Measuring downside risk — realised semivariance

OLE E. BARNDORFF-NIELSEN
The T.N. Thiele Centre for Mathematics in Natural Science, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark oebn@imf.au.dk

SILJA KINNEBROCK
Oxford-Man Institute, University of Oxford, Blue Boar Court, 9 Alfred Street, Oxford OX1 4EH, UK & Merton College, University of Oxford silja.kinnebrouck@oxford-man.ox.ac.uk

NEIL SHEPHARD
Oxford-Man Institute, University of Oxford, Blue Boar Court, 9 Alfred Street, Oxford OX1 4EH, UK & Department of Economics, University of Oxford neil.shephard@economics.ox.ac.uk

January 9, 2008

Abstract

We propose a new measure of risk, based entirely on downwards moves measured using high frequency data. Realised semivariances are shown to have important predictive qualities for future market volatility. The theory of these new measures is spelt out, drawing on some new results from probability theory.

Keywords: Market frictions; Quadratic variation; Realised variance; Semimartingale; Semivariance.
“It was understood that risk relates to an unfortunate event occurring, so for an investment this corresponds to a low, or even negative, return. Thus getting returns in the lower tail of the return distribution constitutes this “downside risk.” However, it is not easy to get a simple measure of this risk. The “semi-variance,” which is the variance measured over the distribution below some value (such as the median) is easy to define but is mathematically intractable. Because of these practical difficulties attention was turned to measuring risk by the variance.” Quoted from Granger (2008).

1 Introduction

A number of economists have wanted to measure downside risk, the risk of prices falling, just using information based on negative returns — a prominent recent example is by Ang, Chen, and Xing (2006). This has been operationalised by quantities such as semivariance, value at risk and expected loss, which are typically estimated using daily returns. In this paper we introduce a new measure of the variation of asset prices based on high frequency data. It is called realised semivariance (RS). We derive its limiting properties, relating it to quadratic variation and, in particular, negative jumps. Further, we show it has some useful properties in empirical work, enriching the standard ARCH models pioneered by Rob Engle over the last 25 years and building on the recent econometric literature on realised volatility.

Realised semivariance extends the influential work of, for example, Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002), on formalising so-called realised variances (RV) which links these commonly used statistics to the quadratic variation process. Realised semivariance measures the variation of asset price falls. At a technical level it can be regarded as a continuation of the work of Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen and Shephard (2006), who showed it is possible to go inside the quadratic variation process and separate out components of the variation of prices into that due to jumps and that due to the continuous evolution. This work has prompted papers by, for example, Andersen, Bollerslev, and Diebold (2007), Huang and Tauchen (2005) and Lee and Mykland (2008) on the importance of this decomposition empirically in economics. Surveys of this kind of thinking are provided by Andersen, Bollerslev, and Diebold (2006) and Barndorff-Nielsen and Shephard (2007), while a lengthy discussion of the relevant probability theory is given in Jacod (2007).

Let us start with statistics and results which are well known. Realised variance (RV) estimates the ex-post variance of asset prices over a fixed time period. We will suppose that this period is 0 to 1. In our applied work it can be thought of as any individual day of interest. Then RV is
defined as

\[ RV = \sum_{j=1}^{t_1 \leq 1} (Y_{t_j} - Y_{t_{j-1}})^2. \]

where 0 = t_0 < t_1 < ... < t_n = 1 are the times at which (trade or quote) prices are available. For arbitrage free-markets, Y must follow a semimartingale. This estimator converges as we have more and more data in that interval to \([Y]_1\), where

\[ [Y]_t = \lim_{n \to \infty} \sum_{j=1}^{t \leq t_j} (Y_{t_j} - Y_{t_{j-1}})^2, \]

(e.g. Protter (2004, p. 66–77)) for any sequence of deterministic partitions 0 = t_0 < t_1 < ... < t_n = t with \(\sup_j \{t_{j+1} - t_j\} \to 0\) for \(n \to \infty\). This limiting operation is often referred to as “in-fill asymptotics” in statistics and econometrics.

One of the initially strange things about realised variance is that it solely uses squares of the data, while the research of, for example, Black (1976), Nelson (1991), Glosten, Jagannathan, and Runkle (1993) and Engle and Ng (1993) has indicated the importance of falls in prices as a driver of conditional variance. The reason for this is clear, as the high frequency data becomes dense, the extra information in the sign of the data can fall to zero. The most elegant framework in which to see this is where Y is a Brownian semimartingale

\[ Y_t = \int_0^t a_s \, ds + \int_0^t \sigma_s \, dW_s, \quad t \geq 0, \]

where \(a\) is a locally bounded predictable drift process and \(\sigma\) is a càdlàg volatility process – all adapted to some common filtration \(\mathcal{F}_t\), implying the model can allow for classic leverage effects. For such a process

\[ [Y]_t = \int_0^t \sigma_s^2 \, ds, \]

and so

\[ d[Y]_t = \sigma_t^2 \, dt, \]

which means for a Brownian semimartingale the QV process tells us everything we can know about the ex-post variation of \(Y\). The signs of the returns are irrelevant in the limit — this is true whether there is leverage or not.

