Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise

Ole E. Barndorff-Nielsen
Peter R. Hansen
Asger Lunde
Neil Shephard

The Oxford-Man Institute, University of Oxford
Working paper, OMI01/07

Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise*

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

Peter Reinhard Hansen
Department of Economics, Stanford University,
Landau Economics Building, 579 Serra Mall,
Stanford, CA 94305-6072, USA
peter.hansen@stanford.edu

Asger Lunde
Department of Marketing and Statistics,
Aarhus School of Business, University of Aarhus,
Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark
alunde@asb.dk

Neil Shephard
Oxford-Man Institute, University of Oxford,
Blue Boar Court, Alfred Road, Oxford OX1 4EH, UK
neil.shephard@economics.ox.ac.uk

July 5, 2007

Abstract

This paper shows how to use realised kernels to carry out efficient feasible inference on the ex-post variation of underlying equity prices in the presence of simple models of market frictions. The weights can be chosen to achieve the best possible rate of convergence and to have an asymptotic variance which is close to that of the maximum likelihood estimator in the parametric version of this problem. Realised kernels can also be selected to (i) be analysed using endogenously spaced data such as that in databases on transactions, (ii) allow for market frictions which are endogenous, (iii) allow for temporally dependent noise. The finite sample performance of our estimators is studied using simulation, while empirical work illustrates their use in practice.

Keywords: Bipower variation; Long run variance estimator; Market frictions; Quadratic variation; Realised variance.

*An initial version of this paper was circulated in the Autumn of 2004 under the title "Regular and modified kernel-based estimators of integrated variance: the case with independent noise." We thank Yacine Aït-Sahalia, Clive Bowsher, Xin Huang, Ilze Kalnina, Silja Kinnebrock, Jeremy Large, Nour Meddahi, Ulrich Müller, Per Mykland, Mark Podolskij, Joe Romano, George Tauchen, Lan Zhang, anonomous referees, and the co-editor, Whitney Newey for valuable comments. Financial support from the Research Network "Microstructure of Financial Markets in Europe" is gratefully acknowledged. The Ox language of Doornik (2001) was used to perform the calculations reported here.

1 Introduction

In the last six years the harnessing of high frequency financial data has lead to substantial improvements in our understanding of financial volatility. The idea behind this is to use quadratic variation as a measure of the ex-post variation of asset prices. Estimators of increments of this quantity can allow us, for example, to improve forecasts of future volatility and estimate parametric models of time varying volatility. The most commonly used estimator of this type is the realised variance (e.g. Andersen, Bollerslev, Diebold, and Labys (2001), Meddahi (2002) and Barndorff-Nielsen and Shephard (2002)), which the recent econometric literature has shown has good properties when applied to 10 to 30 minute return data for frequently traded assets.

A weakness with realised variance is that it can be unacceptably sensitive to market frictions when applied to returns recorded over shorter time intervals such as 1 minute, or even more ambitiously, 1 second (e.g. Zhou (1996), Fang (1996) and Andersen, Bollerslev, Diebold, and Labys (2000)). In this paper we study the class of *realised kernel* estimators of quadratic variation. We show how to design these estimators to be robust to certain types of frictions and to be efficient.

The problem of estimating the quadratic variation is, in some ways, similar to the estimation of the long-run variance in stationary time series. For example, the realized variance is analogous to the sum-of-squares variance estimator. The moving average filter of Andersen, Bollerslev, Diebold, and Ebens (2001) and Hansen, Large, and Lunde (2006) and the autoregressive filter of Bollen and Inder (2002), are estimators that use pre-whitening techniques — see also Bandi and Russell (2005a). Aït-Sahalia, Mykland, and Zhang (2005) and Oomen (2005) propose parametric estimators. The two scale estimator of Zhang, Mykland, and Aït-Sahalia (2005) was the first consistent nonparametric estimator for stochastic volatility plus noise processes. It is related to the earlier work of Zhou (1996) on scaled Brownian motion plus noise. The multiscale estimator of Zhang (2006) is more efficient than the two scale estimator. An alternative is due to Large (2005), whose alternation estimator applies when prices move by a sequence of single ticks. Finally, Delattre and Jacod (1997) studied the effect of rounding on realised variances.

More formally, our interest will be in inference for the ex-post variation of log-prices over some arbitrary fixed time period, such as a day, using estimators of realised kernel type. In order to focus on the core issue we represent this period as the single interval $[0, t]$. For a continuous time log-price process X and time gap $\delta > 0$ our flat-top *realised kernels* take on the following form

$$K(X_\delta) = \gamma_0(X_\delta) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left\{\gamma_h(X_\delta) + \gamma_{-h}(X_\delta)\right\}.$$

Here the non-stochastic $k(x)$ for $x \in [0, 1]$ is a weight function and the h-th realised autocovariance is

$$\gamma_h(X_\delta) = \sum_{j=1}^{n} \left(X_{\delta j} - X_{\delta(j-1)}\right)\left(X_{\delta(j-h)} - X_{\delta(j-h-1)}\right),$$

with $h = -H, ..., -1, 0, 1, ..., H$ and $n = \lfloor t/\delta \rfloor$. We will think of δ as being small and so $X_{\delta j} - X_{\delta(j-1)}$ represents the j-th high frequency return, while $\gamma_0(X_\delta)$ is the realised variance of X. Here $K(X_\delta) - \gamma_0(X_\delta)$ is the realised kernel correction to realised variance for market frictions.

We show that if $k(0) = 1$, $k(1) = 0$ and $H = c_0 n^{2/3}$ then the resulting estimator is asymptotically mixed Gaussian, converging at rate $n^{1/6}$. Here c_0 is a estimable constant which can be optimally chosen as a function of k, the variance of the noise and a function of the volatility path, to minimise the asymptotic variance of the estimator. The special case of a so-called flat-top Bartlett kernel, where $k(x) = 1 - x$, is particularly interesting as its asymptotic distribution is the same as that of the two scale estimator.

When we additionally require that $k'(0) = 0$ and $k'(1) = 0$ then by taking $H = c_0 n^{1/2}$ the resulting estimator is asymptotically mixed Gaussian, converging at rate $n^{1/4}$, which we know is the fastest possible rate. When $k(x) = 1 - 3x^2 + 2x^3$ this estimator has the same asymptotic distribution as the multiscale estimator.

We use our novel realised kernel framework to make three innovations to the literature: (i) we design a kernel to have an asymptotic variance which is smaller than the multiscale estimator, (ii) we design $K(X_\delta)$ for data with endogenously spaced data, such as that in databases on transactions (see Renault and Werker (2005) for the importance of this), (iii) we cover the case where the market frictions are endogenous. All of these results are new and the last two of them are essential from a practical perspective.

Clearly these realised kernels are related to so-called HAC estimators discussed by, for example, Gallant (1987), Newey and West (1987), and Andrews (1991). The flat-top of the kernel, where a unit weight is imposed on the first autocovariance, is related to the flat-top literature initiated by Politis and Romano (1995) and Politis (2005). However, the realised kernels are not scaled by the sample size, which has a great number of technical implications and makes their analysis subtle.

The econometric literature on realised kernels was started by Zhou (1996) who proposed $K(X_\delta)$ with $H = 1$. This suffices for eliminating the bias caused by frictions under a simple model for frictions where the population values of higher-order autocovariances of the market frictions are zero. However, the estimator is inconsistent. Hansen and Lunde (2006) use realised kernel type estimators, with $k(x) = 1$ for general H to characterize the second order properties of market microstructure noise. Again these are inconsistent estimators. Some analysis of the finite sample performance of a type of inconsistent realised kernel is provided by Bandi and Russell (2006), who focus on the selection of H in the case where $k(x) = 1 - x$, the Bartlett kernel.

In Section 2 we detail our notation and assumptions about the efficient price process, market frictions and realised kernels. In Section 3 we give a central limit theory for $\gamma_h(X_\delta)$. Section 4 then looks at the corresponding properties of realised kernels. Here we also analyse the realised kernels with an asymptotic scheme that takes the level of market frictions local to zero. In Section 5 we study the effect irregularly spaced data has on our theory and extend the analysis of realised kernels to the case with jumps and the case where the noise is temporally dependent and endogenous. Section 6 performs a Monte Carlo experiment to assess the accuracy of our feasible central limit theory. In Section 7 we apply the theory to some data taken from the New York stock exchange and in Section 8 we draw conclusions. Some intermediate results on stable convergence is presented in Appendix A and a lengthy Appendix B details the proofs of the results given in the paper.

2 Notation, definitions and background

2.1 Semimartingales and quadratic variation

The fundamental theory of asset prices says that the log-price at time t, Y_t, must, in a frictionless arbitrage free market, obey a *semimartingale* process (written $Y \in \mathcal{SM}$) on some filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq T^*}, P\right)$, where $T^* \leq 0$. Introductions to the economics and mathematics of semimartingales are given in Back (1991) and Protter (2004). It is unusual to start the clock of a semimartingale before time 0, but this raises no technical difficulty and eases the exposition. We think of 0 as the start of an economic day and sometimes it is useful to use data from the previous day. Alternatively we could define $\gamma_h(X_\delta)$ as using data from time 0 to t by changing the range of the summation to $j = H+1$ and $n-H$ and then scaling the resulting estimator. All the theoretical properties we discuss in this paper would then follow in the same way as here.

Crucial to semimartingales, and to the economics of financial risk, is the *quadratic variation* (QV) process of $Y \in \mathcal{SM}$. This can be defined as

$$[Y]_t = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^{t_j \leq t} \left(Y_{t_j} - Y_{t_{j-1}}\right)^2, \tag{1}$$

(e.g. Protter (2004, p. 66–77) and Jacod and Shiryaev (2003, p. 51)) 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 $n \to \infty$. Discussion of the case of stochastic spacing $\{t_j\}$ will be given in Section 5.3.

The most familiar semimartingales are of *Brownian semimartingale* type ($Y \in \mathcal{BSM}$)

$$Y_t = \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u, \tag{2}$$

where a is a predictable locally bounded drift, σ is a càdlàg volatility process and W is a Brownian motion. This rules out jumps in Y, an issue addressed in Section 5.6. For reviews of the econometrics of $Y \in \mathcal{BSM}$ see, for example, Ghysels, Harvey, and Renault (1996) and Shephard (2005). If $Y \in \mathcal{BSM}$ then

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

In some of our asymptotic theory we also assume, for simplicity of exposition, that

$$\sigma_t = \sigma_0 + \int_0^t a_u^\# \mathrm{d}u + \int_0^t \sigma_u^\# \mathrm{d}W_u + \int_0^t v_u^\# \mathrm{d}V_u, \tag{3}$$

where $a^{\#}$, $\sigma^{\#}$ and $v^{\#}$ are adapted càdlàg processes, with $a^{\#}$ also being predictable and locally bounded and V is Brownian motion independent of W. Moreover, σ^2 is assumed to be (a.s.) pathwise positive on every compact interval. Much of what we do here can be extended to allow for jumps in σ, following the details discussed in Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006), but we will not address that here.

2.2 Assumptions about noise

We write the effects of market frictions as U, so that we observe the process

$$X = Y + U, \tag{4}$$

and think of $Y \in \mathcal{BSM}$ as the efficient price. Our scientific interest will be in estimating $[Y]_t$. In the main part of our work we will assume that $Y \perp\!\!\!\perp U$ where, in general, $A \perp\!\!\!\perp B$ denotes that A and B are independent. From a market microstructure theory viewpoint this is a strong assumption as one may expect U to be correlated with increments in Y, (see e.g. Kalnina and Linton (2006)). However, the empirical work of Hansen and Lunde (2006) suggests this independence assumption is not too damaging statistically when we analyse data in thickly traded stocks recorded every minute. In Section 5.5 we will show realised kernels are consistent when this assumption is relaxed.

Furthermore we mostly work under a white noise assumption about the U process ($U \in \mathcal{WN}$) which we assume has

$$\mathrm{E}(U_t) = 0, \quad \mathrm{Var}(U_t) = \omega^2, \quad \mathrm{Var}(U_t^2) = \lambda^2\omega^4, \quad U_t \perp\!\!\!\perp U_s$$

for any $t, s, \lambda \in \mathbb{R}^+$. This white noise assumption is unsatisfactory from a number of viewpoints (e.g. Phillips and Yu (2006) and Kalnina and Linton (2006)) but is a useful starting point if we think of the market frictions as operating in tick time (e.g. Bandi and Russell (2005b), Zhang, Mykland, and Aït-Sahalia (2005) and Hansen and Lunde (2006)). A feature of $U \in \mathcal{WN}$ is that $[U]_t = \infty$. Thus $U \notin \mathcal{SM}$ and so in a frictionless market would allow arbitrage opportunities. Hence it only makes sense to add processes of this type when there are frictions to be modelled. In Section 5.4 we will show our kernel can be made to be consistent when the when the white noise assumption is dropped. This type of property has been achieved earlier by the two scale estimator of Aït-Sahalia, Mykland, and Zhang (2006). Further, Section 4.7 provides a small-ω^2 asymptotic analysis which provides an alternative prospective on the properties of realised kernels.

