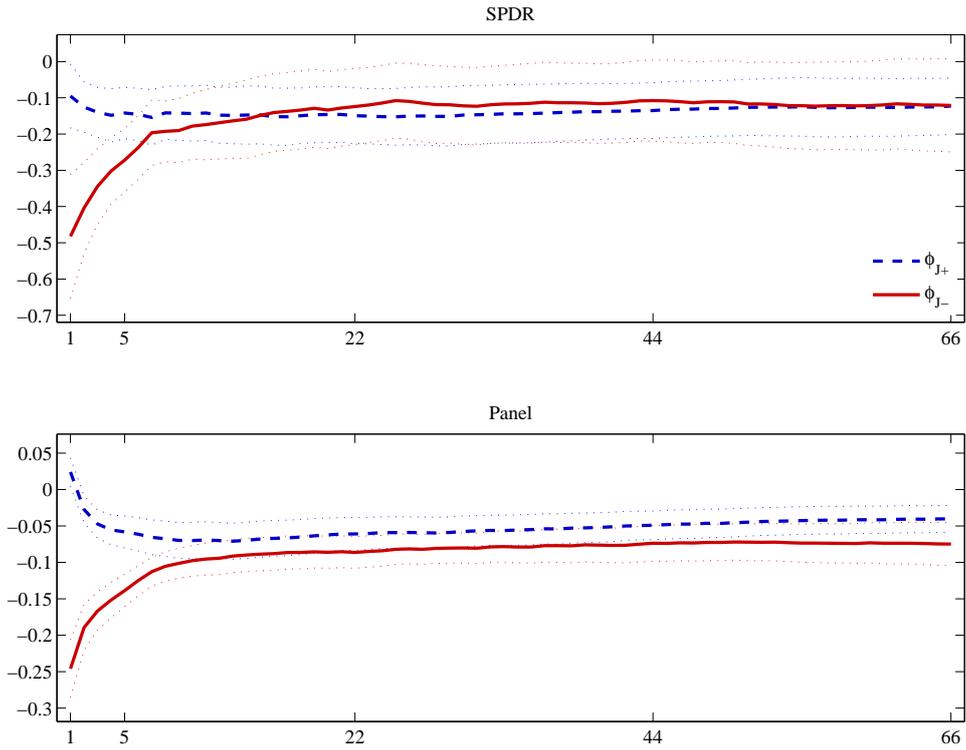
Figure 2: Fit coefficients from the levels model that included both signed jump variation and bipower variation, $\overline{RV}_{h,i,t+h} = \mu_i + \phi^{J+}\left(RS_{i,t}^+ - RS_{i,t}^-\right)I_{i,t} + \phi^{J-}\left(RS_{i,t}^+ - RS_{i,t}^-\right)\left(1 - I_{i,t}\right) + \phi^C BV_{i,t}^- + \phi_5\overline{RV}_{5,i,t} + \phi_{22}\overline{RV}_{22,i,t} + \epsilon_{i,t}$., along with 95% confidence intervals for the panel of individual volatilities. 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.
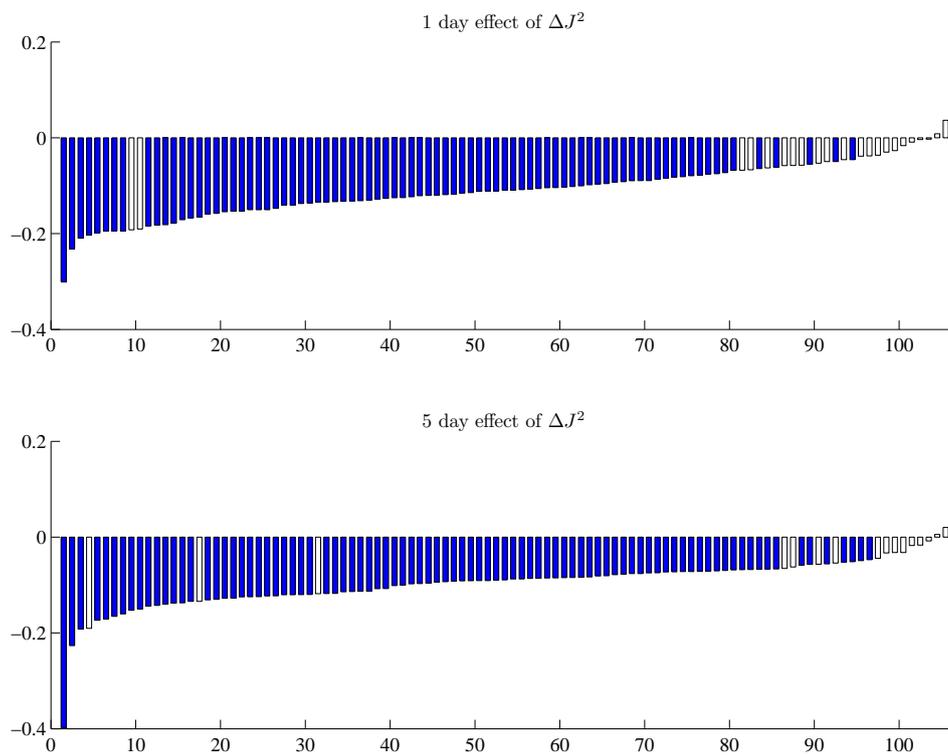
Figure 3: Sorted effects of the signed jump variation in the individual firm volatilities. The magnitude of the coefficient on ΔJ^2 is indicated as distance from the horizontal axis. Solid bars indicate significance at the 5% level.