\footnote{When there are market frictions it is possible to correct this statistic for their effect using the two scale estimator of Zhang, Mykland, and Aït-Sahalia (2005), the realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) or the pre-averaging based statistic of Jacod, Li, Mykland, Podolskij, and Vetter (2007).}
If there are jumps in the process there are additional things to learn than just the QV process. Let

\[ Y_t = \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t, \]

where \( J \) is a pure jump process. Then, writing jumps in \( Y \) as \( \Delta Y_t = Y_t - Y_{t-} \), then

\[ [Y]_t = \int_0^t \sigma_s^2 ds + \sum_{s \leq t} (\Delta Y_s)^2, \]

and so QV aggregates two sources of risk. Even when we employ bipower variation (Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen and Shephard (2006)), which allows us to estimate \( \int_0^t \sigma_s^2 ds \) robustly to jumps, this still leaves us with estimates of \( \sum_{s \leq t} (\Delta J_s)^2 \). This tells us nothing about the asymmetric behaviour of the jumps — which is important if we wish to understand downside risk.

In this paper we introduce the realised semivariances

\[ RS = \sum_{j=1}^{t_j \leq 1} (Y_{t_j} - Y_{t_{j-1}})^2 1_{Y_{t_j} - Y_{t_{j-1}} \leq 0}, \]

where \( 1_y \) is the indicator function taking the value 1 of the argument \( y \) is true. We will study the behaviour of this statistic under in-fill asymptotics. In particular we will see that

\[ RS \xrightarrow{p} \frac{1}{2} \int_0^t \sigma_s^2 ds + \sum_{s \leq 1} (\Delta Y_s)^2 1_{\Delta Y_s \leq 0}, \]

under in-fill asymptotics. Hence RS provides a new source of information, one which focuses on squared negative jumps.\(^3\)

Semivariances, or more generally measures of variation below a threshold (target semivariance) have a long history in finance. The first references are probably Markowitz (1959), Mao (1970b), Mao (1970a), Hogan and Warren (1972) and Hogan and Warren (1974). Examples include the work of Fishburn (1977) and Lewis (1990). Sortino ratios (which are an extension of Sharpe ratios and were introduced by Sortino and van der Meer (1991)), and the so-called post-modern portfolio theory by, for example, Rom and Ferguson (1993), have attracted considerable attention. Sortino and Satchell (2001) look at recent developments and provide a review, while Pedersen and Satchell (2002) look at the economic theory of this measure of risk. Of course these types of measures are likely not to be very informative for exchange rate investments or for the individual holdings of

\(^2\)Threshold based decompositions have also been suggested in the literature, examples of this include Mancini (2001), Jacob (2007) and Lee and Mykland (2008).

\(^3\)This type of statistic relates to the work of Babsiria and Zakolskian (2001) who built separate ARCH type conditional variance models of daily returns using positive and negative daily returns. It also resonates with the empirical results in a recent paper by Chen and Ghysels (2007) on news impact curves estimated through semi-parametric MIDAS regressions.
hedge funds which can go either short (hence a fall in price can lead to a positive return and hence is desirable) or long. Our innovation is to bring high frequency analysis to bear on this measure of risk.

The empirical essence of daily realised semivariance can be gleaned from Figure 1 which shows an analysis of trades on General Electric (GE) carried out on the New York Stock Exchange\textsuperscript{4} from 1995 to 2005 (giving us 2,616 days of data). In graph (a) we show the path of the trades drawn in trading time on a particular randomly chosen day in 2004, to illustrate the amount of daily trading which is going on in this asset. Notice by 2004 the tick size has fallen to one cent.

Graph (b) shows the open to close returns, measured on the log-scale and multiplied by 100, which indicates some moderation in the volatility during the last and first piece of the sample period. The corresponding daily realised volatility (the square root of the realised variance) is plotted in graph (c), based upon returns calculated every 15 trades. The Andersen, Bollerslev, Diebold, and Labys (2000) variance signature plot is shown in graph (d), to assess the impact of noise on the calculation of realised volatility. It suggests statistics computed on returns calculated every 15 trades should not be too sensitive to noise for GE. Graph (e) shows the same but focusing on daily RS and RV-RS. Throughout, the statistics are computed using returns calculated every 15 trades. It indicates they are pretty close to one another on average over this sample period. This component signature plot is in the spirit of the analysis pioneered by Andersen, Bollerslev, Diebold, and Labys (2001) in their analysis of realised variance. Graph (f) shows the correlogram for the realised semivariance and the realised variance and suggests the realised semivariance has much more dependence in it than RV-RS. Some summary statistics for this data are available in Table 2, which will be discussed in some detail in Section 3.

In the realised volatility literature, authors have typically worked out the impact of using realised volatilities on volatility forecasting using regressions of future realised variance on lagged realised variance and various other explanatory variables\textsuperscript{5}. Engle and Gallo (2006) prefers a different route, which is to add lagged realised quantities as variance regressors in Engle (2002) and Bollerslev (1986) GARCH type models of daily returns — the reason for their preference is that it is aimed at a key quantity, a predictive model of future returns, and is more robust to the heteroskedasticity inherent in the data. Typically when Engle generalises to allow for leverage he uses the Glosten, Jagannathan, and Runkle (1993) (GJR) extension. This is the method we follow here. Throughout we will use the subscript $i$ to denote discrete time.

\textsuperscript{4}This data is taken from the TAQ database, managed through WRDS. Although information on trades is available from all the different exchanges in the U.S., we solely study trades which are made at the exchange in New York.