2.3 Defining the realised autocovariation process

We measure returns over time spans of length δ. Define, for any processes X and Z,

$$\gamma_h(Z_\delta, X_\delta) = \sum_{j=1}^{n} \left(Z_{j\delta} - Z_{(j-1)\delta}\right)\left(X_{(j-h)\delta} - X_{(j-h-1)\delta}\right), \quad h = -H, ..., -1, 0, 1, 2, ..., H.$$

We call $\gamma_h(X_\delta) = \gamma_h(X_\delta, X_\delta)$ the realised autocovariation process, while noting that

$$\gamma_h(X_\delta) = \gamma_h(Y_\delta) + \gamma_h(Y_\delta, U_\delta) + \gamma_h(U_\delta, Y_\delta) + \gamma_h(U_\delta). \tag{5}$$

The daily increments of the realised QV, $\gamma_0(X_\delta)$, are called realised variances, their square root the realised volatility. Realised volatility has a very long history. It appears in, for example, Rosenberg (1972), Merton (1980) and French, Schwert, and Stambaugh (1987), with Merton (1980) making the implicit connection with the case where $\delta \downarrow 0$ in the pure scaled Brownian motion plus drift case. For more general processes a closer connection between realised QV and QV, and its use for econometric purposes, was made in Andersen, Bollerslev, Diebold, and Labys (2001), Comte and Renault (1998) and Barndorff-Nielsen and Shephard (2002).

2.4 Defining the realised kernel

We study the realised kernel

$$K(X_\delta) = \gamma_0(X_\delta) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\{\gamma_h(X_\delta) + \gamma_{-h}(X_\delta)\}, \tag{6}$$

when $k(0) = 1$ and $k(1) = 0$, noting that $K(X_\delta) = K(Y_\delta) + K(Y_\delta, U_\delta) + K(U_\delta, Y_\delta) + K(U_\delta)$. Throughout we will write "$\intercal$" to denote a transpose,

$$\begin{aligned}\gamma(X_\delta) &= \{\gamma_0(X_\delta), \gamma_1(X_\delta) + \gamma_{-1}(X_\delta), ..., \gamma_H(X_\delta) + \gamma_{-H}(X_\delta)\}^\intercal, \\ \gamma(Y_\delta, U_\delta) &= (\gamma_0(Y_\delta, U_\delta), \gamma_1(Y_\delta, U_\delta) + \gamma_{-1}(Y_\delta, U_\delta), ..., \gamma_H(Y_\delta, U_\delta) + \gamma_{-H}(Y_\delta, U_\delta))^\intercal.\end{aligned}$$

2.5 Maximum likelihood estimator of QV

In order to put non-parametric results in context, it is helpful to have a parametric benchmark. In this subsection we recall the behaviour of the maximum likelihood (ML) estimator of $\sigma^2 = [Y]_1$ when $Y_t = \sigma W_t$ and where the noise is Gaussian. All the results we state here are already known.

Given $Y \perp\!\!\!\perp U$ and taking $t = 1$ it follows that

$$\begin{pmatrix} X_{1/n} - X_0 \\ X_{2/n} - X_{1/n} \\ \vdots \\ X_1 - X_{(n-1)/n} \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \frac{\sigma^2}{n} I + \begin{pmatrix} 2\omega^2 & \bullet & \bullet & \bullet \\ -\omega^2 & 2\omega^2 & \bullet & \bullet \\ 0 & -\omega^2 & 2\omega^2 & \bullet \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}\right\}.$$

Let $\hat{\sigma}^2_{\mathrm{ML}}$ and $\hat{\omega}^2_{\mathrm{ML}}$ denote the ML estimators. Their asymptotic properties are given from classical results about the MA(1) process. By adopting the expression given in Aït-Sahalia, Mykland, and Zhang (2005, Proposition 1) to our notation, we have that for $\omega^2 > 0$

$$\left\{\begin{array}{c} n^{1/4}\left(\hat{\sigma}^2_{\mathrm{ML}} - \sigma^2\right) \\ n^{1/2}\left(\hat{\omega}^2_{\mathrm{ML}} - \omega^2\right) \end{array}\right\} \xrightarrow{L} N\left\{0, \begin{pmatrix} 8\omega\sigma^3 & 0 \\ 0 & 2\omega^4 \end{pmatrix}\right\}. \tag{7}$$

The slow rate of convergence of $\hat{\sigma}^2_{\mathrm{ML}}$ is a familiar result from the work of, for example, Stein (1987) and Gloter and Jacod (2001a, 2001b).

Interestingly, Aït-Sahalia, Mykland, and Zhang (2005) have shown that the asymptotic distribution of $\hat{\sigma}^2_{\mathrm{ML}}$ does not depend on the actual distribution of U. Hence, if U is non-Gaussian distributed (making $\hat{\sigma}^2_{\mathrm{ML}}$ a quasi maximum likelihood estimator) we continue to have $n^{1/4}\left(\hat{\sigma}^2_{\mathrm{ML}} - \sigma^2\right) \xrightarrow{L} N(0, 8\omega\sigma^3)$. So $n^{1/4}$ is the fastest possible rate of convergence unless additional assumptions are made about the distribution of U. For example, if U is assumed to have a two point distribution it is then possible to recover the convergence rate of $n^{1/2}$ by carrying out ML estimation on this alternative parametric model.

The special case where there is no market microstructure noise, (i.e. the true value of $\omega^2 = 0$) results in faster rates of convergence for $\hat{\sigma}^2_{\mathrm{ML}}$, since $n^{1/2}\left(\hat{\sigma}^2_{\mathrm{ML}} - \sigma^2\right) \xrightarrow{L} N\left(0, 6\sigma^4\right)$. When ω^2 is also known a priori to be zero, and so is not estimated, then

$$n^{1/2}\left(\hat{\sigma}^2_{\mathrm{ML}} - \sigma^2\right) \xrightarrow{L} N\left(0, 2\sigma^4\right). \tag{8}$$

2.6 Notation and jittering

To simplify the exposition of some results we redefine the price measurements at the two end-points, X_0 and X_t, to be an average of m distinct observations in the intervals $(-\delta, \delta)$ and $(t-\delta, t+\delta)$, respectively. This jittering can be used to eliminate end-effects that would otherwise appear in the asymptotic variance of $K(U_\delta)$, in some cases. The jittering does not affect consistency, rate of convergence, or the asymptotic results concerning $K(Y_\delta)$ and $K(Y_\delta, U_\delta)$.

In the following we consider kernel weight functions, $k(x)$, that are two times continuously differentiable on $[0,1]$, and define

$$k_\bullet^{0,0} = \int_0^1 k(x)^2 \mathrm{d}x, \quad k_\bullet^{1,1} = \int_0^1 k'(x)^2 \mathrm{d}x, \quad k_\bullet^{2,2} = \int_0^1 k''(x)^2 \mathrm{d}x, \tag{9}$$

where we, as usual, write derivatives using primes. The kernels for which $k'(0)^2 + k'(1)^2 = 0$ are particularly interesting in this context, and we shall refer to this class of kernel as *smooth kernels.*

3 Central limit theory for $\gamma(X_\delta)$

Readers uninterested in the background theory of realised kernels can skip this section and go immediately to Section 4.

3.1 Background result

Here we will study the large sample behaviour of the contributions to $\gamma(X_\delta)$. These results will be used in the proofs of the next Section's results on the properties of $K(X_\delta)$ and so to select k to produce good estimators of $[Y]$. Throughout this paper $\xrightarrow{L_Y}$ will denote convergence in law stably with respect to the σ-field, $\sigma(Y)$, generated by the process Y.

Theorem 1 *Suppose that $Y \in \mathcal{BSM}$ with σ of the form (3) and that $U \in \mathcal{WN}$ with $U \perp\!\!\!\perp Y$. Let*

$$\Gamma_{\delta,H} = \left(\gamma_0 \left(Y_\delta\right) - \int_0^t \sigma_u^2 \mathrm{d}u, \gamma_1 \left(Y_\delta\right) + \gamma_{-1} \left(Y_\delta\right), \ldots, \gamma_H \left(Y_\delta\right) + \gamma_{-H} \left(Y_\delta\right) \right)^\top .$$

As $n \to \infty$, the random variates

$$\delta^{-1/2} \Gamma_{\delta,H}, \quad \gamma \left(Y_\delta, U_\delta\right) + \gamma \left(U_\delta, Y_\delta\right), \quad \delta^{1/2} \left\{ \gamma \left(U_\delta\right) - E\gamma \left(U_\delta\right) \right\}$$

converge jointly in law and $\sigma(Y)$-stably. The limiting laws are as follows:

$$\Gamma_{\delta,H} \xrightarrow{L_Y} MN \left(0, 2 \int_0^t \sigma_u^4 \mathrm{d}u \times A \right), \quad A = \mathrm{diag}\left(1, 2, \ldots, 2\right),$$

where MN denotes a mixed normal distribution; $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta) \xrightarrow{L_Y} MN\left(0, 8\omega^2 [Y] B\right)$, where B is a $(H+1) \times (H+1)$ symmetric matrix with block structure

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \; B_{11} = \begin{pmatrix} 1 & \bullet \\ -1 & 2 \end{pmatrix}, \; B_{21} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \; B_{22} = \begin{pmatrix} 2 & \bullet & \bullet & \bullet \\ -1 & 2 & \bullet & \bullet \\ \ddots & \ddots & \ddots & \bullet \\ \cdots & 0 & -1 & 2 \end{pmatrix}$$

$B_{12} = B_{21}^\intercal$; and

$$\mathrm{E}\left\{\gamma(U_\delta)\right\} = 2\omega^2 n \left(1, -1, 0, 0, ..., 0\right)^\intercal, \quad \mathrm{Cov}\left\{\gamma(U_\delta)\right\} = 4\omega^4 \left(nC + D + m^{-1}E\right),$$

where C, D, and E are symmetric $(H+1) \times (H+1)$ matrices; C with the block structure:

$$C_{11} = \begin{pmatrix} 1 + \lambda^2 & \bullet \\ -2 - \lambda^2 & 5 + \lambda^2 \end{pmatrix}, \; C_{21} = \begin{pmatrix} 1 & -4 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \; C_{22} = \begin{pmatrix} 6 & \bullet & \bullet & \bullet \\ -4 & 6 & \bullet & \bullet \\ 1 & -4 & 6 & \bullet \\ 0 & \ddots & \ddots & \ddots \end{pmatrix},$$

where $C_{12} = C_{21}^\intercal$;

$$D = \begin{pmatrix} -\lambda^2 - 2 & \bullet & \bullet & \bullet & \bullet & \bullet \\ \lambda^2 + 4 & -\lambda^2 - \frac{21}{2} & \bullet & \bullet & \bullet & \bullet \\ -\frac{4}{2} & 9 & -15 & \bullet & \bullet & \bullet \\ 0 & -\frac{5}{2} & 11 & -18 & \bullet & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\frac{H+2}{2} & 2H+5 & -3(H+3) \end{pmatrix};$$

and E with the block structure: $E_{12} = E_{21}^\intercal$,

$$E_{11} = \begin{pmatrix} \frac{\frac{\lambda^2}{2} + (m-1)}{m^2} + 2 & \bullet \\ -\frac{\frac{\lambda^2}{2} + (m-1)}{m^2} - 3 & \frac{\frac{\lambda^2}{2} + (m-1)}{m^2} + 7 \end{pmatrix}, \; E_{21} = \begin{pmatrix} 1 & -5 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}, \; E_{22} = \begin{pmatrix} 8 & \bullet & \bullet & \cdots \\ -5 & 8 & \bullet & \ddots \\ 0 & -5 & 8 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Finally, $n^{-1/2} \left\{\gamma(U_\delta) - \mathrm{E}\gamma(U_\delta)\right\} \xrightarrow{L} N(0, 4\omega^4 C)$.

Using $\gamma_h(U_\delta) + \gamma_{-h}(U_\delta)$ in the construction of our realised kernels, rather than $2\gamma_h(U_\delta)$, say, is essential for obtaining a consistent estimator. The explanation is simple. Part of the realised variance, $\gamma_0(U_\delta)$, is given by $U_0^2 + 2\sum_{j=1}^{n-1} U_{\delta j}^2 + U_t^2$, and the corresponding terms in $\gamma_1(U_\delta)$ and $\gamma_{-1}(U_\delta)$ are given by $U_0^2 + \sum_{j=1}^{n-1} U_{\delta j}^2$ and $\sum_{j=1}^{n-1} U_{\delta j}^2 + U_t^2$, respectively. So both $\gamma_1(U_\delta)$ and $\gamma_{-1}(U_\delta)$ are needed in order to eliminate the two end-terms, U_0^2 and U_t^2, because these terms do not appear in $\gamma_h(U_\delta)$ for $|h| \geq 2$.