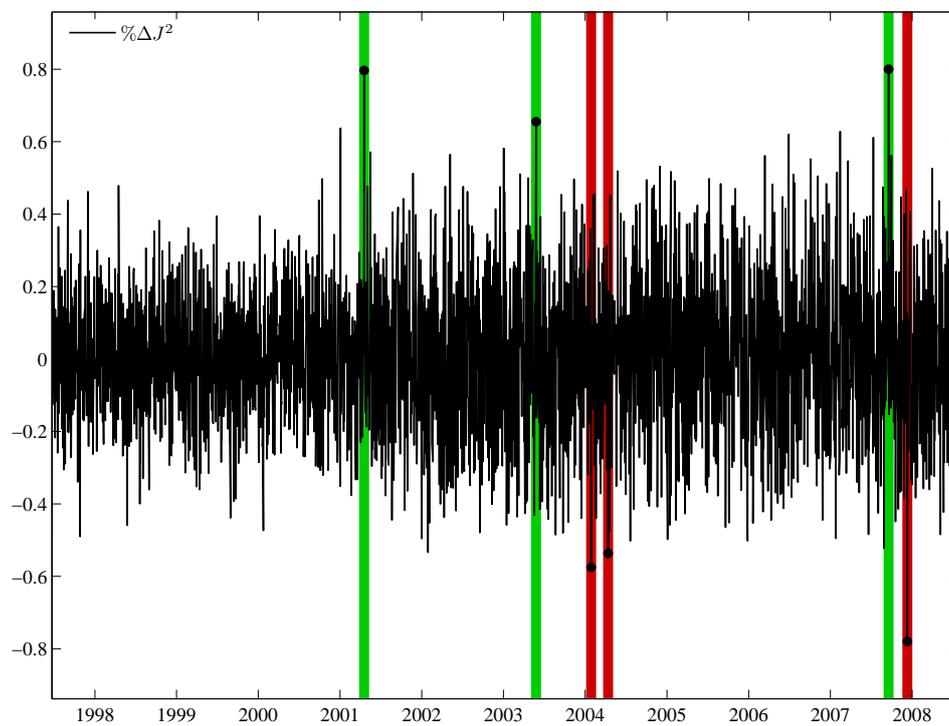
Figure 4: Time series of the percentage signed jump variation ($\%\Delta J_t^2 = (RS_t^+ - RS_t^-)/RV_t$) for the S&P 500 SPDR. Green (red) highlighting corresponds to the "good" ("bad") volatility days depicted in Figure 5.
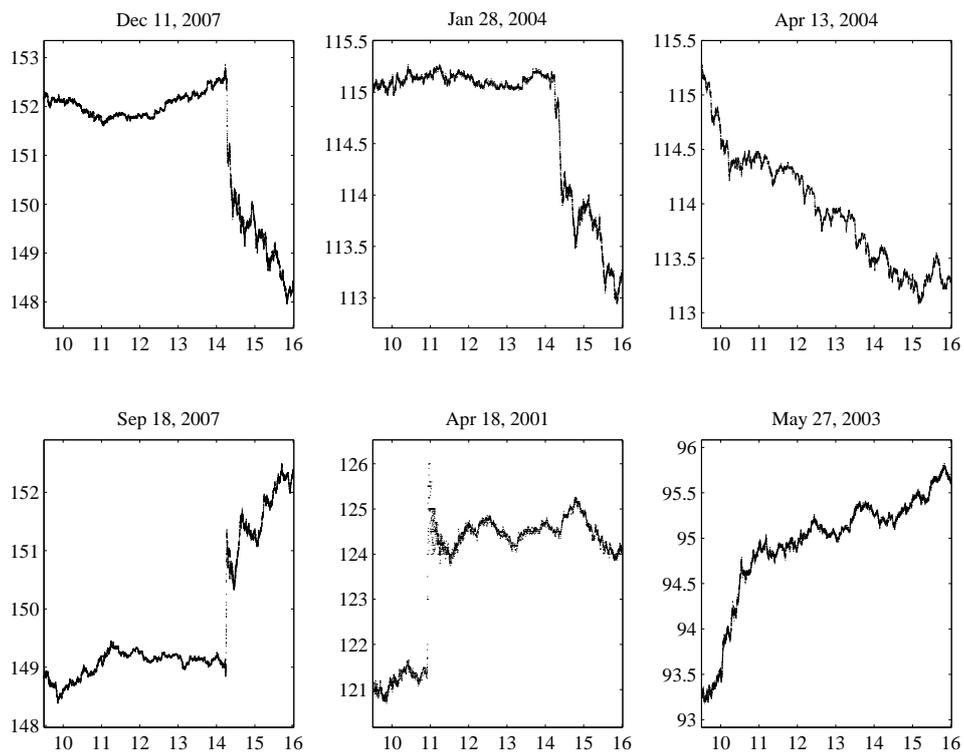
Figure 5: Intra-daily prices for the S&P 500 SPDR on the six dates identified in figure 4. The top row corresponds to "bad volatility", where the leftmost figure had the lowest $\%\Delta J_t^2$ across all dates, and the bottom row corresponds to "good volatility" where the leftmost figure had the highest $\%\Delta J_t^2$.