\textsuperscript{5}Leading references include Andersen, Bollerslev, Diebold, and Labys (2001) and Andersen, Bollerslev, and Meddahi (2004).
Analysis of the General Electric share price from 1995 to 2005

Figure 1: Analysis of trades on General Electric carried out on the NYSE from 1995 to 2005. (a): path of the trades drawn in trading time on a random day in 2004. (b): daily open to close returns \( r_i \), measured on the log-scale and multiplied by 100. The corresponding daily realised volatility \( \sqrt{RV_i} \), is plotted in graph (c), based upon returns calculated every 15 trades. (d): variance signature plot in trade time to assess the impact of noise on the calculation of realised variance \( RV_i \). (e): same thing, but for the realised semivariance \( RS_i \) and \( RV_i - RS_i \). (f) correlogram for \( RS_i \), \( RV_i \) and \( RV_i - RS_i \). Code: downside.ox.

We model daily open to close returns \( \{ r_i; i = 1, 2, ..., T \} \) as

\[
E( r_i | G_{i-1} ) = \mu, \\
h_i = \text{Var} ( r_i | G_{i-1} ) = \omega + \alpha ( r_{i-1} - \mu)^2 + \beta h_{i-1} + \delta ( r_{i-1} - \mu)^2 I_{r_{i-1} - \mu < 0} + \gamma' z_{i-1},
\]

and then use a standard Gaussian quasi-likelihood to make inference on the parameters, e.g. Bollerslev and Wooldridge (1992). Here \( z_{i-1} \) are the lagged daily realised regressors and \( G_{i-1} \) is the information set generated by discrete time daily statistics available to forecast \( r_i \) at time \( i - 1 \).

Table 1 shows the fit of the GE trade data from 1995-2005. It indicates the lagged RS beating out of the GARCH model (\( \delta = 0 \)) and the lagged RV. Both realised terms yield large likelihood improvements over a standard daily returns based GARCH. Importantly there is a vast shortening in the information gathering period needed to condition on, with the GARCH memory parameter \( \beta \) dropping from 0.953 to around 0.7. This makes fitting these realised based models much easier in practice, allowing their use on relatively short time series of data.
When the comparison with the GJR model is made, which allows for traditional leverage effects, the results are more subtle, with the realised semivariance significantly reducing the importance of the traditional leverage effect while the high frequency data still has an important impact on improving the fit of the model. In this case the RS and RV play similar roles, with RS no longer dominating the impact of the RV in the model.

When the comparison with the GJR model is made, which allows for traditional leverage effects, the results are more subtle, with the realised semivariance significantly reducing the importance of the traditional leverage effect while the high frequency data still has an important impact on improving the fit of the model. In this case the RS and RV play similar roles, with RS no longer dominating the impact of the RV in the model.

<table>
<thead>
<tr>
<th>ARCH type models and lagged realised semivariance and variance</th>
<th>GARCH</th>
<th>GJR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged RS</td>
<td>0.685 (2.78)</td>
<td>0.499 (2.86)</td>
</tr>
<tr>
<td>Lagged RV</td>
<td>-0.114 (−1.26)</td>
<td>0.228 (3.30)</td>
</tr>
</tbody>
</table>
| AR
| GARCH                                                      | 0.040 (2.23) | 0.036 (2.068) |
| GJR                                                       | 0.711 (7.79) | 0.691 (7.071) |

Table 1: Gaussian quasi-likelihood fit of GARCH and GJR models fitted to daily open to close returns on General Electric share prices, from 1995 to 2005. We allow lagged daily realised variance (RV) and realised semivariance (RS) to appear in the conditional variance. They are computed using every 15th trade. T-statistics, based on robust standard errors, are reported in small font and in brackets. Code: GARCH_analysis.ox

The rest of this paper has the following structure. In Section 2 we will discuss the theory of realised semivariance, deriving a central limit theory for it under some mild assumptions. In Section 3 we will deepen the empirical work reported here, looking at a variety of stocks and also both trade and quote data. In Section 4 we will discuss various extensions and areas of possible future work.

2 Econometric theory

2.1 The model and background

We start this section by repeating some of the theoretical story from Section 1.

Consider a Brownian semimartingale $Y$ given as

$$Y_t = \int_0^t a_s \, ds + \int_0^t \sigma_s \, dW_s,$$

where $a$ is a locally bounded predictable drift process and $\sigma$ is a càdlàg volatility process. For such a process

$$[Y]_t = \int_0^t \sigma_s^2 \, ds,$$
and so \( d[Y]_t = \sigma^2_t dt \), which means that when there are no jumps the QV process tells us everything we can know about the ex-post variation of \( Y \).

When there are jumps this is no longer true, in particular let

\[
Y_t = \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t,
\]

where \( J \) is a pure jump process. Then

\[
[Y]_t = \int_0^t \sigma^2_s ds + \sum_{s \leq t} (\Delta J_s)^2,
\]

and \( d[Y]_t = \sigma^2_t dt + (\Delta Y_t)^2 \). Even when we employ devices like bipower variation (Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen and Shephard (2006))

\[
\{Y\}^{[1,1]}_t = \mu_1^{-2} \frac{1}{n} \lim_{n \to \infty} \sum_{j=2}^{t_j \leq t} |Y_{t_j} - Y_{t_{j-1}}| |Y_{t_{j-1}} - Y_{t_{j-2}}|,
\]

we are able to estimate \( \int_0^t \sigma^2_s ds \) robustly to jumps, but this still leaves us with estimates of \( \sum_{s \leq t} (\Delta J_s)^2 \). This tells us nothing about the asymmetric behaviour of the jumps.