3.2 Comments

3.2.1 Stable convergence

The concept and role of stable convergence may be unfamiliar to some readers and we therefore add some words of explanation. The formal definition of stable convergence (given in the appendix) conceals a key property of stable convergence, which is a useful joint convergence. Let $\mathcal{Y}_n$ denote a random variate on $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. $\mathcal{Y}_n$ *converges* $\mathcal{G}$*-stably in law* to $\mathcal{Y}$, written $\mathcal{Y}_n \overset{L_\mathcal{G}}{\to} \mathcal{Y}$, if and only if $(\mathcal{Y}_n, Z) \overset{L}{\to} (\mathcal{Y}, Z)$ for all $\mathcal{G}$-measurable Z random variables Z and some random variate $\mathcal{Y}$. When $\mathcal{G} = \sigma(X)$ we will write $\overset{L_X}{\to}$ in place of $\overset{L_\mathcal{G}}{\to}$.

Consider the following simple example where

$$\mathcal{Y}_n = \delta^{-1/2}\left(\gamma_0(Y_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right) \overset{L_X}{\to} MN(0, 2Z) \qquad \text{and} \qquad Z = \int_0^t \sigma_u^4 \mathrm{d}u, \tag{10}$$

Our focus is on $\mathcal{Y}_n/\sqrt{Z}$ and if Z is $\mathcal{G}$-measurable then convergence $\mathcal{G}$-stably in law implies that

$$\delta^{-1/2}\left(\gamma_0(Y_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right) / \sqrt{\int_0^t \sigma_u^4 \mathrm{d}u} \overset{L}{\to} N(0, 2), \tag{11}$$

a result that cannot be deduced from the convergence in law to a mixed Gaussian variable in (10) without stable convergence.

3.2.2 Related results

The asymptotic distribution (10) appears in Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002). This estimator has the efficiency of the ML estimator (8) in the pure Brownian motion case. The extension of the limiting results to deal with more general realised autocovariances is new. The first part of Theorem 1 implies that the simple kernel

$$\delta^{-1/2}\left(\gamma_0(Y_\delta) + \gamma_1(Y_\delta) + \gamma_{-1}(Y_\delta) - [Y]_t\right) \overset{L_X}{\to} MN\left(0, 6\int_0^t \sigma_u^4 \mathrm{d}u\right),$$

in the no-noise case. This increases the asymptotic variance by a factor of three relative to that in (10). So in the absence of noise there will be no gains from realised kernels.

The main impact of the noise is through the $\gamma(U_\delta)$ term. The mean and variance of $\gamma_0(U_\delta)$ was studied by, for example, Fang (1996), Bandi and Russell (2005b) and Zhang, Mykland, and Aït-Sahalia (2005). Note that both the mean and variance of $\gamma_0(U_\delta)$ explode as $n \to \infty$. Of course these features are passed onto $\gamma_0(X_\delta)$ making it inconsistent, thus motivating this literature. The bias of $\gamma_0(U_\delta)$ is exactly balanced by that of $\gamma_1(U_\delta)+\gamma_{-1}(U_\delta)$, so producing the asymptotically unbiased but inconsistent estimator $\gamma_0(X_\delta)+\gamma_1(X_\delta)+\gamma_{-1}(X_\delta)$ with (e.g. Zhou (1996)) $\mathrm{E}\left\{\gamma_0(U_\delta)+\gamma_1(U_\delta)+\gamma_{-1}(U_\delta)\right\}=0$ and $\mathrm{Var}\left\{\gamma_0(U_\delta)+\gamma_1(U_\delta)+\gamma_{-1}(U_\delta)\right\}=4\omega^4\left(2n-1.5\right).$

The higher-order autocovariances, $h\geq 2$, are noisy estimates of zero as $\mathrm{E}\left\{\gamma_h(U_\delta)+\gamma_{-h}(U_\delta)\right\}=0$ and $\mathrm{Var}\left\{\gamma_h(U_\delta)+\gamma_{-h}(U_\delta)\right\}\propto n$. Yet, including them can reduce the variance, and this is essential for obtaining a consistent estimator. Thus the higher-order autocovariances play the role of control variables (e.g. Ripley (1987, p.118)). For example, one can show that $\mathrm{Var}\left\{K(U_\delta)\right\}\simeq\frac{n}{H^2}8\omega^4$ when the Bartlett kernel is employed, and this shows that increasing H with n makes it possible to reduce the variance induced by the noise.

The structure of the matrices, A, B, C, D, and E, is key for the asymptotic properties of our realised kernel, and we have the following result.

Theorem 2 *Write* $w=\left(1,1,k\left(\frac{1}{H}\right),...,k\left(\frac{H-1}{H}\right)\right)^\intercal$. *Then as* H *increases,*

$$\begin{aligned}
w^\intercal Aw &= 2Hk_\bullet^{0,0}+O(1),\\
w^\intercal Bw &= H^{-1}k_\bullet^{1,1}+O(H^{-2}),\\
w^\intercal Cw &= \begin{cases} H^{-2}\left\{k'(0)^2+k'(1)^2\right\}+O(H^{-3}), & \textit{if } k'(0)^2+k'(1)^2\neq 0,\\ H^{-3}k_\bullet^{2,2}+O(H^{-4}), & \textit{if } k'(0)^2+k'(1)^2=0,\end{cases}\\
w^\intercal Dw &= -H^{-1}\tfrac{1}{2}k'(1)^2+O(H^{-2}),\\
w^\intercal Ew &= H^{-1}k_\bullet^{1,1}+O(H^{-2}).
\end{aligned}$$

It may be interesting to note that the Bartlett kernel minimises, $w^\intercal Bw$, while the cubic kernel function, $k(x)=1-3x^2+2x^3$, minimises the asymptotic contribution from $w^\intercal Cw$.

4 Behaviour of kernels

4.1 Core result

In this Section we derive the asymptotic behaviour of arbitrary realised kernels. In Section 4.3 we derive a way of choosing the number of terms to use in the kernel, which is indexed by ω^2 and $\int_0^t\sigma_u^4\mathrm{d}u$. Subsequently we provide estimators of these quantities, implying the feasible asymptotic distribution of the realised kernel can be applied in practice to form confidence intervals for $[Y]$.

Recalling the definition of $k_\bullet^{0,0}$, $k_\bullet^{1,1}$, and $k_\bullet^{2,2}$ in (9) we have

Theorem 3 *As* $n,H\to\infty$ *and* $H/n\to 0$

$$\begin{aligned}
\sqrt{\frac{n}{H}}\left\{K(Y_\delta)-\int_0^t\sigma_u^2\mathrm{d}u\right\} &\overset{L_Y}{\to} MN\left(0,4k_\bullet^{0,0}t\int_0^t\sigma_u^4\mathrm{d}u\right),\\
\sqrt{H}\left\{K(Y_\delta,U_\delta)+K(U_\delta,Y_\delta)\right\} &\overset{L_Y}{\to} MN\left(0,k_\bullet^{1,1}8\omega^2\int_0^t\sigma_u^2\mathrm{d}u\right),
\end{aligned}$$

$$\sqrt{\frac{H^2}{n}}\left\{K(U_\delta)\right\} \quad \xrightarrow{L} \quad N\left[0, 4\omega^4\left\{k'(0)^2 + k'(1)^2\right\}\right].$$

When $k'(0)^2 + k'(1)^2 = 0$ *the asymptotic variance of* $K(U_\delta)$ *is* $4\omega^4\left(\frac{n}{H^3}k_\bullet^{2,2} + \frac{1}{Hm}k_\bullet^{1,1}\right)$, *and*

$$\sqrt{\frac{H^3}{n}}\left\{K(U_\delta)\right\} \xrightarrow{L} N\left(0, 4\omega^4 k_\bullet^{2,2}\right), \qquad \text{if } H^2/(mn) \to 0.$$

It is useful to define

$$\xi^2 = \omega^2 \Big/ \sqrt{t\int_0^t \sigma_u^4 \mathrm{d}u} \quad \text{and} \quad \rho = \int_0^t \sigma_u^2 \mathrm{d}u \Big/ \sqrt{t\int_0^t \sigma_u^4 \mathrm{d}u},$$

to be a noise-to-signal ratio and a measure of heteroskedasticity, respectively. Note that $\rho = 1$ corresponds to the case with constant volatility, and by Cauchy-Schwartz inequality we have $\rho \leq 1$.

The large n and large H asymptotic variance of $K(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u$ is

$$4t\int_0^t \sigma_u^4 \mathrm{d}u\left[\tfrac{H}{n}k_\bullet^{0,0} + 2\tfrac{k_\bullet^{1,1}}{H}\rho\xi^2 + n\left\{\tfrac{k'(0)^2 + k'(1)^2}{H^2} + \tfrac{k_\bullet^{2,2}}{H^3}\right\}\xi^4 + \tfrac{k_\bullet^{1,1}}{Hm}\xi^4\right]. \tag{12}$$

If we now relate H to n, we see that $k'(0)^2 + k'(1)^2 = 0$ is an important special case. This is spelled out in the following Theorem.

Theorem 4 *When* $H = c_0 n^{2/3}$ *we have*

$$n^{1/6}\left\{K(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \xrightarrow{L_X} MN\left(0, 4t\int_0^t \sigma_u^4 \mathrm{d}u\left[c_0 k_\bullet^{0,0} + c_0^{-2}\left\{k'(0)^2 + k'(1)^2\right\}\xi^4\right]\right). \tag{13}$$

When $k'(0)^2 + k'(1)^2 = 0$, $m \to \infty$, *and* $H = c_0 n^{1/2}$ *we have*

$$n^{1/4}\left\{K(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \xrightarrow{L_X} MN\left\{0, 4t\int_0^t \sigma_u^4 \mathrm{d}u\left(c_0 k_\bullet^{0,0} + c_0^{-1}2k_\bullet^{1,1}\rho\xi^2 + c_0^{-3}k_\bullet^{2,2}\xi^4\right)\right\}. \tag{14}$$

The result (14) is interesting because we have seen, in (7), that this is the best rate of convergence that can be achieved for this problem.

The requirement that $m \to \infty$ for the result (14) is due to end-effects. When m is fixed an additional term appears in the asymptotic variance. Its relative contribution to the asymptotic variance is proportional to ξ^2/m. In our empirical analysis, we find ξ^2 to be quite small, about 10^{-3}, so the last term can reasonably be ignored even when $m = 1$. This argument will be spelled out in Section 4.7 where we consider a small-ω^2 asymptotic scheme. Under the alternative asymptotic experiment, this term vanishes at a sufficiently fast rate without the need for a jittering of the end-points.

4.2 Special cases with $n^{1/6}$

When $H = c(\xi^2 n)^{2/3}$ we have the asymptotic distribution given in (13) by setting $c_0 = c\xi^{4/3}$. For this class of kernels the value of c which minimises the asymptotic variance in (13) is

$$c^* = \left[2\left\{k'(0)^2 + k'(1)^2\right\}/k_\bullet^{0,0}\right]^{1/3},$$

and the lower bound for the asymptotic variance is

$$4c^*\omega^{4/3}\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3}\left[k_\bullet^{0,0}+c^{*-3}\left\{k'(0)^2+k'(1)^2\right\}\right]=6c^*k_\bullet^{0,0}\omega^{4/3}\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3}. \tag{15}$$

Hence $c^*k_\bullet^{0,0}$ controls the asymptotic efficiency of estimators in this class.

Three flat-top cases of this setup are analysed in Table 1. The flat-top Bartlett kernel puts $k(x)=1-x$, Epanechnikov kernel puts $k(x)=1-x^2$, while the second order kernel has $k(x)=1-2x+x^2$. The Bartlett kernel has the same asymptotic distribution as the two scale estimator. It is more efficient than the Epanechnikov kernel but less good than the second order kernel.