HAR estimation results for the SPDR, cumulative volatility

$$\overline{RV}_{h,i,t+h} = \mu_i + \phi_1 RV_{i,t} + \phi_1^+ RS_{i,t}^+ + \phi_1^- RS_{i,t}^- + \gamma RV_{i,t} I_{[r_{i,t-1}<0]} + \phi_5 \overline{RV}_{5,i,t} + \phi_{22}\overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	ϕ_1	ϕ_1^+	ϕ_1^-	γ	ϕ_5	ϕ_{22}	R^2
$h=1$	0.607 (17.0)				0.268 (8.1)	0.120 (4.9)	0.532
		−0.012 (−0.3)	0.591 (13.0)		0.291 (9.3)	0.120 (4.9)	0.611
		0.018 (0.4)	0.532 (7.4)	0.025 (1.4)	0.293 (9.3)	0.121 (5.0)	0.611
$h=5$	0.425 (14.7)				0.409 (8.6)	0.158 (4.0)	0.563
		−0.015 (−0.7)	0.431 (13.4)		0.421 (8.8)	0.155 (4.0)	0.620
		0.036 (1.1)	0.325 (6.6)	0.046 (2.5)	0.424 (8.8)	0.157 (4.0)	0.619
$h=22$	0.305 (11.8)				0.357 (7.7)	0.265 (4.8)	0.468
		−0.005 (−0.3)	0.314 (9.9)		0.359 (7.5)	0.261 (4.8)	0.508
		−0.006 (−0.2)	0.317 (5.8)	−0.001 (−0.1)	0.359 (7.6)	0.261 (4.8)	0.508
$h=66$	0.203 (8.4)				0.256 (7.4)	0.299 (5.3)	0.282
		−0.033 (−2.2)	0.250 (7.3)		0.253 (6.8)	0.294 (5.3)	0.313
		−0.060 (−1.8)	0.311 (4.1)	−0.027 (−1.3)	0.251 (6.8)	0.292 (5.2)	0.315

Table 1a: Reference, base and asymmetric model parameter estimates using h-day cumulative variance as the dependent variance (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 R^2 measure is constructed using the WLS parameter estimates and the original data.