### 2.2 Realised semivariance

The empirical analysis we carry out throughout this paper is based in trading time, so data arrives into our database at irregular points in time. However, these irregularly spaced observations can be thought of as being equally spaced observations on a new time-changed process, in the same stochastic class, as argued by, for example, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). Thus there is no intellectual loss in initially considering equally spaced returns

\[
y_i = \frac{Y_n - Y_{i-1}}{n}, \quad i = 1, 2, ..., n.
\]

We study the functional

\[
V(Y, n) = \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{\lfloor nt \rfloor} \left( \begin{array}{c} y_i^2 1\{y_i \geq 0\} \\ y_i^2 1\{y_i < 0\} \end{array} \right).
\]

The main results then come from an application of some limit theory of Kinnebrock and Podolskij (2007) for bipower variation. This work can be seen as an important generalisation of Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006) who studied bipower type statistics of the form

\[
\frac{1}{n} \sum_{j=2}^{n} g(\sqrt{n}y_j) h(\sqrt{n}y_{j-1}),
\]

when \( g \) and \( h \) were assumed to be even functions. Kinnebrock and Podolskij (2007) give the extension to the uneven case, which is essential here.\(^6\)

\(^6\)It is also useful in developing the theory for realised autocovariance under a Brownian motion, which is important in the theory of realised kernels developed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).
Proposition 1 Suppose (1) holds, then
\[ V(Y, n) \overset{p}{\rightarrow} \frac{1}{2} \int_0^t \sigma_s^2 ds \left( \begin{array}{c} 1 \\ 1 \end{array} \right). \]

Proof. Trivial application of Theorem 1 in Kinnebrock and Podolskij (2007).

Corollary 1 Suppose
\[ Y_t = \int_0^t a_s ds + \int_0^t \sigma_s dW_s + J_t, \]
holds, where \( J \) is a finite activity jump process then
\[ V(Y, n) \overset{p}{\rightarrow} \frac{1}{2} \int_0^t \sigma_s^2 ds \left( \begin{array}{c} 1 \\ 1 \end{array} \right) + \sum_{s \leq t} \left( \frac{(\Delta Y_s)^2 1_{\{\Delta Y_s \geq 0\}}}{(\Delta Y_s)^2 1_{\{\Delta Y_s \leq 0\}}} \right). \]

Remark. The above means that
\[ (1, -1) V(Y, n) \overset{p}{\rightarrow} \sum_{s \leq t} (\Delta Y_s)^2 1_{\{\Delta Y_s \geq 0\}} - (\Delta Y_s)^2 1_{\{\Delta Y_s \leq 0\}}, \]
the difference in the squared jumps. Hence this statistic allows us direct econometric evidence on the importance of the sign of jumps. Of course, by combining with bipower variation
\[ V(Y, n) - \frac{1}{2} [Y]_t^{[1,1]} \overset{p}{\rightarrow} \sum_{s \leq t} \left( \frac{(\Delta Y_s)^2 1_{\{\Delta Y_s \geq 0\}}}{(\Delta Y_s)^2 1_{\{\Delta Y_s \leq 0\}}} \right), \]
we can straightforwardly estimate the QV of just positive or negative jumps.

In order to derive a central limit theory we need to make two assumptions on the volatility process.

(H1). If there were no jumps in the volatility then it would be sufficient to employ
\[ \sigma_t = \sigma_0 + \int_0^t a^*_u du + \int_0^t \sigma^*_u dW_u + \int_0^t v^*_u dV_u. \quad (4) \]
Here \( a^*, \sigma^*, v^* \) are adapted càdlàg processes, with \( a^* \) also being predictable and locally bounded. \( V \) is a Brownian motion independent of \( W \).

(H2). \( \sigma_t^2 > 0 \) everywhere.

The assumption (H1) is rather general from an econometric viewpoint as it allows for flexible leverage effects, multifactor volatility effects, jumps, non-stationarities, intraday effects, etc. Indeed we do not know of a continuous time continuous sample path volatility model used in financial economics which is outside this class. Kinnebrock and Podolskij (2007) also allow jumps in the volatility under the usual (in this context) conditions introduced by Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006) and discussed by, for example, Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006) but we will not detail this here.

The assumption (H2) is also important, it rules out the situation where the diffusive component disappears.

9
Proposition 2 Suppose (1), (H1) and (H2) hold, then
\[ \sqrt{n} \left\{ \frac{1}{2} \int_0^t \sigma_s^2 ds \left( \begin{array}{c} 1 \\ 1 \end{array} \right) \right\} \Rightarrow V_t \]
where
\[ V_t = \int_0^t \alpha_s (1) ds + \int_0^t \alpha_s (2) dW_s + \int_0^t \alpha_s (3) dW'_s, \]
\[ \alpha_s (1) = \frac{1}{\sqrt{2\pi}} \alpha_s \sigma_s \left( \begin{array}{c} 1 \\ -1 \end{array} \right), \]
\[ \alpha_s (2) = \frac{1}{2\sqrt{2\pi}} \sigma_s^2 \left( \begin{array}{c} 1 \\ -1 \end{array} \right), \]
\[ A_s = \frac{1}{4} \sigma_s^4 \left( \begin{array}{cc} 5 & -1 \\ -1 & 5 \end{array} \right), \]
\[ \alpha_s (3) \alpha_s (3)' = A_u - \alpha_s (2) \alpha_s (2)', \]
where \( \alpha_s (3) \) is a \( 2 \times 2 \) matrix. Here \( W' \perp \perp (W, V) \), the Brownian motions which appears in the Brownian semimartingale (1) and (H1).

