# Robust variation estimation using kernels in financial econometrics

NEIL SHEPHARD
*Nuffield College, University of Oxford*
(joint with Ole E. Barndorff-Nielsen,
Peter Hansen and Asger Lunde)
`www.nuffield.ox.ac.uk/users/shephard/`

# 1 Introduction

- Realised quantities & irregularly spaced data
- Realised autocorrelation
- Market frictions and kernels

# 2 Brownian semimartingale

Work with univariate $(\mathcal{BSM})$

$$Y_t = \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_{u-} \mathrm{d}W_u$$

$a$ is predictable, $\sigma$ is càdlàg volatility process and $W$ is BM.

Quadratic variation (QV) process of $Y$ is (EX-POST)

$$[Y]_t = \mathrm{p-}\lim_{n\to\infty} \sum_{j=1}^{t_j \le t} \left(Y_{t_j} - Y_{t_{j-1}}\right)^2, \tag{1}$$

for $0 = t_0 < t_1 < ... < t_n = T$ with $\sup_j \{t_{j+1} - t_j\} \to 0$ for $n \to \infty$.

In $Y \in \mathcal{BSM}$ then

$$[Y]_t = \int_0^t \sigma_u^2 \mathrm{d}u.$$

The realised QV estimator

$$[Y_\delta]_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} \left( Y_{j\delta} - Y_{(j-1)\delta} \right)^2,$$

Define the daily QV

$$V_i = [Y]_i - [Y]_{(i-1)}, \quad i = 1, 2, ...$$

estimated by the realised daily QV

$$\widehat{V}_i = [Y_\delta]_i - [Y_\delta]_{(i-1)}, \quad i = 1, 2, ....$$

$\widehat{V}_i$ is called the realised variance, square root is its realised volatility in finance.

Realised volatility has a long history. It appears in Rosenberg (1972), Merton (1980), Schwert (1989) and Schwert (1998). Of course, in probability theory QV was discussed as early as Wiener (1924). Closer connection between realised QV and QV, and its use for econometric purposes, was made in Comte and Renault (1998), Barndorff-Nielsen and Shephard (2001) and Andersen, Bollerslev, Diebold, and Labys (2001).

Substantial literature on writing derivatives on realised volatility. Neuberger (1990), Carr and Madan (1998), Demeterfi, Derman, Kamal, and Zou (1999), Carr and Lewis (2004). Brockhaus and Long (1999), Javaheri, Wilmott, and Haug (2002), Howison, Rafailidis, and Rasmussen (2004), Carr, Geman, Madan, and Yor (2005), Carr and Lee (2003). See also the overview of Branger and Schlag (2005).

CLT for $[Y_\delta]_t$ was developed in a series of papers by Jacod (1994), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002).

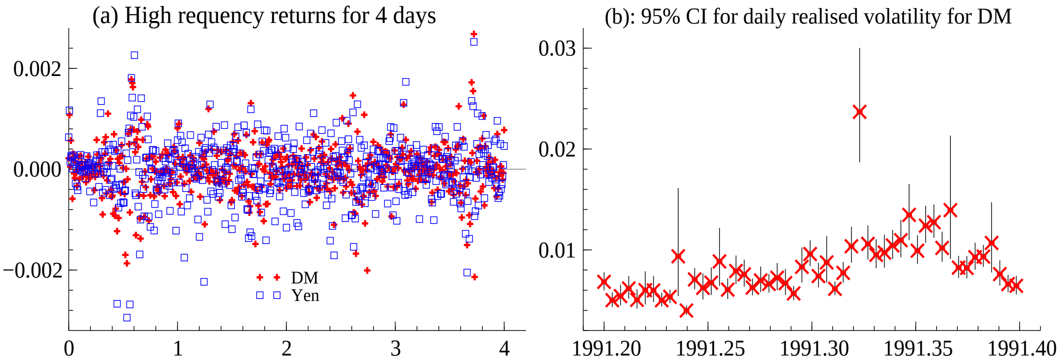
$$\delta^{-1/2}\left([Y_\delta]_t - [Y]_t\right) \xrightarrow{L} MN\left(0, 2\int_0^t \sigma_u^4 \mathrm{d}u\right), \tag{2}$$

where $MN$ denotes a mixed Gaussian distribution. Barndorff-Nielsen and Shephard (2002) named $\int_0^t \sigma_u^4 \mathrm{d}u$ *integrated quarticity.* Can be consistently estimated using $(1/3)\left\{Y_\delta\right\}_t^{[4]}$ where realised quarticity

$$[Y_\delta]_t^{[4]} = \delta^{-1} \sum_{j=1}^{\lfloor t/\delta \rfloor} \left(Y_{j\delta} - Y_{(j-1)\delta}\right)^4. \tag{3}$$