	$k(x)$	$k'(0)$	$k'(1)$	$k_\bullet^{0,0}$	$k_\bullet^{1,1}$	$k_\bullet^{2,2}$	c^*	$ck_\bullet^{0,0}$
Bartlett	$1-x$	-1	-1	$\frac{1}{3}$	1	0	2.28	0.76
2nd order	$1-2x+x^2$	-2	0	$\frac{1}{5}$	$\frac{4}{3}$	4	3.42	0.68
Epanechnikov	$1-x^2$	0	-2	$\frac{8}{15}$	$\frac{4}{3}$	4	2.46	1.31

Table 1: *Properties of some $n^{1/6}$ flat-top realised kernels. Bartlett kernel has the same asymptotic distribution as the two scale estimator. In the last column, $c^*k_\bullet^{0,0}$, measures the relative asymptotic efficiency of the realised kernels in this class.*

4.3 Special cases with $n^{1/4}$

When $H=c\xi\sqrt{n}$ and $m\to\infty$ the asymptotic variance in (14) is proportional to

$$4t\int_0^t\sigma_u^4\mathrm{d}u\left(ck_\bullet^{0,0}\xi+2c^{-1}k_\bullet^{1,1}\rho\xi+c^{-3}k_\bullet^{2,2}\xi\right)=\omega\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{\frac{3}{4}}\underbrace{4\left(ck_\bullet^{0,0}+2c^{-1}\rho k_\bullet^{1,1}+c^{-3}k_\bullet^{2,2}\right)}_{g(c)}.$$

To determine the c that minimises the asymptotic variance we simply minimise $g(c)$. Writing $x=c^2$ the first order condition is $k_\bullet^{0,0}x^2-2\rho k_\bullet^{1,1}x-3k_\bullet^{2,2}=0$. Taking the square root of the positive root yields

$$c^*=\sqrt{\rho\frac{k_\bullet^{1,1}}{k_\bullet^{0,0}}\left\{1+\sqrt{1+3d/\rho}\right\}},\quad \text{where } d=\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{\left(k_\bullet^{1,1}\right)^2}.$$

With the optimal value for c the asymptotic variance can be expressed as $g\times\omega\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{3/4}$ where

$$g=g(c^*)=\frac{16}{3}\sqrt{\rho k_\bullet^{0,0}k_\bullet^{1,1}}\left\{\frac{1}{\sqrt{1+\sqrt{1+3d/\rho}}}+\sqrt{1+\sqrt{1+3d/\rho}}\right\}.$$

From the properties of the maximum likelihood estimator, (7), in the parametric version of the problem, $\rho=1$, we should expect that $g\geq 8$. It can be shown that g increases as ρ decreases, so in the heteroskedastic case, $\rho<1$, we should expect $g>8$.

Eight flat-top cases of this setup are analysed in Table 2, and four kernel functions are plotted in Figure 1. The first is derived by thinking of a cubic kernel $k(x)=1+a_1x+a_2x^2+a_3x^3$, where

$k(x)$		$k_\bullet^{0,0}$	$k_\bullet^{1,1}$	$k_\bullet^{2,2}$	c^*	g
Cubic kernel	$1-3x^2+2x^3$	0.371	1.20	12.0	3.68	9.04
5-th order kernel	$1-10x^3+15x^4-6x^5$	0.391	1.42	17.1	3.70	10.2
6-th order kernel	$1-15x^4+24x^5-10x^6$	0.471	1.55	22.8	3.97	12.1
7-th order kernel	$1-21x^5+35x^6-15x^7$	0.533	1.71	31.8	4.11	13.9
8-th order kernel	$1-28x^6+48x^7-21x^8$	0.582	1.87	43.8	4.31	15.7
Parzen	$\begin{cases}1-6x^2+6x^3 & 0\le x\le 1/2\\ 2(1-x)^3 & 1/2\le x\le 1\end{cases}$	0.269	1.50	24.0	4.77	8.54
Tukey-Hanning$_1$	$\sin^2\{\pi/2\,(1-x)\}$	0.375	1.23	12.1	3.70	9.18
Tukey-Hanning$_2$	$\sin^2\{\pi/2\,(1-x)^2\}$	0.219	1.71	41.7	5.74	8.29
Tukey-Hanning$_5$	$\sin^2\{\pi/2\,(1-x)^5\}$	0.097	3.50	489.0	8.07	8.07
Tukey-Hanning$_{16}$	$\sin^2\{\pi/2\,(1-x)^{16}\}$	0.032	10.26	14374.0	39.16	8.02

Table 2: *Properties of some $n^{1/4}$ flat-top realised kernels. The cubic kernel has the same asymptotic distribution as the multiscale estimator. g is computed for the case $\rho = 1$* and *measures the relative asymptotic efficiency of the realised kernels in this class — 8 being the parametric efficiency bound.*

a_1, a_2, a_3 are constants. We can choose a_1, a_2, a_3 by imposing the conditions $k'(0)^2 + k'(1)^2 = 0$, and that $k(0) = 1$ and $k(1) = 0$. The resulting cubic kernel has $k(x) = 1 - 3x^2 + 2x^3$, which has some of the features of cardinal cubic splines (e.g. Park and Schowengerdt (1983)) and quadratic mother kernels (e.g. Phillips, Sun, and Jin (2003)). As stated earlier, the cubic kernel minimises $k_\bullet^{2,2}$ within the class of smooth kernels, thus in general we have $k_\bullet^{2,2} \geq 12$. It is noteworthy that the realised kernel based on the cubic kernel has the same asymptotic distribution as the multiscale estimator. Naturally, minimising $k_\bullet^{2,2}$ need not minimize g, and a well known kernel that has a smaller asymptotic variance is the flat-top Parzen kernel, which places

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1. \end{cases}$$

We also consider flat-top Tukey-Hanning$_p$ kernel, defined by

$$k(x) = \sin^2\left\{\tfrac{\pi}{2}(1-x)^p\right\}. \tag{16}$$

We call this the modified Tukey-Hanning kernel because the case $p = 1$, $\sin^2\{\pi/2\,(1-x)\} = \{1 + \cos(\pi x)\}/2$, corresponds to the usual Tukey-Hanning kernel.

Table 2 shows that the performance of the Tukey-Hanning$_1$ kernel is almost identical to that of the cubic kernel. The Parzen kernel outperforms the cubic kernel, but is not as good as the modified Tukey-Hanning kernel, (16), when $p \geq 2$. While none of the standard kernels reach the parametric efficiency bound, we see that the modified Tukey-Hanning kernel approaches the lower bound as we increase p. This kernel utilize more lags as p increases, and later we will relax the requirement that $k(1) = 0$ and consider kernels that utilize all lags, such as the quadratic spectral kernel, see e.g. Andrews (1991).

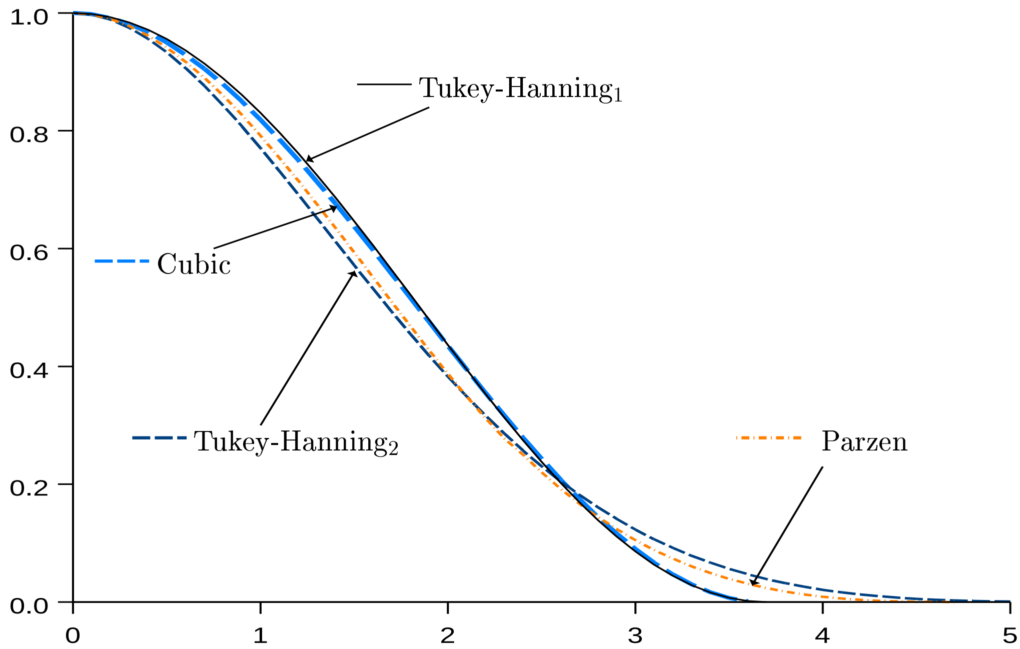

Figure 1: *Kernel functions, $k(x/c^*)$, scaled by their respective c^** to make them comparable.

4.4 Finite sample behaviour

It is important to ask whether the approximation suggested by Theorem 3 and our special cases thereof provides a useful guide to finite sample behaviour? Table 3 gives

$$g_n = \mathrm{Var}\left\{n^{1/4}K(X_\delta)\right\}\omega^{-1}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{-3/4},$$

listed against n for the case where $\rho = 1$. An empirically realistic value for ξ^2 is around 0.001 for the types of data we study later in this paper. The Table also includes results for an optimal selection of k, computed numerically.[1] This indicates that there does exist a realised kernel which can achieve the ML efficiency bound of 8 in this case. More generally the Table shows that the asymptotics provides a good approximation to the finite sample case, especially when n is over $1,000$ and when ξ^2 is moderate to large. The Table also shows that even though the Bartlett kernel converges at the slow $n^{1/6}$ rate, it is only mildly inefficient even when n is $4,000$. When ξ^2 is small the asymptotic expressions provide a poor approximation in all cases unless n is $4,000$ or so.[2] Of course, in that case the realised kernels are quite precise as the asymptotic variance is proportional to ξ. Relatively small values for ξ^2 and n result in small values for H, and this explains that the asymptotic approximation is poor in this situation. The reason is that the asymptotic approximations of Theorem 2, such as $w^\intercal Bw \simeq H^{-1}k_\bullet^{1,1}$, are inaccurate when H is small. So in our simulations and empirical analysis we compute the variance of $K(X_\delta)$ using the matrix

[1] The asymptotically optimal kernel weights, than can be computed numerically, are given by $w^* = (v_1', v_2')'$ where $v_1 = (1,1)'$ and $v_2 = -(n^{-1}A_{22} + 2\xi^2\rho B_{22} + n\xi^4 C_{22})^{-1}(2\xi^2\rho B_{21} + n\xi^4 C_{21})v_1$. The matrices A_{22}, B_{22}, C_{22}, B_{21} and C_{21} are given in Theorem 1

[2] This result has subsequently been verified by Bandi and Russell (2006) who analyse and compare the finite sample properties of several realised kernels.

	$\xi^2 = 0.1$					$\xi^2 = 0.01$				
n	Opt	TH_2	Par	Cubic	Bart	Opt	TH_2	Par	Cubic	Bart
256	8.52	9.11	9.39	9.60	10.7	9.63	10.6	10.8	10.7	10.6
$1,024$	8.30	8.76	9.03	9.37	11.9	8.73	9.43	9.73	9.81	10.3
$4,096$	8.19	8.58	8.85	9.26	13.9	8.34	8.86	9.13	9.40	10.9
$16,384$	8.14	8.49	8.76	9.21	16.8	8.17	8.58	8.84	9.22	12.5
$65,536$	8.12	8.45	8.71	9.19	20.6	8.08	8.44	8.70	9.13	14.8
$1,048,576$	8.10	8.41	8.68	9.17	31.9	8.02	8.33	8.59	9.07	22.2
∞		8.29	8.54	9.04	∞		8.29	8.54	9.04	∞
	$\xi^2 = 0.001$					$\xi^2 = 0.0001$				
n	Opt	TH_2	Par	Cubic	Bart	Opt	TH_2	Par	Cubic	Bart
256	15.1	15.4	16.2	16.1	16.9	38.7	38.8	38.8	38.8	38.8
$1,024$	10.8	11.8	12.1	12.1	11.7	21.0	21.1	21.2	23.2	21.5
$4,096$	9.22	10.0	10.3	10.4	10.5	13.2	14.0	15.0	14.9	14.0
$16,384$	8.55	9.19	9.47	9.61	10.4	10.1	11.1	11.6	11.3	11.0
$65,536$	8.26	8.73	9.00	9.31	11.3	8.93	9.69	10.0	10.0	10.2
$1,048,576$	8.06	8.40	8.66	9.10	15.8	8.20	8.64	8.90	9.25	11.9
∞		8.29	8.54	9.04	∞		8.29	8.54	9.04	∞

Table 3: *Flat-top realised kernels.* $Var\left\{n^{1/4}K(X_\delta)\right\}\omega^{-1}\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{-3/4}$ *listed against* n. *Asymptotic lower bound is 8. 'Opt' refers to kernel weights that were selected numerically to minimise the finite sample variance of a flat-top realised kernel. 'Cubic' refers to* $k(x) = 1 - 3x^2 + 2x^3$. *'*TH_2*' denotes the modified Tukey-Hanning with* $p = 2$, see *(16).*

expressions directly, rather than the asymptotic expressions of Theorem 2. This greatly improves the finite sample behaviour of confidence interval and related quantities that depend on an estimate of the asymptotic variance. Given the simple structure of the matrices this is not computationally burdensome even for large values of H.