Proof. Consider the framework of Theorem 2 in Kinnebrock and Podolskij (2007) and choose
\[ g (x) = \left( \begin{array}{c} g_1 (x) \\ g_2 (x) \end{array} \right) = \left( \begin{array}{c} x^2 1_{\{x \geq 0\}} \\ x^2 1_{\{x \leq 0\}} \end{array} \right), \quad h (x) = 1, \]
and use the notation that
\[ \rho_{\sigma} (g) = E \{ g(\sigma U) \}, \quad U \sim N(0, 1) \]
\[ \rho_{\sigma}^{(1)} (g) = E \{ U g(\sigma U) \}, \quad U \sim N(0, 1) \]
\[ \tilde{\rho}_{\sigma}^{(1,1)} (g) = E \left\{ g(\sigma W_1) \int_0^1 W_s dW_s \right\}, \]
where \( W \) is standard Brownian motion. We note that \( g \) is continuously differentiable and so their theory applies directly.

Then we obtain the result with
\[ \rho_{\sigma} (h) = 1, \quad \rho_{\sigma} (g_1) = \rho_{\sigma} (g_2) = \frac{1}{2} \sigma_s^2, \]
\[ \rho_{\sigma}^{(1)} (h) = 0, \quad \rho_{\sigma}^{(1)} (g_1) = \frac{1}{2\sqrt{2\pi}} \sigma_s^2 = -\rho_{\sigma}^{(1)} (g_2), \]
\[ \rho_{\sigma} \left( \frac{\partial h}{\partial x} \right) = 0, \quad \rho_{\sigma} \left( \frac{\partial g_1}{\partial x} \right) = \frac{1}{\sqrt{2\pi}} \sigma_s^2 = -\rho_{\sigma} \left( \frac{\partial g_2}{\partial x} \right), \]
\[ \rho_{\sigma}^{(1)} \left( \frac{\partial h}{\partial x} \right) = 0, \quad \rho_{\sigma}^{(1)} \left( \frac{\partial g_1}{\partial x} \right) = \rho_{\sigma}^{(1)} \left( \frac{\partial g_2}{\partial x} \right) = \sigma_s, \]
\[ \tilde{\rho}_{\sigma}^{11} \left( \frac{\partial g_1}{\partial x} \right) = \tilde{\rho}_{\sigma}^{11} \left( \frac{\partial g_2}{\partial x} \right) = \frac{1}{2\sqrt{2\pi}}, \]
\[ \tilde{\rho}_{\sigma}^{11} (h) = 0. \]
This completes the proof.

**Remark.** When we look at

$$RV = (1, 1) V(Y, n),$$

then we produce the well known result

$$\sqrt{n} \left\{ RV - \int_0^t \sigma_s^2 ds \right\} \overset{D_{st}}{\rightarrow} \int_0^t 2\sigma_s^2 dW'_s$$

which appears in Jacod (1994) and Barndorff-Nielsen and Shephard (2002).

**Remark.** Write

$$V(Y, n) = \sqrt{n} \left\{ V(Y, n) - \frac{1}{2} \int_0^t \sigma_s^2 ds \left( \begin{array}{c} 1 \\ 1 \end{array} \right) \right\},$$

then if \(a, \sigma \perp \perp W\) then

$$V(Y, n) \overset{D_{st}}{\rightarrow} MN \left( \frac{1}{\sqrt{2\pi}} \int_0^t a_s \sigma_s ds \left( \begin{array}{c} 1 \\ -1 \end{array} \right), \frac{1}{4} \int_0^t \sigma_s^4 ds \left( \begin{array}{cc} 5 & -1 \\ -1 & 5 \end{array} \right) \right).$$

If there is no drift then the mean of this mixed Gaussian distribution is zero and we could use this limit result to construct confidence intervals on these quantities. When the drift is not zero we cannot use this result as we do not have a method for estimating the bias which is a scaled version of

$$\frac{1}{\sqrt{n}} \int_0^t a_s \sigma_s ds.$$

Of course in practice this bias will be tiny. The asymptotic variance of \((1, -1) V(Y, n)\) is \(\frac{3}{n} \int_0^t \sigma_s^4 ds\), but obviously not mixed Gaussian.

**Remark.** When the \(a, \sigma \perp \perp W\) result fails, we do not know how to construct confidence intervals even if the drift is zero. This is because \(V\) depends upon \(W\). All we know is that the asymptotic variance is again

$$\frac{1}{4n} \int_0^t \sigma_s^4 ds \left( \begin{array}{cc} 5 & -1 \\ -1 & 5 \end{array} \right).$$

Notice, throughout the asymptotic variance of \(RS\) is

$$\frac{5}{4n} \int_0^t \sigma_s^4 ds$$

so is less than that of the RV.
More empirical work

3.1 More on GE trade data

For the GE trade data, Table 2 reports basic summary statistics for squared open to close daily returns, realised variance and realised semivariance. Much of this is familiar, with the average level of squared returns and realised variance being roughly the same, while the mean of the realised semivariance is around one half that of the realised variance. The most interesting results are that the RS statistic has a correlation with RV of around 0.86 and that it is negatively correlated with daily returns. The former correlation is modest for an additional volatility measure and indicates it may have additional information not in the RV statistic. The latter result shows that large daily semivariances are associated with contemporaneous downward moves in the asset price — which is not surprising of course.