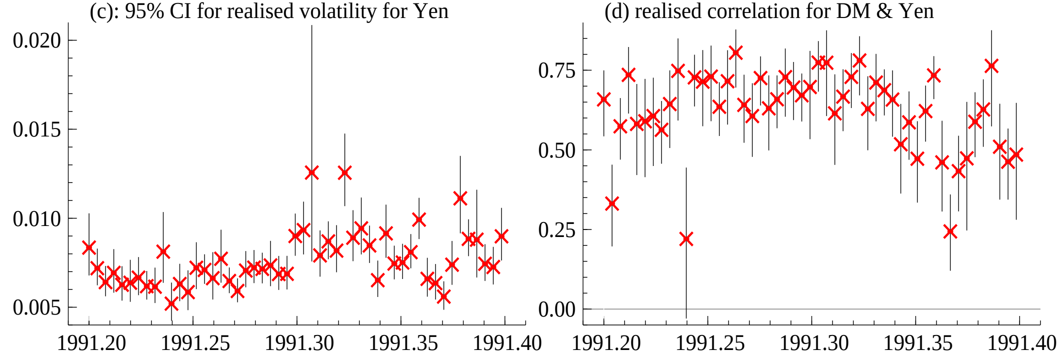
$$\frac{\delta^{-1/2}\left(\log[Y_\delta]_t - \log[Y]_t\right)}{\sqrt{\frac{2}{3}\frac{[Y_\delta]_t^{[4]}}{\left([Y_\delta]_t\right)^2}}} \xrightarrow{L} N(0,1). \tag{4}$$

# 10 minute returns, DM/US and Yen/Dollar



(a) High requency returns for 4 days

(b): 95% CI for daily realised volatility for DM



(c): 95% CI for realised volatility for Yen

(d) realised correlation for DM & Yen

$$Y_t = \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_{u-} \mathrm{d}W_u + \sum_{j=1}^{N_t} C_j$$

$$[Y]_t = \int_0^t \sigma_u^2 \mathrm{d}u + \sum_{j=1}^{N_t} C_j^2.$$

Realised bipower

$$\left\{Y_\delta^l\right\}_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} \left| Y_{\delta(j-1)}^l - Y_{\delta(j-2)}^l \right| \left| Y_{\delta j}^l - Y_{\delta(j-1)}^l \right|,$$

in general is robust to jumps

$$\mu_1^{-2} \{Y_\delta\}_t \xrightarrow{p} \int_0^t \sigma_u^2 \mathrm{d}u.$$

Barndorff-Nielsen, Graversen, Jacod, and Shephard (2005) under $Y \in \mathcal{BSM}$

$$\delta^{-1/2}\left(\mu_1^{-2}\{Y_\delta\}_t - \int_0^t \sigma_u^2 \mathrm{d}u\right) \xrightarrow{L} MN\left(0, (2+\vartheta)\int_0^t \sigma_u^4 \mathrm{d}u\right).$$

Also work of Mancini (2004) in this context — uses a shrinking threshold on squared returns. More efficient — but how to choose the threshold? Aït-Sahalia and Jacod (2005) have additional insights on that.

What to do when we have irregularly spaced data?

Spacing is endogenous in econometrics!

Mykland and Zhang (2005) on realised QV, but assume independence of spacing and Y. See HIGHLY likely that their theory will go in more general context, but the proof of this is not easy. What about other multipower variations, realised range?

We following a straightforward approach, which must be well known! We do not claim originality, but we like the results!

# 3 Irregularly spaced data

Think of irregularly spaced data as regularly spaced data on a new process, for $\delta > 0$,

$$X_{\delta j} = Y_{T_{\delta j}} = (Y \circ T)_{\delta j} \, ,$$

where $T$ is a time-change (non-decreasingly process, $T_0 = 0$). Need to be clearly about the information sets here

$$(Y \circ T, T)$$

adapted to $\mathcal{F}$. Assume that $T$ is absolutely continuous

$$T_t = \int_0^t \tau_u^2 \mathrm{d}u,$$

where $\tau$ is càdlàg.