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

$$\overline{RV}_{h,i,t+h} = \mu_i + \phi_1 RV_{i,t} + \phi_1^+ RS_{i,t}^+ + \phi_1^- RS_{i,t}^- + \gamma RV_{i,t} I_{[r_{i,t-1}<0]} + \phi_5 \overline{RV}_{5,i,t} + \phi_{22} \overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	ϕ_1	ϕ_1^+	ϕ_1^-	γ	ϕ_5	ϕ_{22}	R^2
$h=1$	0.488 (39.7)				0.315 (28.1)	0.172 (16.2)	0.550
		0.134 (15.6)	0.352 (24.5)		0.317 (28.7)	0.172 (16.4)	0.560
		0.158 (17.1)	0.304 (18.9)	0.023 (6.8)	0.317 (28.7)	0.172 (16.5)	0.561
$h=5$	0.357 (23.2)				0.357 (16.3)	0.247 (10.2)	0.616
		0.079 (11.4)	0.276 (19.0)		0.359 (16.4)	0.247 (10.3)	0.630
		0.105 (14.5)	0.222 (19.6)	0.026 (7.7)	0.359 (16.5)	0.248 (10.3)	0.631
$h=22$	0.241 (14.8)				0.312 (11.9)	0.360 (10.3)	0.544
		0.045 (7.3)	0.194 (12.8)		0.314 (11.9)	0.360 (10.4)	0.556
		0.063 (9.4)	0.157 (13.8)	0.018 (5.1)	0.314 (11.9)	0.360 (10.4)	0.557
$h=66$	0.161 (12.2)				0.236 (10.7)	0.431 (12.0)	0.451
		0.022 (3.9)	0.138 (8.8)		0.238 (10.6)	0.431 (12.1)	0.462
		0.031 (5.9)	0.117 (8.5)	0.010 (3.8)	0.238 (10.6)	0.432 (12.1)	0.463

Table 1b: Reference, base and asymmetric model parameter estimates using h-day cumulative variance as the dependent variance (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 R^2 measure is constructed using the WLS parameter estimates and the original data and are the average of the 105 R^2s for the individual assets.

HAR estimation results for the SPDR, one-day volatility

$$RV_{i,t+h} = \mu_i + \phi_1 RV_{i,t} + \phi_1^+ RS_{i,t}^+ + \phi_1^- RS_{i,t}^- + \gamma I_{[r_{i,t-1}<0]} RV_{i,t} + \phi_5 I_{[r_{i,t-1}<0]} \overline{RV}_{5,i,t} + \phi_{22} I_{[r_{i,t-1}<0]} \overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	ϕ_1	ϕ_1^+	ϕ_1^-	γ	ϕ_5	ϕ_{22}	R^2
$h = 5$	0.270 (6.5)				0.526 (6.0)	0.177 (3.3)	0.327
		−0.007 (−0.3)	0.279 (6.1)		0.528 (6.1)	0.173 (3.3)	0.341
		0.036 (1.0)	0.185 (2.7)	0.042 (1.9)	0.532 (6.1)	0.175 (3.3)	0.340
$h = 22$	0.148 (3.2)				0.410 (4.1)	0.250 (2.5)	0.128
		−0.052 (−1.5)	0.220 (4.0)		0.401 (4.0)	0.245 (2.5)	0.138
		−0.055 (−1.1)	0.225 (2.5)	−0.002 (−0.1)	0.401 (4.0)	0.245 (2.5)	0.138
$h = 66$	0.038 (1.9)				0.159 (2.7)	0.257 (3.2)	0.021
		−0.053 (−1.7)	0.110 (2.6)		0.150 (2.6)	0.253 (3.2)	0.022
		−0.066 (−1.8)	0.143 (2.5)	−0.015 (−0.8)	0.149 (2.6)	0.252 (3.1)	0.022

Table 2a: Reference, base and asymmetric model parameter estimates using day-h variance as the dependent variable (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 R^2 measure is constructed using the WLS parameter estimates and the original data.