The serial correlation in the daily statistics are also presented in Table 2. They show the RV statistic has some predictability through time, but that the autocorrelation in the realised semivariance is much higher. Together with the negative correlation between returns and contemporaneous RS (which is consistent for a number of different assets), this suggests one should be able to modestly predict returns using past RS.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Summary information for daily statistics for GE trade data</th>
<th>Correlation matrix</th>
<th>ACF1</th>
<th>ACF20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_i$</td>
<td>Mean 0.01, S.D. 1.53</td>
<td>1.00</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$r_i^2$</td>
<td>Mean 2.34, S.D. 5.42</td>
<td>0.06 1.00</td>
<td>0.17</td>
<td>0.07</td>
</tr>
<tr>
<td>$RV_i$</td>
<td>Mean 2.61, S.D. 3.05</td>
<td>0.03 0.61 1.00</td>
<td>0.52</td>
<td>0.26</td>
</tr>
<tr>
<td>upRV$_i$</td>
<td>Mean 1.33, S.D. 2.03</td>
<td>0.20 0.61 0.94 1.00</td>
<td>0.31</td>
<td>0.15</td>
</tr>
<tr>
<td>RS$_i$</td>
<td>Mean 1.28, S.D. 1.28</td>
<td>-0.22 0.47 0.86 0.66 1.00</td>
<td>0.65</td>
<td>0.37</td>
</tr>
<tr>
<td>BPV$_i$</td>
<td>Mean 2.24, S.D. 2.40</td>
<td>0.00 0.54 0.95 0.84 0.93 1.00</td>
<td>0.64</td>
<td>0.34</td>
</tr>
<tr>
<td>BPDV$_i$</td>
<td>Mean 0.16, S.D. 0.46</td>
<td>-0.61 -0.10 -0.08 -0.34 0.34 -0.01 1.00</td>
<td>0.06</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for daily GE data computed using trade data. $r_i$ denotes daily open to close returns, $RV_i$ is the realised variance, $RS_i$ is the realised semivariance, $upRV = RV - RS$ and $BPV_i$ is the daily realised bipower variation.

Table 3 shows the regression fit of $r_i$ on $r_{i-1}$ and $RS_{i-1}$ for the GE trade data. The t-statistic on lagged RS is just significant and positive. Hence a small amount of the variation in the high frequency falls of price in the previous day are associated with rises in future asset prices — presumably because the high frequency falls increase the risk premium. The corresponding t-statistics for the impact of $RS_{i-1}$ for other series are given in Table 6, they show a similar weak pattern.

The RS statistic has a similar dynamic pattern to the bipower variation statistic\(^7\). The mean

\(^7\)This is computed using not one but two lags, which reduces the impact of market microstructure, as shown by
GE trade data: Regression of returns on lagged realised semivariance and returns

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.009</td>
<td>0.03</td>
<td>-0.061</td>
<td>-1.43</td>
<td>-0.067</td>
<td>-1.56</td>
</tr>
<tr>
<td>( r_{i-1} )</td>
<td>-0.012</td>
<td>0.01</td>
<td>-0.001</td>
<td>-0.06</td>
<td>0.016</td>
<td>0.67</td>
</tr>
<tr>
<td>( RS_{i-1} )</td>
<td></td>
<td></td>
<td>0.054</td>
<td>2.28</td>
<td>0.046</td>
<td>1.85</td>
</tr>
<tr>
<td>( RS_{i-1} - 0.5BPV_{i-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.109</td>
<td>1.26</td>
</tr>
<tr>
<td>( \log L )</td>
<td>-4802.2</td>
<td></td>
<td>-4799.6</td>
<td></td>
<td>-4798.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Regression of returns \( r_{i} \) on lagged realised semivariance \( RS_{i-1} \) and returns \( r_{i-1} \) for daily returns based on the GE trade database.

and standard deviation of the RS statistic is slightly higher than half the realised BPV one. The difference estimator

\[
BPDV_i = RS_i - 0.5BPV_i,
\]

which estimates the squared negative jumps, is highly negatively correlated with returns but not very correlated with other measures of volatility. Interestingly this estimator is slightly autocorrelated, but at each of the first 10 lags this correlation is positive which means it has some forecasting potential.

### 3.2 Other trade data

Results in Table 3 show that broadly the same results hold for a number of frequently traded assets - American Express (AXP), Walt Disney (DIS) and IBM. Table 5 shows the log-likelihood improvements by including RV and RS statistics into the GARCH and GJR models based on trades. The conclusion is clear for GARCH models. By including RS statistics in the model there is little need to include a traditional leverage effect. Typically it is only necessary to include RS in the information set, adding RV plays only a modest role. For GJR models, the RV statistic becomes more important and is sometimes slightly more effective than the RS statistic.

### 3.3 Quote data

We have carried out the same analysis based on quote data, looking solely at the series for offers to buy placed on the New York Stock Exchange. The results are given in Tables 6 and 7. The results are in line with the previous trade data. The RS statistic is somewhat less effective for quote data, but the changes are marginal.