The rest of this Section generalises the theory in various directions, and can be skipped during a first reading of the paper.

4.5 Realised kernels with infinite Lags

If we extend the kernel function beyond the unit interval, and set $k(x) = 0$ for $x \geq 1$, then the realised kernels can be expressed as

$$K(X_\delta) = \gamma_0(X_\delta) + \sum_{h=1}^{n-1} k\left(\tfrac{h-1}{H}\right)\left\{\gamma_h(X_\delta) + \gamma_{-h}(X_\delta)\right\},$$

as all autocovariance of orders higher than H are assigned zero weight. Here we consider kernels that need not have $k(x) = 0$ for $x > 1$. Such kernels can potentially assign non-zero weight to all autocovariances. So we replace the requirement "$k(x) = 0$ for $x > 1$" with "$k(x) \to 0$ as $x \to \infty$ and $k(x)$ is twice differentiable on $[0, \infty)$". An inspection of our proofs reveals that such kernels can be accommodated by our results with minor modifications. Not surprisingly, we still need $k'(0) = 0$ to achieve the fast rate of convergence and need to redefine $k_\bullet^{0,0}$, $k_\bullet^{1,1}$, and $k_\bullet^{2,2}$, to represent integrals over the whole positive axis, e.g., $k_\bullet^{0,0} = \int_0^\infty k(x)^2 \mathrm{d}x$.

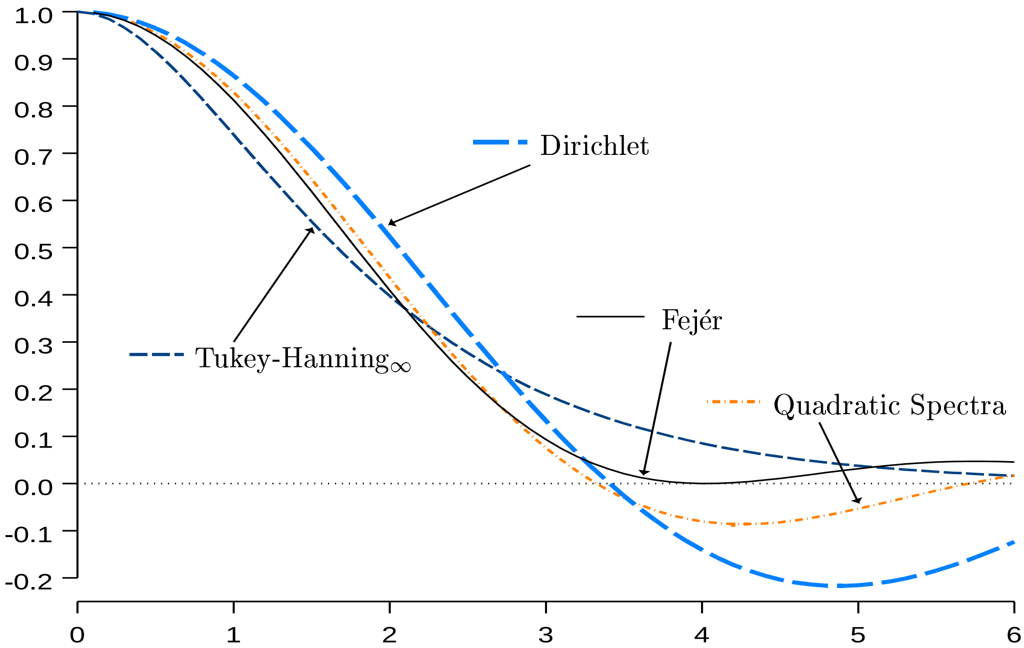

Figure 2: *Kernel functions,* $k(x/c^*)$, *scaled by their respective* c^* to make them comparable.

Four kernels of this type are given in Table 4 and Figure 2. First the Tukey-Hanning$_\infty$ kernel which is limit of (16) as $p \to \infty$. This is the kernel that gets very close to the parametric lower bound. Second, the quadratic-spectral kernel. The last two kernels are related to the Fourier-based estimators that have been used in this literature, see Malliavin and Mancino (2002). The Fourier estimator by Malliavin and Mancino (2002) is closely related to the realised kernel using the Dirichlet kernel weights,

$$k_N(z) = \frac{1}{2N+1} \frac{\sin\left((N+\frac{1}{2})z\right)}{\sin\left(\frac{1}{2}z\right)} \to k(x) = \frac{\sin(x)}{x} \text{ as } N \to \infty, \quad x = (N+1/2)\,z.$$

Mancino and Sanfelici (2007) introduce another variant of the Fourier estimator, which also has a realised kernel representation. The implied asymptotic weights for this estimator are given by the Fejér kernel: $k(x) = \sin^2(x)/x^2$. A practical drawback of these kernels, is that they require a very large number of out-of-period intraday returns. The reason is that the h-th autocovariance estimator need h intraday returns before time 0 and after time t. Because the gains in precision from these estimators is relative small we will not utilize these "infinite-lag" estimators in our simulations and empirical analysis.

$k(x)$		$k_\bullet^{0,0}$	$k_\bullet^{1,1}$	$k_\bullet^{2,2}$	c^*	g
Tukey-Hanning$_\infty$	$\sin^2\left\{\frac{\pi}{2}\exp\left(-x\right)\right\}$	0.52	$\frac{\pi^2}{16}$	$\frac{\pi^2(1+\pi^2)}{32}$	2.3970	8.0124
Quadratic Spectral	$\frac{3}{x^2}\left(\frac{\sin x}{x}-\cos x\right)$	$\frac{3\pi}{5}$	$\frac{3\pi}{35}$	$\frac{\pi}{35}$	0.7395	9.3766
Dirichlet$_\infty$	$\frac{\sin x}{x}$	$\frac{\pi}{2}$	$\frac{\pi}{6}$	$\frac{\pi}{10}$	1.0847	11.662
Fejér	$\left(\frac{\sin x}{x}\right)^2$	$\frac{\pi}{3}$	$\frac{2\pi}{15}$	$\frac{16\pi}{105}$	1.2797	8.8927

Table 4: *Properties of some $n^{1/4}$ flat-top infinite-lag realised kernels. g measures the relative asymptotic efficiency of the realised kernels in this class — 8 being the parametric efficiency bound.*

4.6 Non-flat-top kernels

The flat-top constraint is imposed on these kernels to eliminate the bias caused by frictions. If we remove the flat-top constraint then the realised kernel becomes

$$\bar{K}(X_\delta)=\gamma_0(X_\delta)+\sum_{h=1}^{H}k\left(\tfrac{h}{H}\right)\left\{\gamma_h(X_\delta)+\gamma_{-h}(X_\delta)\right\},$$

where we assume $k(0)=1$ and $k(1)=0$. Now the bias in the Bartlett case $k(x)=1-x$ is $O(n/H)=O(n^{1/3})$. In the cubic case it is $O(n/H^2)=O(1)$, which is better but not satisfactory. To remove the flat-top condition we need a kernel which is flatter to a higher order near zero, so the bias becomes negligible. For this we add the additional constraint that $k''(0)=k''(1)=0$. Simple polynomials of this type, $k(x)=1+ax^j+bx^{j+1}+cx^{j+2}$, $j=3,4,...$ yield $c=-\left(j+j^2\right)/2$, $b=2j+j^2$, $a=-1-3j/2-j^2/2$. Examples of this include

$$k(x)=\begin{cases}1-10x^3+15x^4-6x^5, & j=3\\ 1-15x^4+24x^5-10x^6, & j=4\\ 1-21x^5+35x^6-15x^7, & j=5\\ 1-28x^6+48x^7-21x^8, & j=6.\end{cases} \tag{17}$$

The bias of these estimators is $O(n/H^j)=O(n^{-(j-2)/2})$ which has no impact on its asymptotic distribution when $j\geq 3$ and should become more robust in finite samples as j increases. We call the j-th case the $j+2$-th order kernel. Table 2 shows that these estimators are less efficient than the realised kernels produced by (16). Table 5 shows the corresponding finite sample behaviour for this realised kernel. In addition to the scaled variance, we also report the scaled squared bias

$$\frac{\left[n^{1/4}\mathrm{E}\left\{\bar{K}(X_\delta)-\int_0^t\sigma_u^2\mathrm{d}u\right\}\right]^2}{\omega\left(\int_0^t\sigma_u^4\mathrm{d}u\right)^{3/4}}=4n^{5/2}\xi^3\left\{1-k\left(\frac{1}{c\xi n^{1/2}}\right)\right\}^2.$$

The Table shows the bias is small when ξ^2 is large and so does not create a distortion for the inference procedure for this realised kernel. However, for small ω^2 the bias dramatically swamps the variance and so inference would be significantly affected.

n	$\xi^2 = 0.01$		$\xi^2 = 0.001$		$\xi^2 = 0.0001$		$\xi^2 = 0.01$		$\xi^2 = 0.001$		$\xi^2 = 0.0001$	
	Var	Bias^2	Var	Bias^2	Var	Bias^2	Var	Bias^2	Var	Bias^2	Var	Bias^2
	5-th order kernel						6-th order kernel					
256	9.97	5.28	8.34	33.1	13.8	4.19	11.9	0.10	13.1	1.33	13.8	4.19
$1,024$	10.1	3.47	9.74	45.4	10.7	33.5	12.0	0.02	12.3	1.22	13.1	15.8
$4,096$	10.2	1.97	10.0	34.9	9.90	189	12.0	0.00	12.0	0.48	11.5	43.0
$16,384$	10.2	1.05	10.1	31.0	9.88	461	12.1	0.00	12.0	0.11	12.1	10.4
$65,536$	10.2	0.57	10.2	17.2	10.1	322	12.1	0.00	12.0	0.02	12.0	3.41
$262,144$	10.2	0.29	10.2	9.07	10.2	254	12.1	0.00	12.0	0.00	12.0	0.71
∞	10.2	0.00	10.2	0.00	10.2	0.00	12.1	0.00	12.1	0.00	12.1	0.00
	7-th order kernel						8-th order kernel					
256	13.6	0.00	14.7	0.27	13.8	4.19	15.0	0.00	15.9	0.05	13.8	4.19
$1,024$	13.8	0.00	13.8	0.09	15.5	6.88	15.5	0.00	15.1	0.00	17.4	2.80
$4,096$	13.9	0.00	13.7	0.01	12.7	8.80	15.6	0.00	15.3	0.00	13.8	1.66
$16,384$	13.9	0.00	13.8	0.00	13.7	0.55	15.7	0.00	15.6	0.00	15.1	0.02
$65,536$	13.9	0.00	13.9	0.00	13.7	0.05	15.7	0.00	15.6	0.00	15.6	0.00
$262,144$	13.9	0.00	13.9	0.00	13.9	0.00	15.7	0.00	15.7	0.00	15.6	0.00
∞	13.9	0.00	13.9	0.00	13.9	0.00	15.7	0.00	15.7	0.00	15.7	0.00

Table 5: *Finite sample value of* $\mathrm{Var}\{n^{1/4}K(X_\delta)\}\omega^{-1}\left(\int_0^t \sigma_u^4 \mathrm{d}u\right)^{-3/4}$ *listed against* n *and scaled squared bias for various order cases when* $\rho = 1$. *In the* $n = 256$ *case, when* ξ^2 *is very small* H *is selected to be zero and so the realised kernel becomes the RV.*

4.7 Small-ω^2 Asymptotic Analysis

Given that ω^2 is estimated to be small relative to the integrated variance, $\int_0^t \sigma_u^2 \mathrm{d}u$, it becomes interesting to analyse the realised kernels with an asymptotic scheme that takes ω^2 to be local to zero. Specifically, we consider the situation where $\omega^2 = \omega_0^2 n^{-\alpha}$ for some $0 \leq \alpha < 1$,[3] and define $\xi_0^2 = \omega_0^2 / \sqrt{t \int_0^t \sigma_u^4 \mathrm{d}u}$.[4] In this situation the asymptotic variance is

$$4t\int_0^t \sigma_u^4 \mathrm{d}u \left\{ \frac{H}{n} k_\bullet^{0,0} + n^{-\alpha} \frac{2\xi_0^2 \rho k_\bullet^{1,1}}{H} + n^{-2\alpha} \xi_0^4 n \frac{k_\bullet^{2,2}}{H^3} + n^{-2\alpha} \frac{k_\bullet^{1,1}}{Hm} \xi_0^4 \right\},$$

when $k'(0)^2 + k'(1)^2 = 0$.

To determine the optimal rate for H, we set $H = c_0 n^\gamma$ and find the four terms of the variance to be $O(n^{\gamma-1})$, $O(n^{-\alpha-\gamma})$, $O(n^{-2\alpha+1-3\gamma})$, and $O(n^{-2\alpha-\gamma}m^{-1})$, respectively. The first three terms are all $O(n^{-\frac{1+\alpha}{2}})$ when $\gamma^* = (1-\alpha)/2$, which is the optimal rate for H. With this rate for H the last term of the asymptotic variance is of order $O(n^{-\frac{1+\alpha}{2}-\alpha}m^{-1})$, which is lower than that of the other terms, even if m is constant. So jittering ($m \to \infty$) is not needed under this form of asymptotics.