HAR estimation results for the panel of 105 individual stocks, one-day volatility

$$RV_{i,t+h} = \mu_i + \phi_1 RV_{i,t} + \phi_1^+ RS_{i,t}^+ + \phi_1^- RS_{i,t}^- + \gamma I_{[r_{i,t-1}<0]} RV_{i,t} + \phi_5 I_{[r_{i,t-1}<0]}\overline{RV}_{5,i,t} + \phi_{22} I_{[r_{i,t-1}<0]}\overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	ϕ_1	ϕ_1^+	ϕ_1^-	γ	ϕ_5	ϕ_{22}	R^2
$h=5$	0.273 (14.3)				0.376 (11.4)	0.299 (8.9)	0.389
		0.059 (5.2)	0.212 (13.1)		0.377 (11.4)	0.299 (8.9)	0.393
		0.081 (7.1)	0.167 (13.1)	0.022 (4.6)	0.378 (11.4)	0.299 (9.0)	0.394
$h=22$	0.161 (7.2)				0.242 (5.6)	0.453 (8.2)	0.217
		0.010 (0.8)	0.150 (7.6)		0.243 (5.7)	0.453 (8.2)	0.220
		0.024 (2.2)	0.118 (6.3)	0.016 (3.7)	0.244 (5.7)	0.454 (8.2)	0.221
$h=66$	0.109 (7.7)				0.173 (5.0)	0.424 (6.7)	0.116
		0.003 (0.2)	0.105 (5.5)		0.174 (5.0)	0.423 (6.7)	0.117
		0.011 (1.1)	0.086 (6.3)	0.010 (1.5)	0.174 (5.0)	0.424 (6.7)	0.118

Table 2b: Reference, base and asymmetric model parameter estimates using day-h variance as the dependent variable (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 R^2 measure is constructed using the WLS parameter estimates and the original data and are the average of the 105 R^2s for the individual assets.

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

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_1^+ RS_{i,t}^+ + \phi_5^+ \overline{RS}_{5,i,t}^+ + \phi_{22}^+ \overline{RS}_{22,i,t}^+ + \phi_1^- RS_{i,t}^- + \phi_5^- \overline{RS}_{5,i,t}^- + \phi_{22}^- \overline{RS}_{22,i,t}^- + \epsilon_{i,t}$$

	RM	ϕ_1^+	ϕ_1^-	ϕ_5^+	ϕ_5^-	ϕ_{22}^+	ϕ_{22}^-	R^2
$h=1$	RV	-0.044 (−1.3)	0.581 (12.7)	-0.090 (−2.5)	0.394 (9.1)	-0.116 (−1.9)	0.268 (3.4)	0.621
	RS^+	-0.084 (−1.9)	0.644 (10.0)	-0.110 (−2.9)	0.404 (8.1)	-0.069 (−1.0)	0.219 (2.3)	0.539
	RS^-	-0.003 (−0.1)	0.518 (14.9)	-0.069 (−1.7)	0.384 (8.0)	-0.163 (−2.3)	0.317 (3.6)	0.621
$h=5$	RV	-0.035 (−1.5)	0.418 (13.4)	0.005 (0.1)	0.408 (5.9)	-0.237 (−2.4)	0.440 (3.7)	0.642
	RS^+	-0.049 (−1.9)	0.447 (12.6)	0.015 (0.2)	0.405 (5.8)	-0.207 (−2.0)	0.403 (3.2)	0.635
	RS^-	-0.022 (−0.9)	0.389 (13.6)	-0.004 (−0.1)	0.410 (5.9)	-0.267 (−2.6)	0.477 (3.9)	0.623
$h=22$	RV	-0.029 (−1.3)	0.291 (10.3)	-0.109 (−1.3)	0.454 (5.9)	-0.460 (−1.8)	0.804 (2.6)	0.565
	RS^+	-0.031 (−1.3)	0.301 (10.2)	-0.107 (−1.2)	0.451 (5.5)	-0.494 (−1.9)	0.851 (2.6)	0.583
	RS^-	-0.027 (−1.3)	0.282 (10.3)	-0.110 (−1.4)	0.457 (6.1)	-0.425 (−1.7)	0.756 (2.5)	0.538
$h=66$	RV	-0.060 (−2.7)	0.224 (7.9)	-0.206 (−2.5)	0.466 (5.0)	-0.418 (−1.3)	0.790 (2.3)	0.364
	RS^+	-0.064 (−2.8)	0.235 (8.0)	-0.218 (−2.5)	0.488 (5.0)	-0.444 (−1.3)	0.819 (2.3)	0.378
	RS^-	-0.055 (−2.7)	0.212 (7.8)	-0.193 (−2.4)	0.444 (5.0)	-0.392 (−1.3)	0.761 (2.4)	0.347