---

Andersen, Bollerslev, and Diebold (2007).
Summary information for daily statistics for other trade data

<table>
<thead>
<tr>
<th>DIS</th>
<th>Mean</th>
<th>S.D.</th>
<th>r_i</th>
<th>1.00</th>
<th>ACF_1</th>
<th>ACF_20</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_i</td>
<td>-0.02</td>
<td>1.74</td>
<td>1.00</td>
<td>-0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>r_i^2</td>
<td>3.03</td>
<td>6.52</td>
<td>0.04</td>
<td>1.00</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>RV_i</td>
<td>3.98</td>
<td>4.69</td>
<td>-0.00</td>
<td>0.53</td>
<td>1.00</td>
<td>0.69</td>
</tr>
<tr>
<td>upRV_i</td>
<td>1.97</td>
<td>2.32</td>
<td>0.19</td>
<td>0.55</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>RS_i</td>
<td>2.01</td>
<td>2.60</td>
<td>-0.18</td>
<td>0.46</td>
<td>0.95</td>
<td>0.81</td>
</tr>
<tr>
<td>BPV_i</td>
<td>3.33</td>
<td>3.97</td>
<td>-0.00</td>
<td>0.53</td>
<td>0.98</td>
<td>0.93</td>
</tr>
<tr>
<td>BPDV_i</td>
<td>0.35</td>
<td>1.03</td>
<td>-0.46</td>
<td>0.13</td>
<td>0.52</td>
<td>0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AXP</th>
<th>Mean</th>
<th>S.D.</th>
<th>r_i</th>
<th>1.00</th>
<th>ACF_1</th>
<th>ACF_20</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_i</td>
<td>0.01</td>
<td>1.86</td>
<td>1.00</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>r_i^2</td>
<td>3.47</td>
<td>7.75</td>
<td>-0.00</td>
<td>1.00</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>RV_i</td>
<td>3.65</td>
<td>4.57</td>
<td>-0.01</td>
<td>0.56</td>
<td>1.00</td>
<td>0.64</td>
</tr>
<tr>
<td>upRV_i</td>
<td>1.83</td>
<td>2.62</td>
<td>0.22</td>
<td>0.52</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>RS_i</td>
<td>1.82</td>
<td>2.30</td>
<td>-0.28</td>
<td>0.53</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>BPV_i</td>
<td>3.09</td>
<td>3.74</td>
<td>-0.04</td>
<td>0.52</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td>BPDV_i</td>
<td>0.27</td>
<td>0.90</td>
<td>-0.63</td>
<td>0.27</td>
<td>0.37</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBM</th>
<th>Mean</th>
<th>S.D.</th>
<th>r_i</th>
<th>1.00</th>
<th>ACF_1</th>
<th>ACF_20</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_i</td>
<td>0.01</td>
<td>1.73</td>
<td>1.00</td>
<td>-0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>r_i^2</td>
<td>3.02</td>
<td>7.25</td>
<td>0.04</td>
<td>1.00</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>RV_i</td>
<td>2.94</td>
<td>3.03</td>
<td>0.03</td>
<td>0.55</td>
<td>1.00</td>
<td>0.65</td>
</tr>
<tr>
<td>upRV_i</td>
<td>1.50</td>
<td>1.81</td>
<td>0.24</td>
<td>0.54</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>RS_i</td>
<td>1.44</td>
<td>1.43</td>
<td>-0.24</td>
<td>0.48</td>
<td>0.91</td>
<td>0.74</td>
</tr>
<tr>
<td>BPV_i</td>
<td>2.62</td>
<td>2.60</td>
<td>0.00</td>
<td>0.51</td>
<td>0.96</td>
<td>0.86</td>
</tr>
<tr>
<td>BPDV_i</td>
<td>0.13</td>
<td>0.49</td>
<td>-0.71</td>
<td>0.05</td>
<td>0.13</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Table 4: Summary statistics for various daily data computed using trade data. \( r_i \) denotes daily open to close returns, \( RV_i \) is the realised variance, \( RS_i \) is the realised semivariance, \( upRV = RV – RS \) and \( BPV_i \) is the daily realised bipower variation. \( BPDV_i \) is the realised bipower downward variation statistic.

**Trades: logL improvements by including lagged RS and RV in conditional variance**

<table>
<thead>
<tr>
<th>Lagged variables</th>
<th>GARCH model</th>
<th>GJR model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AXP</td>
<td>DIS</td>
</tr>
<tr>
<td>RV, RS &amp; BPV</td>
<td>59.9</td>
<td>66.5</td>
</tr>
<tr>
<td>RV &amp; BPV</td>
<td>53.2</td>
<td>63.7</td>
</tr>
<tr>
<td>RS &amp; BPV</td>
<td>59.9</td>
<td>65.7</td>
</tr>
<tr>
<td>BPV</td>
<td>46.2</td>
<td>57.5</td>
</tr>
<tr>
<td>RV &amp; RS</td>
<td>59.8</td>
<td>66.3</td>
</tr>
<tr>
<td>RV</td>
<td>53.0</td>
<td>63.5</td>
</tr>
<tr>
<td>RS</td>
<td>59.6</td>
<td>65.6</td>
</tr>
<tr>
<td>None</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5: Improvements in the Gaussian quasi-likelihood by including lagged realised quantities in the conditional variance over standard GARCH and GJR models. Fit of GARCH and GJR models for daily open to close returns on four share prices, from 1995 to 2005. We allow lagged daily realised variance (RV), realised semivariance (RS) realised bipower variation (BPV) to appear in the conditional variance. They are computed using every 15th trade. Code: [GARCHanalysis.ox](https://example.com/GARCHanalysis.ox)
Table 6: The t-statistics on realised semivariance calculated by regressing daily returns $r_i$ on lagged daily returns and lagged daily semivariances ($RS_{i-1}$). This is carried out for a variety of stock prices using trade and quote data. The RS statistics are computed using every 15th high frequency data point.