[3]When $\alpha > 1$ the asymptotic analysis is essentially the same as $\omega^2 = 0$ – the case without noise. Note that as $\alpha < 1$ then $[U] = \infty$ so $U \notin \mathcal{SM}$.

[4]A similiar local to zero asymptotics was used by Kalnina and Linton (2006) in the context of the two scale estimator.

The rate of convergence is now slightly faster, for setting $H = c\xi_0 n^{\frac{1-\alpha}{2}}$ we have

$$n^{\frac{1+\alpha}{4}}\left\{K(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \xrightarrow{L_X} MN\left\{0, \omega_0\left(4t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{\frac{3}{4}}\left(ck_\bullet^{0,0} + c^{-1}2\rho k_\bullet^{1,1} + c^{-3}k_\bullet^{2,2}\right)\right\}, \quad (18)$$

which implies the relative efficiency of different kernels in the class is not effected by changing the asymptotic experiment to $\omega^2 = \omega_0^2 n^{-\alpha}$. From (18) it follows that the optimal c is the same as under the fixed-ω^2 asymptotics. Our estimator of $t\int_0^t \sigma_u^4 \mathrm{d}u$ continues to be consistent, whereas the estimator $\hat{\omega}^2 = \gamma_0(X_\delta)/(2n)$ will decay to zero at the same rate as ω^2. Our estimator $\hat{\xi}^2 = \hat{\omega}^2 \Big/ \widehat{IQ}_{\delta,S}$ will be such that

$$1 - \hat{\xi}^2 / (\xi_0^2 n^{-\alpha}) = o_p(1).$$

It now follows that our data dependent selection of the lag length, $\hat{H} = c\hat{\xi}n^{1/2} \simeq c\xi_0 n^{(1-\alpha)/2}$, is robust, in the the sense that it consistently selects the optimal rate for H under both the fixed-ω^2 and small-ω^2 asymptotic schemes.

Finally, our plug-in estimate of the asymptotic variance is

$$\begin{aligned}
\hat{\varpi} &= \frac{\hat{H}}{n} 4\widehat{IQ}_{\delta,S} k_\bullet^{0,0} + 8\tfrac{1}{\hat{H}}\hat{\omega}^2 K(X_\delta) k_\bullet^{1,1} + 4\tfrac{n}{\hat{H}^3}\hat{\omega}^4 k_\bullet^{2,2} \\
&= n^{-\frac{1+\alpha}{2}}\omega_0 4t\int_0^t \sigma_u^4 \mathrm{d}u \left(c\xi_0 k_\bullet^{0,0} + c^{-1}2\rho\xi_0 k_\bullet^{1,1} + c^{-3}\xi_0 k_\bullet^{2,2}\right) + o_p(n^{-\frac{1+\alpha}{2}}) \\
&= n^{-\frac{1+\alpha}{2}}\omega_0\left(4t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{\frac{3}{4}}\left(ck_\bullet^{0,0} + c^{-1}2\rho k_\bullet^{1,1} + c^{-3}k_\bullet^{2,2}\right) + o_p(n^{-\frac{1+\alpha}{2}}).
\end{aligned}$$

So that $n^{(1+\alpha)/2}\hat{\varpi}$ is consistent for the appropriate asymptotic variance given in (18).

If we set the kernel weight for the first-order autocovariance to be $k(H^{-1})$ rather than one (i.e. removing the flat-top restriction), the bias due to noise is

$$\omega_0^2 n^{1-\alpha}\left\{1 - k(H^{-1})\right\} = \omega_0^2 n^{1-\alpha}\left\{k'(0)H^{-1} + \tfrac{1}{2}k''(0)H^{-2} + O(H^{-3})\right\}.$$

When $k'(0) = 0$ and we use the optimal rate, $H = c\xi_0 n^{\frac{1-\alpha}{2}}$, the bias is $\frac{1}{2}\omega_0^2\xi_0 + o(1)$. So the bias does not vanish under this scheme either, unless we impose the flat-topness, or some other remedy.

When the kernel is "kinked", in the sense that $k'(0)^2 + k'(1)^2 \neq 0$, one can show $H \propto n^{\frac{2}{3}(1-\alpha)}$ is optimal and that the best rate of convergence is $n^{\frac{1+2\alpha}{6}}$. This reveals that kinked kernels are somewhat less inefficient when ω^2 is local-to-zero. For $\alpha = 0$ we recall that the fastest rates of convergence are 0.50 and 0.333 for smooth and kinked kernels, respectively. The difference between the two rates is smaller when $\alpha > 0$, e.g. with $\alpha = \frac{5}{6}$ the corresponding convergence rates are about 0.458 and 0.444.

5 Related issues

Some of our limit theories depend upon integrated quarticity $\int_0^t \sigma_u^4 \mathrm{d}u$ and the noise's variance ω^2. We now discuss estimators of these quantities.

5.1 Estimating ω^2

To estimate ω^2 Oomen (2005) suggested using the unbiased $\tilde{\omega}^2 = -\left\{\gamma_1(X_\delta) + \gamma_{-1}(X_\delta)\right\}/2n$, while, for example, Bandi and Russell (2005a) suggest $\hat{\omega}^2 = \gamma_0(X_\delta)/2n$ which has a bias of $\int_0^t \sigma_u^2 \mathrm{d}u/2n$. Both estimators have their shortcoming, as $\tilde{\omega}^2$ may be negative and $\hat{\omega}^2$ can be severely biased because $\int_0^t \sigma_u^2 \mathrm{d}u/2n$ may be large relative to ω^2. Using Theorem 1 we have that

$$\operatorname{Var}\left\{n^{1/2}\left(\tilde{\omega}^2 - \omega^2\right)\right\} = \omega^4\left(5 + \lambda^2\right), \quad \operatorname{Var}\left\{n^{1/2}\left(\hat{\omega}^2 - \omega^2\right)\right\} = \omega^4\left(1 + \lambda^2\right).$$

In the Gaussian case $\lambda^2 = 2$, and so $\tilde{\omega}^2$ and $\hat{\omega}^2$ have variances which are around 3.5 and 1.5 times that of the ML estimator in the parametric case given in (7). Although it is possible to derive a kernel style estimator to estimate ω^2 efficiently, we resist the temptation to do so here as the statistical gains are minor.

Instead we propose a simple bias correction of $\hat{\omega}^2$ that is guaranteed to produce a non-negative estimate. We have that

$$\log \mathrm{E}\left(\hat{\omega}^2\right) = \log \omega^2 + \log\left\{1 + \int_0^t \sigma_u^2 \mathrm{d}u \bigg/ \left(2n\omega^2\right)\right\}.$$

Substituting the consistent estimators, $K(X_\delta)$ and $\gamma_0(X_\delta)/2n$, for $\int_0^t \sigma_u^2 \mathrm{d}u$ and ω^2, respectively, yields our preferred estimator

$$\breve{\omega}^2 = \exp\left\{\log \hat{\omega}^2 - K(X_\delta)/\gamma_0(X_\delta)\right\}. \tag{19}$$

Note that $K(X_\delta)/\gamma_0(X_\delta)$ is an estimate of the relative bias of $\hat{\omega}^2$, which vanishes as $n \to \infty$, so that $\breve{\omega}^2 - \hat{\omega}^2 \xrightarrow{p} 0$. Throughout our simulations and empirical work we will use (19) to estimate ω^2.

5.2 Estimation of integrated quarticity, $\int_0^t \sigma_u^4 \mathrm{d}u$

Estimating integrated quarticity reasonably efficiently is a tougher problem than estimating $[Y]$. We do not know of any existing research which has solved this problem in the context with noise. Define the subsampled squared returns, for some $\tilde{\delta} > 0$, $x_{j,\cdot}^2 = \frac{1}{S}\sum_{s=0}^{S-1}\left(X_{\tilde{\delta}\left(j+\frac{s}{S}\right)} - X_{\tilde{\delta}\left(j-1+\frac{s}{S}\right)}\right)^2$, $j = 1, 2, ..., \tilde{n}$, where $\tilde{n} = \lfloor t/\tilde{\delta} \rfloor$. This allows us to define a bipower variation type estimator of integrated quarticity

$$\left\{X_{\tilde{\delta}}, \omega^2; S\right\}^{[2,2]} = \frac{t}{\tilde{n}}\sum_{j=1}^{\tilde{n}} \tilde{\delta}^{-2}\left(x_{j,\cdot}^2 - 2\omega^2\right)\left(x_{j-2,\cdot}^2 - 2\omega^2\right).$$

The no noise case of this statistic was introduced by Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen and Shephard (2006) and studied in depth by Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006). See also Mykland (2006).

Detailed calculations show that when $\tilde{\delta}$ is small and S is large then the conditional variance of $\left\{X_{\tilde{\delta}}, \omega^2; S\right\}^{[2,2]}$ is approximately $72\omega^8\tilde{n}^3/S^2$, so that $\tilde{\delta}^{3/2}S \to \infty$ leads to consistency.[5] An interesting research problem is how to make this type of estimator more efficient by using kernel type estimators. For now we use moderate values of $\tilde{n}$ and high values of S in our Monte Carlos and empirical work.

The finite sample performance of our estimator can be greatly improved by using the inequality $t\int_0^t \sigma_u^4 \mathrm{d}u \geq \left(\int_0^t \sigma_u^2 \mathrm{d}u\right)^2$. This is useful as we have a very efficient estimator of $\int_0^t \sigma_u^2 \mathrm{d}u$. Thus our preferred way of estimating $t\int_0^t \sigma_u^4 \mathrm{d}u$ is

$$\widehat{IQ}_{\delta,S} = \max\left[\left\{K(X_\delta)\right\}^2, \left\{X_{\tilde{\delta}}, \breve{\omega}^2; S\right\}^{[2,2]}\right].$$

5.3 Effect of endogenous and stochastically spaced data

So far our analysis has been based on measuring prices at regularly spaced intervals of length δ. In some ways it is more natural to work with returns measured in tick time and so it would be attractive if we could extend the above theory to cover stochastically spaced data. The convergence result inside QV is known to hold under very wide conditions that allow the spacing to be stochastic and endogenous. This is spelt out in, for example, Protter (2004, pp. 66-77) and Jacod and Shiryaev (2003, p. 51). It is important, likewise, to be able to derive central limit theorems for stochastically spaced data without assuming the times of measurement are independent of the underlying $\mathcal{BSM}$, which is the assumption used by Phillips and Yu (2006), Mykland and Zhang (2006) and Barndorff-Nielsen and Shephard (2005). This is emphasised by Renault and Werker (2005) in both their theoretical and empirical work.