What process does

$$X = (Y \circ T),$$

follow? Can show $X \in \mathcal{BSM}$. The important point is that the spot volatility of $X$ is

$$\sigma_{T_t} \tau_t.$$

In particular

$$[X]_t = ([Y] \circ T)_t = \int_0^{T_t} \sigma_u^2 \mathrm{d}u.$$

Example: Define the realised QV

$$\begin{aligned}[X_\delta]_t &= \sum_{j=1}^{\lfloor t/\delta \rfloor} \left(X_{\delta j} - X_{\delta(j-1)}\right)^2 = \sum_{j=1}^{\lfloor t/\delta \rfloor} \Big((Y \circ T)_{\delta j} - (Y \circ T)_{\delta(j-1)}\Big)^2 \\ &= [Y_T]_t.\end{aligned}$$

Think of $T_1 = 1$, then the Jacod (1994), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002) CLT can be used to produce

$$\delta^{-1/2}\left([Y_T]_1 - [Y]_1\right) \xrightarrow{L} MN\left(0, 2\int_0^1 \sigma_{T_u}^4 \tau_u^4 \mathrm{d}u\right),$$

while (see also Barndorff-Nielsen and Shephard (2005))

$$\frac{[Y_T]_1 - [Y]_1}{\sqrt{\frac{2}{3}\sum_{j=1}^{\lfloor 1/\delta \rfloor}\Big((Y \circ T)_{\delta j} - (Y \circ T)_{\delta(j-1)}\Big)^4}} \xrightarrow{L} N(0,1).$$

Example: Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen, Graversen, Jacod, and Shephard (2005) CLT can be used to produce

$$\delta^{-1/2}\left(\mu_1^{-2}\left\{Y_T\right\}_t^{[1,1]}-[Y]_1\right)\xrightarrow{L} MN\left(0,(2+\vartheta)\int_0^1\sigma_{T_u}^4\tau_u^4\mathrm{d}u\right),$$

where $\vartheta \simeq 0.6$. Can be used to test for jumps using irregularly spaced data.

## 3.1 Realised autocovariance

Suppose $Y \in \mathcal{BSM}$ and define, for $h = 0, 1, 2, ..., H$.

$$\widetilde{\gamma}_h(Y_\delta)_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} \left(Y_{\delta j} - Y_{\delta(j-1)}\right)\left(Y_{\delta(j-h)} - Y_{\delta(j-h-1)}\right).$$

Such processes are traded! As $\delta \downarrow 0$ so

$$\delta^{-1/2}\begin{pmatrix} [Y_\delta]_t - \int_0^t \sigma_u^2 \mathrm{d}u \\ \widetilde{\gamma}_1(Y_\delta) \\ \vdots \\ \widetilde{\gamma}_H(Y_\delta) \end{pmatrix} \xrightarrow{L} MN\left(0, \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \int_0^t \sigma_u^4 \mathrm{d}u \right). \tag{5}$$

A consequence of the theorem is that as $\delta \downarrow 0$ the sample autocorrelations have

$$\delta^{-1/2}\begin{pmatrix} \widetilde{\gamma}_1(Y_\delta)_t/[Y_\delta]_t \\ \vdots \\ \widetilde{\gamma}_H(Y_\delta)_t/[Y_\delta]_t \end{pmatrix} \xrightarrow{L} MN\left(0, I\frac{\int_0^t \sigma_u^4 \mathrm{d}u}{\left(\int_0^t \sigma_u^2 \mathrm{d}u\right)^2}\right), \qquad (6)$$

which differs from the result of Bartlett (1946), inflating the usual standard errors as well as making inference multivariate mixed Gaussian.

## 3.2 Kernel

Define

$$\widetilde{\gamma}(Y_\delta) = \left([Y_\delta]_t, 2\widetilde{\gamma}_1(Y_\delta), 2\widetilde{\gamma}_2(Y_\delta), ..., 2\widetilde{\gamma}_H(Y_\delta)\right)'.$$

then the kernel is

$$\widetilde{K}_w(Y_\delta)_t = w'\widetilde{\gamma}(Y_\delta), \tag{7}$$

where the weights $w = (w_0, w_1, ..., w_H)'$ are non-stochastic.

Often used in statistics and econometrics as long-run variance estimators. Such approaches goes back to at least Maurice Bartlett.