Quotes: LogL improvements by including lagged RS and RV in conditional variance

Table 7: Quote data: Improvements in the Gaussian quasi-likelihood by including lagged realised quantities in the conditional variance. Fit of GARCH and GJR models for daily open to close returns on four share prices, from 1995 to 2005. We allow lagged daily realised variance (RV) and realised semivariance (RS) to appear in the conditional variance. They are computed using every 15th trade. Code: GARCH\_analysis.ox

4 Additional remarks

4.1 Bipower variation

We can build on the work of Barndoff-Nielsen and Shephard (2004), Barndorf-Nielsen and Shephard (2006), Andersen, Bollerslev, and Diebold (2007) and Huang and Tauchen (2005) by defining

$$BPDV = \sum_{j=1}^{t_i \leq 1} (Y_{t_j} - Y_{t_{j-1}})^2 \mathbb{1}_{Y_{t_j} - Y_{t_{j-1}} \leq 0} - \frac{1}{2} \nu_1^{-2} \sum_{j=2}^{t_i \leq 1} |Y_{t_j} - Y_{t_{j-1}}| |Y_{t_{j-1}} - Y_{t_{j-2}}|$$

$$\sum_{s \leq t} (\Delta Y_s)^2 I_{\Delta Y_s \leq 0},$$

the realised bipower downward variation statistic (upward versions are likewise trivial to define). This seems a novel way of thinking about jumps — we do not know of any literature which has identified $\sum_{s \leq t} (\Delta Y_s)^2 I_{\Delta Y_s}$ before. It is tempting to try to carry out jump tests based upon it to test for the presence of downward jumps against a null of no jumps at all. However, the theory developed in Section 2 suggests that this is going to be hard to carry out solely based on in-fill asymptotics without stronger assumptions than we usually like to make due to the presence of the drift term in the limiting result and the non-mixed Gaussian limit theory (we could do testing if we assumed the drift was zero and there is no leverage term). Of course, it would not stop us from testing things based on the time series dynamics of the process - see the work of Corradi and
Further, a time series of such objects can be used to assess the factors which drive downward jumps, but simply building a time series model for it, conditioning on explanatory variables.

An alternative to this approach is to use higher order power variation statistics (e.g. Barndorff-Nielsen and Shephard (2004)),

\[ \sum_{j=1}^{t_j \leq 1} |Y_{t_j} - Y_{t_j-1}| r 1_{Y_{t_j} - Y_{t_j-1} \leq 0} \overset{p}{\longrightarrow} \sum_{s \leq t} |\Delta Y_s|^r I_{\Delta Y_s \leq 0}, \quad r > 2, \]

as \( n \to \infty \). The difficulty with using these high order statistics is that they will be more sensitive to noise than the BPDV estimator.

### 4.2 Effect of noise

Suppose instead of seeing \( Y \) we see \( X = Y + U \), and think of \( U \) as noise. Let us focus entirely on

\[
\sum_{i=1}^{n} x_i^2 1\{x_i \leq 0\} = \sum_{i=1}^{n} y_i^2 1\{y_i \leq -u_i\} + \sum_{i=1}^{n} u_i^2 1\{y_i \leq u_i\} + 2 \sum_{i=1}^{n} y_i u_i 1\{y_i \leq u_i\} \approx \sum_{i=1}^{n} y_i^2 1\{u_i \leq 0\} + \sum_{i=1}^{n} u_i^2 1\{u_i \leq 0\} + 2 \sum_{i=1}^{n} y_i u_i 1\{u_i \leq 0\}.
\]

If we use the framework of Zhou (1996), where \( U \) is white noise, uncorrelated with \( Y \), with \( \mathbb{E}(U) = 0 \) and \( \text{Var}(U) = \omega^2 \) then it is immediately apparent that the noise will totally dominate this statistic in the limit as \( n \to \infty \).

Pre-averaging based statistics of Jacod, Li, Mykland, Podolskij, and Vetter (2007) could be used here to reduce the impact of noise on the statistic.

### 5 Conclusions

This paper has introduced a new measure of variation called “realised semivariance.” It is determined solely by high frequency downward moves in asset prices. We have seen it is possible to carry out an asymptotic analysis of this statistic and see that its limit only contains downward jumps.

We have assessed the effectiveness of this new measure using it as a conditioning variable for a GARCH model of daily open to close returns. Throughout, for non-leverage based GARCH models, realised semivariance is more informative than the usual realised variance statistic. When a leverage term is introduce it is hard to tell the difference.

Various extensions to this work were suggested.
6 Acknowledgements

This paper was prepared for a conference in honour of Robert F. Engle to be held in San Diego in June 2008. We are extremely grateful to Tim Bollerslev, Jeffrey Russell and Mark Watson for the invitation to give it.

The ARCH models fitted in this paper were computed using G@ARCH 5.0, the package of Laurent and Peters (2002). Throughout programming was carried out using the Ox language of Doornik (2001) within the OxMetrics 5.0 environment.

We are very grateful for the help of Asger Lunde in preparing some of the data we used in this analysis and advice on various issues. We also would like to thank Anthony Ledford for helpful suggestions at various points.

References