Let $Y \in \mathcal{BSM}$ and assume we have measurements at times $t_j = T_{\delta j}$, $j = 1, 2, ..., n$, where $0 = t_0 < t_1 < ... < t_n = T_1$ and where T is a stochastic process of the form $T_t = \int_0^t \tau_u^2 \mathrm{d}u$, with τ having strictly positive, càdlàg sample paths. Then we can construct a new process $Z_t = Y_{T_t}$, so at the measurement times $Z_{\delta j} = Y_{T_{\delta j}}$ $j = 1, 2, ..., n$. Performing the analysis on observations of Z made at equally spaced times then allows one to analyse irregularly spaced data on Y. The following

[5] Let $\varepsilon_j = \frac{1}{S}\sum_{s=0}^{S-1}\left[\left(U_{\tilde{\delta}(j+\frac{s}{S})} - U_{\tilde{\delta}(j-1+\frac{s}{S})}\right)^2 - 2\omega^2 + 2\left(U_{\tilde{\delta}(j+\frac{s}{S})} - U_{\tilde{\delta}(j-1+\frac{s}{S})}\right)\left(Y_{\tilde{\delta}(j+\frac{s}{S})} - Y_{\tilde{\delta}(j-1+\frac{s}{S})}\right)\right]$ and $R = \left[\sum_{j=1}^n \left(x_{j,.}^2 - 2\omega^2\right)\left(x_{j-2,.}^2 - 2\omega^2\right)\right] - \sum_{j=1}^n y_{j,.}^2 y_{j-2,.}^2 = \sum_{j=1}^n y_{j,.}^2 \varepsilon_{j-2} + \sum_{j=1}^n \varepsilon_j y_{j-2,.}^2 + \sum_{j=1}^n \varepsilon_j\varepsilon_{j-2}$. Then $R \simeq \sum_{j=1}^n \varepsilon_j\left(y_{j-2,.}^2 + y_{j+2,.}^2\right) + \sum_{j=1}^n \varepsilon_j\varepsilon_{j-2}$. Now $\mathrm{Var}\left(\sum_{j=1}^n \varepsilon_j\left(y_{j-2,.}^2 + y_{j+2,.}^2\right)|Y\right) \simeq \frac{12\omega^4}{S}\sum_{j=1}^n\left(y_{j-2,.}^2 + y_{j+2,.}^2\right)^2 = O(n^{-1}S^{-1})$. So

$$\begin{aligned}
\mathrm{Var}(R|Y) &\simeq \mathrm{Var}\left(\sum_{j=1}^n \varepsilon_j\varepsilon_{j-2}|Y\right) = \sum_{j=1}^n \mathrm{Var}\left(\varepsilon_j\varepsilon_{j-2}|Y\right) + 2n\mathrm{Cov}\left(\varepsilon_j\varepsilon_{j-2}, \varepsilon_{j-1}\varepsilon_{j-3}|Y\right) \\
&= \sum_{j=1}^n \mathrm{E}\left(\varepsilon_j^2|Y\right)\mathrm{E}\left(\varepsilon_{j-2}^2|Y\right) + 2n\mathrm{Cov}\left(\varepsilon_j\varepsilon_{j-2}, \varepsilon_{j-1}\varepsilon_{j-3}|Y\right) \\
&\simeq \sum_{j=1}^n\left\{\frac{8\omega^4}{S} + \frac{8\omega^2}{S}\left(y_j^2\right)_.\right\}\left\{\frac{8\omega^4}{S} + \frac{8\omega^2}{S}\left(y_{j-2}^2\right)_.\right\} + \frac{n}{S^2}2\left(2\omega^2\right)^2 + ... = \frac{72\omega^8 n}{S^2} + O(S^{-2}).
\end{aligned}$$

argument shows that $Z \in \mathcal{BSM}$ with spot volatility $\sigma_{T_t}\tau_t$ and so the analysis is straightforward. In particular, the feasible CLT is implemented by recording data every 5 transactions, say, but then analysing it as if the spacing had been equidistant.

Write $Z = Y \circ T$ and $S_t = \int_0^t \sigma_u^2 \mathrm{d}u$. We assume that Y and T are adapted to a common filtration $\mathcal{F}_t$, which includes the history of the paths of T_u and $Y \circ T_u$ for $0 \leq u \leq t$. This assumption implies that σ_{u-} is in $\mathcal{F}_t$ for $0 \leq u \leq T_t$. Recall the key result (e.g. Revuz and Yor (1999, p. 181)) $[Z] = S \circ T$, while $Z \in \mathcal{M}_{loc}$. The following proposition shows that $[Z]$ is absolutely continuous and implies by the martingale representation theorem that Z is a stochastic volatility process with spot volatility of $\sigma_{T_t}\tau_t$.

Proposition 1 *Let $\upsilon_t = \sigma_{T_t}\tau_t$ and*

$$\Upsilon_t = \int_0^t \upsilon_u^2 \mathrm{d}u. \tag{20}$$

Then υ is a càdlàg process and $\Upsilon = S \circ T$.

The implication of this for kernels is that we can write $Z_t = \int_0^t a_{T_u}\tau_u \mathrm{d}u + \int_0^t \sigma_{T_u}\tau_u \mathrm{d}B_u^{\#}$, where $B^{\#}$ is Brownian motion. Hence if we define a tick version of the kernel estimator

$$\gamma_h(Z_n)_t = \sum_{j=1}^{\lfloor t/\delta \rfloor} \left(Y \circ T_{\delta j} - Y \circ T_{\delta(j-1)}\right)\left(Y \circ T_{\delta(j-h)} - Y \circ T_{\delta(j-h-1)}\right),$$

$$K(Z_n)_t = \gamma_0(Z_n)_t + \sum_{h=1}^{H} k\left(\tfrac{h-1}{H}\right)\left\{\gamma_h(Z_n)_t + \gamma_{-h}(Z_n)_t\right\},$$

then the theory for this process follows from the previous results. Thus using the symmetric kernel allows consistent inference on $[Z]_t = [Y]_{T_t}$.

5.4 Effect of serial dependence

So far we have assumed that $U \in \mathcal{WN}$. Now we will relax the assumption $U_s \perp\!\!\!\perp U_r$ and allow U to be serial dependent to the extend that

$$\sum\nolimits_{h=1}^{H} a_{h,H} U_{h\delta} = O_p(1), \qquad \text{for any } \sum\nolimits_{h=1}^{H} a_{h,H}^2 = O(1). \tag{21}$$

Proposition 2 *Suppose (21) holds. If $k'(0) = k'(1) = 0$, then*

$$K(U_\delta) = -2H^{-2}\sum\nolimits_{h=1}^{H} k''\left(\tfrac{h}{H}\right)\sum_{i=1}^{n} U_{i\delta}U_{(i-h)\delta} + O_p(nH^{-3}) + O_p(H^{-1/2}).$$

With dependent noise, it is no longer true that $H^{-1/2}\sum_{h=1}^{H} k''\left(\frac{h}{H}\right) n^{-1/2}\sum_{i=1}^{n} U_{i\delta}U_{(i-h)\delta} \xrightarrow{L} N(0, k_\bullet^{2,2}\omega^4)$. However, despite the serial dependence, this term may be $O_p(1)$, in which case the noise will not have any impact on the asymptotic distribution of $K(X_\delta)$, if we use an inefficient rate for H, such as $H \propto n^{2/3}$.

Proposition 3 *We assume $k''(0) = 0$, that $|k'''(0)| < \infty$, and that $U_{j\delta}$, $j = \ldots, 0, 1, 2, \ldots$ is an AR(1) process with persistence parameter φ, $|\varphi| < 1$, then*

$$(nH)^{-1/2} \sum_{h=1}^{H} k''\left(\tfrac{h}{H}\right) \sum_{i=1}^{n} U_{i\delta} U_{(i-h)\delta} \xrightarrow{L} N\left(0, \omega^4 \tfrac{1+\varphi^2}{1-\varphi^2} k_{\bullet}^{2,2}\right).$$

This means that

$$K(U_\delta) = O_p\left(\tfrac{n^{1/2}}{H^{3/2}}\right) + O_p\left(\tfrac{n}{H^3}\right) + O_p(m^{-1/2} H^{-1/2}).$$

So if $H \propto n^{1/2}$ then $K(X_\delta) = O_p\left(n^{-1/4}\right)$.

If we assume that $Y \perp\!\!\!\perp U$ then temporal dependence in U makes no difference to the asymptotic behaviour of $\gamma_h(U_\delta, Y_\delta)$ as $\delta \downarrow 0$ for the limit behaviour is driven by the local martingale difference behaviour of the increments of the Y process, we simply need to redefine $\omega^2 = \lim_{n\to\infty} \mathrm{Var}(n^{-1/2} \sum_{j=1}^{n} U_{j\delta})$.

The above results mean that if we set $H \propto n^{1/2}$, then $K(U_\delta) = O_p(n^{-1/4})$ and so the rate of convergence of the realised kernel is not changed by this form of serial dependence, but the asymptotic distribution is altered.

5.5 Endogenous noise

One of our key assumptions has been that $Y \perp\!\!\!\perp U$, that is the noise can be regarded as an exogenous process. Hence it is interesting to ask if our realised kernels continue to be consistent when U is endogenous. We do this under a simple linear model of endogeneity

$$U_{\delta i} = \sum_{h=0}^{\bar{H}} \beta_h \left(Y_{\delta(i-h)} - Y_{\delta(i-1-h)}\right) + \bar{U}_{\delta i},$$

where $Y \perp\!\!\!\perp \bar{U}$ and for simplicity we assume that $\bar{U} \in \mathcal{WN}$. Now

$$\gamma_h(Y_\delta, U_\delta) = \sum_{j=0}^{\bar{H}} \beta_j \gamma_{h+j}(Y_\delta) - \sum_{j=0}^{\bar{H}} \beta_j \gamma_{h+j+1}(Y_\delta) + \gamma_h(Y_\delta, \bar{U}_\delta).$$

Hence our asymptotic methods for studying the distribution of realised kernels under exogenous noise can be used to study the impact of endogenous noise on realised kernels through the limit theory we developed for $\gamma_h(Y_\delta)$ and $\gamma_h(Y_\delta, \bar{U}_\delta)$. In particular

$$\gamma_h(Y_\delta, U_\delta) - \gamma_h(Y_\delta, \bar{U}_\delta) = \begin{cases} \beta_0[Y] + O_p(n^{-1/2}), & h = 0, \\ -\beta_0[Y] + O_p(n^{-1/2}), & h = -1, \\ O_p(n^{-1/2}), & |h| \neq 1. \end{cases}$$

Hence realised kernels will be robust to this type of endogenous noise. An alternative approach to dealing with endogenous noise has been independently proposed by Kalnina and Linton (2006) using multiscale estimators.

5.6 Jumps

In this section we study the impact of jumps. The observed price process is now

$$R = X + D = Y + D + U,$$

where D_s is a jump process which has jumped once at time $\tau \times t$ where $\tau \in (0,1)$, and that $D \perp\!\!\!\perp Y \perp\!\!\!\perp U$. We write (intraday) returns as $r_j = y_j + d_j + u_j$, where, for example, $d_j = D_{j\delta} - D_{(j-1)\delta}$. Thus

$$K(R_\delta) - K(X_\delta) - \sum_{j=1}^{n} d_j^2 = L_\delta + M_\delta,$$

where $L_\delta = d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h y_{\lfloor n\tau \rfloor + h}$, $M_\delta = d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h u_{\lfloor n\tau \rfloor + h}$, and we as usual, set $w_0 = 1$ and $w_h = k\left(\frac{|h|-1}{H}\right)$ for $h = \pm 1, \ldots, \pm H$. We have established the asymptotic properties of $K(X_\delta) - \int_0^t \sigma_s^2 \mathrm{d}s$, so the asymptotic properties of $K(R_\delta) - \int_0^t \sigma_s^2 \mathrm{d}s - \sum_{j=1}^{n} d_j^2$ hinges on those of $L_\delta + M_\delta$.

Conditioning on d and σ, we have with $z_j = y_j/\sigma_{j\delta}$ that

$$L_\delta \simeq \sigma_{\tau t} d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h z_{\lfloor n\tau \rfloor + h} \sim N\left(0, \frac{2H}{n} \sigma_{\tau t}^2 d_{\lfloor n\tau \rfloor}^2 \frac{1}{H} \sum_{h=-H}^{H} w_h^2\right).$$

So if $H = c\xi n^{1/2}$ then conditionally

$$n^{1/4} \left(d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h y_{\lfloor n\tau \rfloor + h}\right) \xrightarrow{L} MN\left(0, 2c\xi k_\bullet^{0,0} \sigma_{\tau t}^2 d_{\lfloor n\tau \rfloor}^2\right).$$

If $H = c(\xi^2 n)^{2/3}$ then conditionally

$$n^{1/6} \left(d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h y_{\lfloor n\tau \rfloor + h}\right) \xrightarrow{L} MN\left(0, 2c\xi^{4/3} k_\bullet^{0,0} \sigma_{\tau t}^2 d_{\lfloor n\tau \rfloor}^2\right).$$

What happens with market microstructure effects? We need to look at

$$\begin{aligned} M_\delta &= d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h u_{\lfloor n\tau \rfloor + h} = d_{\lfloor n\tau \rfloor} \sum_{h=-H}^{H} w_h \left(U_{(\lfloor n\tau \rfloor + h)\delta} - U_{(\lfloor n\tau \rfloor + h - 1)\delta}\right) \\ &= d_{\lfloor n\tau \rfloor} \sum_{h=-H-1}^{H} \left\{k\left(\tfrac{h-1}{H}\right) - k\left(\tfrac{h}{H}\right)\right\} U_{(\lfloor n\tau \rfloor + h)\delta}. \end{aligned}$$

Then

$$\mathrm{Var}\left(\sqrt{H} M_\delta \,\middle|\, d_{\lfloor n\tau \rfloor}\right) = H\omega^2 d_{\lfloor n\tau \rfloor}^2 \sum_{h=-H-1}^{H} \left\{k\left(\tfrac{h-1}{H}\right) - k\left(\tfrac{h}{H}\right)\right\}^2 \to \omega^2 d_{\lfloor n\tau \rfloor}^2 2k_\bullet^{1,1}.$$

This suggests the realised kernel is consistent for the quadratic variation, $[Y]_t$, at the same rate of convergence as before. The asymptotic distribution is, of course, not easy to calculate even in the pure $\mathcal{BSM}$ plus jump case (e.g. Jacod (2006)). The extension to allow for finite activity jump processes is straightforward.