Table 3a: Extended model where RV at all lags is decomposed into positive and negative semivariance (t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates which variable is 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 data.

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

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_1^+ RS_{i,t}^+ + \phi_5^+ \overline{RS}_{5,i,t}^+ + \phi_{22}^+ \overline{RS}_{22,i,t}^+ + \phi_1^- RS_{i,t}^- + \phi_5^- \overline{RS}_{5,i,t}^- + \phi_{22}^- \overline{RS}_{22,i,t}^- + \epsilon_{i,t}$$

	RM	ϕ_1^+	ϕ_1^-	ϕ_5^+	ϕ_5^-	ϕ_{22}^+	ϕ_{22}^-	R^2
$h=1$	RV	0.131 (15.3)	0.348 (24.3)	0.053 (5.5)	0.265 (25.1)	0.033 (2.5)	0.145 (11.0)	0.564
	RS^+	0.144 (14.5)	0.339 (18.3)	0.061 (6.0)	0.253 (22.6)	0.058 (4.1)	0.118 (8.3)	0.520
	RS^-	0.118 (13.5)	0.357 (28.7)	0.045 (4.1)	0.277 (23.4)	0.007 (0.5)	0.172 (11.1)	0.519
$h=5$	RV	0.076 (10.7)	0.271 (19.2)	0.068 (3.7)	0.289 (16.5)	0.017 (0.5)	0.239 (7.9)	0.638
	RS^+	0.079 (11.0)	0.272 (18.0)	0.079 (4.3)	0.279 (16.4)	0.038 (1.3)	0.214 (7.6)	0.636
	RS^-	0.073 (10.0)	0.270 (20.0)	0.058 (3.1)	0.298 (15.9)	-0.005 (-0.2)	0.265 (7.8)	0.609
$h=22$	RV	0.042 (6.4)	0.189 (13.2)	0.029 (1.3)	0.281 (11.9)	0.011 (0.2)	0.362 (5.2)	0.579
	RS^+	0.045 (6.8)	0.189 (12.8)	0.035 (1.6)	0.275 (11.9)	0.015 (0.3)	0.358 (5.2)	0.591
	RS^-	0.039 (5.9)	0.188 (13.5)	0.022 (1.0)	0.286 (11.8)	0.008 (0.1)	0.366 (5.2)	0.558
$h=66$	RV	0.019 (2.9)	0.133 (9.4)	-0.000 (-0.0)	0.234 (7.1)	0.060 (0.7)	0.385 (4.7)	0.490
	RS^+	0.020 (3.1)	0.134 (9.2)	0.004 (0.1)	0.233 (7.0)	0.066 (0.8)	0.378 (4.5)	0.497
	RS^-	0.017 (2.6)	0.132 (9.5)	-0.005 (-0.2)	0.235 (7.2)	0.053 (0.7)	0.392 (4.9)	0.478

Table 3b: Extended model where RV at all lags is decomposed into positive and negative semivariance (t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates which variable is 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 data and is the average across the R^2s of the 105 individual assets.