Here the interest is in dealing with the effect of market frictions. Start by studying when there are no frictions!

It is clear that the CLT for a general kernel when $Y \in \mathcal{BSM}$ is

$$(K_H \delta)^{-1/2} \left( \widetilde{K}_w(Y_\delta)_t - w_0 \int_0^t \sigma_u^2 \mathrm{d}u \right) \xrightarrow{L} MN\left(0, \int_0^t \sigma_u^4 \mathrm{d}u \right), \quad (8)$$

where

$$K_H = 2\left( w_0^2 + 2 \sum_{h=1}^{H} w_h^2 \right).$$

When multiplied by integrated quarticity $K_H$ reveals the scaled asymptotic variance of $\widetilde{K}_w(Y_\delta)_t$.

Zhou (1996) suggested estimating $[Y]$ by using a kernel with $w_0 = w_1 = 1$ and $w_h = 0$ for $h > 1$. In that case

$$\delta^{-1/2}\left([Y_\delta]_t + 2\widetilde{\gamma}_1(Y_\delta)_t - \int_0^t \sigma_u^2 \mathrm{d}u\right) \xrightarrow{L} MN\left(0, 6\int_0^t \sigma_u^4 \mathrm{d}u\right).$$

Hence the inclusion of a single lag increases the asymptotic variance by a factor of three.

Another important example is where $w$ follows Bartlett (1950) weights. These were highlighted in econometrics by the work of Newey and West (1987). It sets

$$w_0 = 1 \quad \text{and} \quad w_h = (H + 1 - h) / (H + 1), \quad h = 1, 2, ....$$

When $H = 1$ then $w_1 = 1/2$ and $K_H = 3$, while when $H = 2$ then $w_1 = 2/3$, $w_2 = 1/3$ and $K_H = 4 + 2/9$. For moderately large $H$ then $K_H \simeq 4(H + 1)/3$. This means that $H$ cannot be sent off to infinity without regard to the rate at which $\delta \downarrow 0$. In particular we need that $K_H \delta = o(1)$, which means that

$$H\delta = o(1).$$

# 4 Market frictions/fictions

The semimartingale model is a fiction: i.e. there is no such thing as "a price." Transations data, best bid/ask. Indeed for many markets the whole order book is available.

The impact of frictions on realised QV is reviewed in Hansen and Lunde (2006). Observe $X$, want $Y$

$$X = Y + U.$$

In US markets impact of frictions has fallen dramatically in recent years due to decimalisation, etc. On a thickly traded stock the following is a good guide.

$$X = Y + U.$$

- If $\delta$ is 10-30 minute returns, above limit theory is a good guide to finite sample behaviour, i.e. $U$ ignorable.
- if $\delta$ is $(1, 10)$ minutes, then frictions will have impact, but $Y \perp\!\!\!\perp U$
- if $\delta < 1$ minute. Very complicated dynamics — market microstructure econometrics.

Research problem. Using $Y \perp\!\!\!\perp U$, exploit data up to 1 minute returns.

$U$ causes bias, quantified by a number of authors, see the survey by Hansen and Lunde (2006).

Change the estimator of $[Y]$ from $[X]$. e.g. subsampler introduced by Zhou (1996) and studied for $Y \in \mathcal{BSM}$ by Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2004) and Aït-Sahalia, Mykland, and Zhang (2005), also point process approach of Large (2005).

Here we study kernels.

Define, for $h = 0, 1, 2, ..., H$,

$$\widetilde{\gamma}_h(Y_\delta, U_\delta)_t = \sum_{j=1}^{n} \left(Y_{\delta j} - Y_{\delta(j-1)}\right)\left(U_{\delta(j-h)} - U_{\delta(j-h-1)}\right),$$

then

$$\widetilde{K}_w(X_\delta) = \widetilde{K}_w(Y_\delta) + \widetilde{K}_w(U_\delta) + 2\widetilde{K}_w(Y_\delta, U_\delta),$$

where

$$\widetilde{K}_w(Y_\delta, U_\delta) = w'\widetilde{\gamma}(Y_\delta, U_\delta)$$

and

$$\widetilde{\gamma}(Y_\delta, U_\delta) = \left(\begin{array}{c} \widetilde{\gamma}_0(Y_\delta, U_\delta), \widetilde{\gamma}_1(Y_\delta, U_\delta) + \widetilde{\gamma}_1(U_\delta, Y_\delta), ... \\ \widetilde{\gamma}_H(Y_\delta, U_\delta) + \widetilde{\gamma}_H(U_\delta, Y_\delta) \end{array}\right).$$

To study other terms we assume strong conditions to see the principles.