Tests on parameters of the vector HAR model

	No RS^+	No RS^-	$\phi^+ = \phi^-$	% Neg.	$\phi_1^+ = \phi_1^-$	% Neg.	$\lambda_1^+ = \lambda_1^-$	% Neg.
Both	81.9	99.0	–	–	–	–	89.6	91.6
RS^+	81.9	99.0	97.2	98.2	61.9	96.9	–	–
RS^-	74.3	99.0	76.2	98.8	78.1	95.1	–	–

Table 4: Results from the VHAR fit to the two semivariance components of realized variance. All values represent the percentage of nulls rejected using a 5% test. The left column indicates the dependent variable used in the test. The first panel contains results from a test that one of the realized semivariance measures could be excluded. The second panel tests equality of the sum of the coefficients on the two realized semivariance measures and % Neg. reports the percentage of the rejections where the sum of coefficients on the positive semivariance was smaller than on the negative semivariance. The third panel contains the same test only using the first lag coefficient. The final panel contains results of a test of equal persistence and the percentage of rejections where positive realized semivariance is less significant than negative realized semivariance.

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

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_J \Delta J_{i,t}^2 + \phi_{J^+} \Delta J_{i,t}^{2+} + \phi_{J^-} \Delta J_{i,t}^{2-} + \phi_C BV_{i,t} + \phi_5 \overline{RV}_{5,i,t} + \phi_{22} \overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_5	ϕ_{22}	R^2
$h=1$	RV	-0.286 (-7.7)			0.610 (18.4)	0.282 (9.0)	0.120 (5.0)	0.613
	BV	-0.275 (-9.5)			0.596 (20.4)	0.278 (10.2)	0.098 (5.0)	0.663
	RV		-0.095 (-2.1)	-0.482 (-5.5)	0.545 (16.8)	0.289 (9.4)	0.120 (5.0)	0.621
$h=5$	RV	-0.204 (-9.0)			0.449 (13.6)	0.406 (8.5)	0.154 (4.0)	0.622
	BV	-0.196 (-9.2)			0.440 (15.0)	0.389 (8.6)	0.137 (3.9)	0.633
	RV		-0.142 (-3.9)	-0.272 (-6.1)	0.426 (11.9)	0.409 (8.6)	0.154 (4.0)	0.622
$h=22$	RV	-0.138 (-7.3)			0.342 (11.2)	0.342 (7.1)	0.260 (4.8)	0.512
	BV	-0.132 (-7.0)			0.333 (10.9)	0.330 (6.8)	0.243 (4.7)	0.505
	RV		-0.150 (-3.8)	-0.124 (-2.3)	0.346 (10.0)	0.341 (7.0)	0.260 (4.8)	0.513
$h=66$	RV	-0.122 (-5.4)			0.240 (9.6)	0.240 (6.3)	0.293 (5.3)	0.312
	BV	-0.117 (-5.1)			0.232 (9.2)	0.231 (6.1)	0.279 (5.3)	0.304
	RV		-0.123 (-3.1)	-0.121 (-1.9)	0.240 (8.3)	0.240 (6.0)	0.293 (5.3)	0.312

Table 5a: Models that includes signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates which variable is 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 data. R^2 values in the final column are the average across the R^2s of the 105 individual assets.

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

$$\overline{RM}_{h,i,t+h} = \mu_i + \phi_J \Delta J^2_{i,t} + \phi_{J^+} \Delta J^{2+}_{i,t} + \phi_{J^-} \Delta J^{2-}_{i,t} + \phi_C BV_{i,t} + \phi_5 \overline{RV}_{5,i,t} + \phi_{22} \overline{RV}_{22,i,t} + \epsilon_{i,t}$$