- $Y \perp\!\!\!\perp U$
- Zero mean, $U_{\delta j}$ are independent over $j$ (white noise in tick time).

$$\mathrm{Var}(U_{\delta j}) = \omega^2, \quad \mathrm{Var}(U_{\delta j}^2) = \lambda^2 \omega^4.$$

Implies

$$\mathrm{E}\left[\widetilde{K}_w(Y_\delta, U_\delta)\right] = 0$$

$$n = \delta^{-1}$$

Then

$$\mathrm{E}\left\{\widetilde{\gamma}(U_\delta)\right\} = 2\omega^2 n \left(1, -1, 0, ..., 0\right)',$$

so kernel is unbiased iff

$$w_0 = -w_1$$

while

$$\mathrm{Cov}\left\{\widetilde{\gamma}(U_\delta)\right\} = 4\omega^4 \left(nA + D\right). \tag{9}$$

Here the $(H+1) \times (H+1)$ matrices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$

where the $(H-1)\times(H-1)$ and $2\times(H-1)$ dimensional matrices

$$A_{22}=\begin{pmatrix} 6 & \bullet & \bullet & \bullet & \bullet \\ -4 & 6 & \bullet & \bullet & \bullet \\ 1 & -4 & 6 & \bullet & \bullet \\ 0 & 1 & -4 & 6 & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad A_{12}=\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -4 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

$A_{21}=A_{12}'$ and

$$D_{22}=\begin{pmatrix} -2 & \bullet & \bullet & \bullet & \bullet \\ 2 & -2 & \bullet & \bullet & \bullet \\ -1 & 2 & -2 & \bullet & \bullet \\ 0 & -1 & 2 & -2 & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad D_{12}=\begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 2 & -1 & 0 & \cdots & 0 \end{pmatrix},$$

where $D_{21} = D_{12}'$. The $2 \times 2$ matrices $A_{11}$ and $D_{11}$ are

$$A_{11} = \begin{pmatrix} 1+\lambda^2 & \bullet \\ -2-\lambda^2 & 5+\lambda^2 \end{pmatrix} \quad \text{and} \quad D_{11} = \begin{pmatrix} -\lambda^2/2 & \bullet \\ 1+\lambda^2/2 & -2 \end{pmatrix}.$$

Can choose $w$ to minimise this, it is not influenced by $\omega^2$.

Ignore $D$ then choose $w$ to minimise variance of kernel. Can drive variance to zero like

$$\simeq c\omega^4 n/H^3,$$

by selecting weights

$$\simeq 1 - 3\left(\frac{i}{H}\right)^2 + 2\left(\frac{i}{H}\right)^3,$$

which is like the Zhang (2004) subsampler. Optimal rate of convergence $n^{-1/4}$.

End conditions D have impact though.

Can show that the minimum variance of unbiased $\widetilde{K}_w(U_\delta)$ is $4\omega^4\lambda^2$, so kernel class is inconsistent unless $\lambda^2 = \mathrm{Var}(U_{\delta j}^2) = 0$.

Of course in practice $\omega^2$ is very small, so this variance is tiny!

As $n \to \infty$ so the cross terms

$$\widetilde{\gamma} \xrightarrow{L} MN\left(0, 2\omega^2[Y]B\right),$$

where

$$B = \begin{pmatrix} 1 & \bullet & \bullet & \bullet & \bullet \\ -1 & 2 & \bullet & \bullet & \bullet \\ 0 & -1 & 2 & \bullet & \bullet \\ \vdots & \ddots & \ddots & \ddots & \bullet \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

Minimised by Bartlett kernel — which is close to the Zhang, Mykland, and Aït-Sahalia (2005) subsampler. Poor rate of convergence $n^{-1/4}$.

# 5 Conclusions

Market frictions are complicated under 1 minute.

Their impact has dramtically fallen in recent years.

Ghysels and Sinko (2005) have shown that these robust measures do not, in any case, really improve our ability to forecast volatility.

This suggests one might be negative about this new stream of work.

Multivariate case is important though, e.g. Epps effects. So although the methods developed here are univariate, their payoff will be in the multivariate case.