	RM	ϕ_J	ϕ_{J^+}	ϕ_{J^-}	ϕ_C	ϕ_5	ϕ_{22}	R^2
$h=1$	RV	-0.108 (-10.5)			0.563 (40.0)	0.327 (29.1)	0.182 (17.2)	0.564
	BV	-0.089 (-10.9)			0.488 (41.9)	0.258 (28.3)	0.136 (16.2)	0.572
	RV		0.024 (2.5)	-0.246 (-12.2)	0.502 (38.1)	0.330 (29.5)	0.182 (17.3)	0.566
$h=5$	RV	-0.097 (-11.5)			0.411 (23.0)	0.366 (16.5)	0.255 (10.5)	0.633
	BV	-0.080 (-11.6)			0.358 (23.6)	0.293 (16.0)	0.196 (10.1)	0.633
	RV		-0.058 (-5.2)	-0.139 (-12.2)	0.392 (20.9)	0.367 (16.6)	0.255 (10.5)	0.634
$h=22$	RV	-0.073 (-8.7)			0.279 (13.9)	0.318 (12.1)	0.366 (10.4)	0.558
	BV	-0.060 (-8.6)			0.244 (13.8)	0.256 (11.4)	0.287 (10.0)	0.553
	RV		-0.061 (-5.3)	-0.086 (-7.8)	0.273 (13.0)	0.319 (12.1)	0.366 (10.4)	0.559
$h=66$	RV	-0.057 (-5.6)			0.180 (11.5)	0.244 (10.7)	0.437 (12.1)	0.461
	BV	-0.046 (-5.4)			0.158 (11.4)	0.194 (10.0)	0.348 (12.4)	0.457
	RV		-0.040 (-4.3)	-0.075 (-4.9)	0.171 (11.5)	0.245 (10.6)	0.437 (12.1)	0.463

Table 5b: Models that includes signed jump information where quadratic variation has been decomposed into signed jump variation, ΔJ^2, and its continuous component using bipower variation, BV (t-statistics in parentheses). Each of the four panels contains results for the forecast horizon indicated at the left. RM indicates which variable is 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 $\Delta J^{2+}_{i,t} = \Delta J^2_{i,t} I_{[RS^+ - RS^- > 0]}$. The R^2 measure is constructed using the WLS parameter estimates and the original data. R^2 values in the final column are the average across the R^2s of the 105 individual assets.

Results for models of individual stock volatility

	Baseline				Jumps		Extended Jumps			
	RS^+		RS^-				J^+		J^-	
h	Sig. +	Sig -	Sig +	Sig -	Sig +	Sig -	Sig +	Sig -	Sig +	Sig -
1	84.8	0.0	99.0	0.0	0.0	79.2	7.5	2.8	0.0	82.1
5	77.1	0.0	99.0	0.0	0.0	84.9	0.9	35.8	0.9	65.1
22	47.6	0.0	99.0	0.0	0.0	81.1	0.9	33.0	0.9	43.4
66	26.7	1.0	99.0	0.0	0.0	62.3	0.9	17.9	0.9	35.8

Table 6: This table contains percentage of null hypotheses rejected for the Baseline, Jump, and Extended Jump model for individually fit models for the 105 S&P 100 constituents. Two columns are presented for each term, the left of which indicates the percentage of parameters that are significantly positive, the other indicating the percentage significantly negative. For example, for 5-step ahead forecasts in the baseline specification (equation 17) the negative realized semivariance was significantly positive in 99% of the series while the positive realized semivariance was significantly positive in 77.1% of the series.

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

	$\widehat{RV}_1 - \widehat{RV}_2$			$\widehat{RV}_1 - \widehat{RV}_2'$			$\widehat{RV}_1' - \widehat{RV}_2'$		
	SPDR	Other Assets		SPDR	Other Assets		SPDR	Other Assets	
	DM	Favor RS	Favor RV	DM	Favor RS	Favor RV	DM	Favor RS	Favor RV
$h=1$	1.48	22.1	2.1	2.69	12.6	3.2	0.02	2.1	9.5
$h=5$	3.16	29.5	2.1	4.29	33.7	5.3	1.25	10.5	3.2
$h=22$	3.96	26.3	2.1	4.41	33.7	1.1	3.07	16.8	2.1
$h=66$	3.64	23.2	1.1	5.23	36.8	1.1	4.27	16.8	3.2

Table 7: Each of the three panels corresponds to results for test of equal predictive accuracy. The left most column contains the Diebold-Mariano-Giacomini-White test statistic for the S&P 500 SPDR. In all cases, a positive test statistic indicates the realized semivariance model outperformed the realized variance model. The remaining two columns contain to the percentage of the 83 individual series which reject the null, and the direction of the rejection using a 5% 2-sided test. The model are $(\widehat{RV}_1)$ standard realized variance HAR, $(\widehat{RV}_1')$ realized variance HAR with a daily asymmetry, $(\widehat{RV}_2)$ realized semivariance HAR and $(\widehat{RV}_2')$ realized semivariance HAR with no positive semivariance.