6 Simulation study

6.1 Goal of the study

In this Section we report simulation results which assess the accuracy of the feasible asymptotic approximation for the realised kernel. A much more thorough analysis is provided in a Web Appendix to this paper available from `http://www.hha.dk/~alunde/bnhls/bnhls.htm`.

Before we turn our attention to feasible asymptotic distributions, we note the Web Appendix also reports on the accuracy of $K(X_\delta)$ as an estimator of $\int_0^t \sigma_u^2 \mathrm{d}u$ and $\widehat{IQ}_{\delta,S}$ as an estimator of $t\int_0^t \sigma_u^4 \mathrm{d}u$. The raw estimator $K(X_\delta)$ may be negative, in which case we always truncate it at zero (the same technique is used for ML estimators of course). The Web Appendix shows this occurrence is extremely rare. In our simulations we generated millions of artificial samples and less than 25 of them resulted in negative value for $K(X_\delta)$, using the Tukey-Hanning$_2$ weights.

In this short section our focus will be assessing the infeasible and feasible central limit theories for $K(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u$. Throughout we simulate over the time interval $[0,1]$. We recall the asymptotic variance of $K(X_\delta)$ is given in (12) which we write as ϖ here. This allows us to compute the asymptotic pivot

$$Z_n = \frac{K(X_\delta) - \int_0^1 \sigma_u^2 \mathrm{d}u}{\sqrt{\varpi}} \xrightarrow{L} N(0,1).$$

An alternative is to use the delta method and base the asymptotic analysis on (e.g. Barndorff-Nielsen and Shephard (2002) and Goncalves and Meddahi (2004))

$$Z_n^{\log} = \frac{\log\{K(X_\delta)+d\} - \log\left\{\int_0^1 \sigma_u^2 \mathrm{d}u + d\right\}}{\sqrt{\varpi}/\{K(X_\delta)+d\}} \xrightarrow{L} N(0,1).$$

The presence of $d \geq 0$ allows for the possibility that $K(X_\delta)$ may be truncated to be exactly zero. By selecting $d > 0$ we have the property that $K(X_\delta)+d$ is not negative.

In the infeasible case our simple rule-of-thumb for the choice of H is $H^* = c^*\omega\sqrt{n/[Y]_t}$, which immediately gives us ϖ. In practice this is less interesting than the feasible version, which puts $\hat{H} = c^*\hat{\omega}\sqrt{n/\gamma_0(X_{\bar{\delta}})}$, where $\gamma_0(X_{\bar{\delta}})$ is the realised variance estimator based on low frequency data, such as 10 minute returns, which should not be too sensitive to market frictions. Having selected H, in the feasible case we can then compute $\hat{\varpi}$ by plugging $K(X_\delta)$, $\check{\omega}^2$, and $\widehat{IQ}_{\delta,S}$ into our expression for ϖ, replacing $\int_0^1 \sigma_u^2 \mathrm{d}u$, ω^2, and $\int_0^1 \sigma_u^4 \mathrm{d}u$ respectively. Monte Carlo results reported in the Appendix suggest taking $S = \sqrt{n}$ in computing $\widehat{IQ}_{\delta,S}$.

6.2 Simulation design

We focus on the Tukey-Hanning$_2$ kernel, because it is near efficient and does not require too many intraday returns outside the $[0,t]$ interval. We simulate data for the unit interval $[0,1]$ and normalize one second to be $1/23400$, so that $[0,1]$ is thought to span 6.5 hours. The X process is generated using an Euler scheme based on $N = 23,400$ of intervals. We then construct sparsely sampled returns $X_{i/n} - X_{(i-1)/n}$, based on sample sizes n. In our Monte Carlo designs n takes on the values 195, 390, 780, 1,560, 4,680, 7,800, 11,700 and 23,400. The case of 1 minute returns is when $n = 390$.

We consider two stochastic volatility models that are commonly used in this literature, see e.g. Huang and Tauchen (2005) and Goncalves and Meddahi (2004). The first is a one-factor model (**SV1F**):

$$\mathrm{d}Y_t = \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad \sigma_t = \exp(\beta_0 + \beta_1 \tau_t), \quad \mathrm{d}\tau_t = \alpha \tau_t \mathrm{d}t + \mathrm{d}B_t, \quad \mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_t) = \varphi.$$

Here φ is a leverage parameter. To make the results comparable to our constant volatility simulations reported in our Appendix we impose that $\mathrm{E}(\sigma_t^2) = 1$ by setting $\beta_0 = \beta_1^2/(2\alpha)$. We utilize the fact that the stationary distribution $\tau_t \sim N\left(0, \frac{-1}{2\alpha}\right)$ to restart the process each day. In these experiments we set $\mu = 0.03$, $\beta_1 = 0.125$, $\alpha = -0.025$ and $\varphi = -0.3$. The variance of σ is comparable to the empirical results found in e.g. Hansen and Lunde (2005).

We also consider a two-factor SV model (**SV2F**):[6]

$$\mathrm{d}Y_t = \mu \mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad \sigma_t = \text{s-}\exp(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}), \quad \mathrm{d}\tau_{1t} = \alpha_1 \tau_{1t} \mathrm{d}t + \mathrm{d}B_{1t},$$
$$\mathrm{d}\tau_{2t} = \alpha_2 \tau_{2t} \mathrm{d}t + (1 + \phi \tau_{2t}) \mathrm{d}B_{2t}, \quad \mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_{1t}) = \varphi_1, \quad \mathrm{corr}(\mathrm{d}W_t, \mathrm{d}B_{2t}) = \varphi_2.$$

We adopt the configuration from Huang and Tauchen (2005) and set $\mu = 0.03$, $\beta_0 = -1.2$, $\beta_1 = 0.04$, $\beta_2 = 1.5$, $\alpha_1 = -0.00137$, $\alpha_2 = -1.386$, $\phi = 0.25$, $\varphi_1 = \varphi_2 = -0.3$. At the start of each interval we initialize the two factors, by drawing the persistent factor from its unconditional distribution, $\tau_{10} \sim N(0, \frac{-1}{2\alpha_1})$, while the strongly mean-reverting factor is simply started at zero, $\tau_{20} = 0$. An important difference between the two volatility models is the extend of heteroskedasticity, because the variation in ρ is much larger for the 2-factor model than for the 1-factor model.

Finally, the market microstructure effects are modelled through ξ^2. This is varied over 0.0001, 0.001 and 0.01, the latter being regarded as a very large effect indeed. These values are taken from the detailed study of Hansen and Lunde (2006).

6.3 Results

Tables 6 shows the Monte Carlo results for the infeasible asymptotic theory for Z_n, knowing a priori the value of ϖ. We can see from the Tables that the results are rather good, although the asymptotics are slightly underestimating the mass of the distribution in the tails. The mean and standard deviations of Z_n show that the Z-statistic is slightly overdispersed.

[6] The function s-exp$[x]$ is defined in the web appendix `www.hha.dk/~alunde/bnhls/bnhls.htm`. This appendix also give a detailed description of our discretisation scheme for the models.

Table 6: **SV1F**: Finite sample properties of Z_n

ξ^2 *= 0.01, number of reps. = 100,000*

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	9.00	0.002	1.052	0.08	1.54	4.17	93.12	95.67	98.43
390	12.0	0.001	1.036	0.17	1.78	4.38	93.50	96.12	98.74
780	17.0	-0.002	1.025	0.18	1.88	4.48	93.84	96.44	98.95
1560	23.0	-0.003	1.015	0.26	1.94	4.49	94.18	96.70	99.09
4680	40.0	-0.006	1.009	0.31	2.12	4.68	94.49	96.99	99.20
7800	51.0	-0.005	1.009	0.34	2.23	4.70	94.57	97.05	99.26
11700	63.0	-0.005	1.010	0.36	2.22	4.76	94.57	97.06	99.28
23400	88.0	-0.002	1.007	0.40	2.30	4.84	94.66	97.19	99.34

ξ^2 *= 0.001, number of reps. = 100,000*

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	3.00	-0.002	1.015	0.15	1.67	4.20	93.94	96.45	98.91
390	4.00	-0.004	1.007	0.21	1.81	4.27	94.22	96.76	99.10
780	6.00	-0.005	1.005	0.27	1.98	4.43	94.40	96.98	99.23
1560	8.00	-0.005	1.002	0.33	2.10	4.59	94.54	97.09	99.28
4680	13.0	-0.005	1.003	0.39	2.32	4.82	94.72	97.27	99.40
7800	17.0	-0.003	1.005	0.39	2.31	4.80	94.75	97.21	99.37
11700	20.0	-0.004	1.004	0.43	2.38	4.93	94.81	97.36	99.39
23400	28.0	-0.002	1.001	0.46	2.42	4.88	94.87	97.39	99.41

ξ^2 *= 0.0001, number of reps. = 100,000*

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	1.00	-0.003	1.003	0.19	1.72	4.15	94.28	96.75	99.06
390	2.00	-0.005	1.003	0.25	1.98	4.41	94.45	96.94	99.21
780	2.00	-0.002	0.998	0.31	2.14	4.50	94.60	97.23	99.34
1560	3.00	-0.005	1.002	0.38	2.28	4.74	94.73	97.22	99.39
4680	4.00	-0.004	1.005	0.45	2.45	5.02	94.83	97.31	99.40
7800	6.00	-0.002	1.003	0.44	2.33	4.84	94.81	97.31	99.41
11700	7.00	-0.002	1.001	0.45	2.39	4.84	94.85	97.37	99.44
23400	9.00	0.000	0.999	0.46	2.37	4.83	94.90	97.43	99.47

Summary statistics and empirical quantiles for the infeasible statistic, Z_n, that employs the asymptotic variance, ϖ. The empirical quantiles are benchmarked against those for the limit distribution, $N(0,1)$. The simulation design is the one-factor model, **SV1F**, and $K(X_\delta)$ based on Tukey-Hanning$_2$ weights.

Table 7 shows the results for the feasible asymptotic theory for $\hat{Z}_n$. This indicates that the asymptotic theory does eventually kick in but it takes very large samples for it to provide anything like a good approximation. The reason for this is clearly that it is difficult to accurately estimate the integrated quarticity. This result is familiar from the literature on realised volatility where the same phenomena is observed.

Table 7: **SV1F**: Finite sample properties of $\hat{Z}_n$

ξ^2 = 0.01, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	8.01	-0.257	1.132	3.89	7.71	10.9	98.85	99.78	100.0
390	10.6	-0.202	1.073	2.78	6.32	9.46	98.27	99.57	99.99
780	14.3	-0.164	1.036	2.07	5.41	8.50	97.82	99.34	99.97
1560	19.8	-0.136	1.015	1.60	4.67	7.65	97.38	99.10	99.94
4680	33.6	-0.104	1.002	1.22	3.97	6.84	96.98	98.78	99.88
7800	43.1	-0.091	1.000	1.11	3.72	6.58	96.73	98.67	99.83
11700	52.7	-0.082	1.000	1.02	3.62	6.41	96.52	98.54	99.82
23400	74.2	-0.067	0.997	0.87	3.39	6.20	96.44	98.41	99.78

ξ^2 = 0.001, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	5.30	-0.242	1.079	3.29	6.88	9.91	98.98	99.81	100.0
390	5.94	-0.181	1.019	2.19	5.38	8.42	98.41	99.57	99.99
780	7.03	-0.140	0.994	1.57	4.42	7.36	97.90	99.32	99.96
1560	8.79	-0.108	0.986	1.27	3.88	6.66	97.36	99.03	99.93
4680	13.7	-0.078	0.985	0.97	3.46	6.10	96.73	98.66	99.84
7800	17.1	-0.066	0.988	0.86	3.27	5.91	96.45	98.45	99.78
11700	20.7	-0.061	0.988	0.82	3.22	5.95	96.40	98.44	99.77
23400	28.8	-0.049	0.987	0.76	3.08	5.65	96.16	98.27	99.74

ξ^2 = 0.0001, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	4.73	-0.235	1.058	2.99	6.55	9.61	99.02	99.81	100.0
390	4.82	-0.172	0.993	1.85	4.88	7.84	98.49	99.63	99.99
780	4.98	-0.126	0.966	1.30	3.93	6.71	98.06	99.39	99.97
1560	5.28	-0.091	0.958	1.03	3.43	6.04	97.53	99.11	99.94
4680	6.31	-0.058	0.965	0.75	3.00	5.57	96.80	98.68	99.84
7800	7.18	-0.046	0.969	0.68	2.79	5.30	96.42	98.42	99.78
11700	8.14	-0.040	0.972	0.65	2.76	5.28	96.29	98.34	99.75
23400	10.5	-0.030	0.976	0.62	2.66	5.19	96.01	98.19	99.72

Summary statistics and empirical quantiles for the feasible statistic, $\hat{Z}_n$, that employs our estimate of the asymptotic variance, $\hat{\varpi}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0,1)$. The simulation design is the one-factor model, **SV1F**, and $K(X_\delta)$ based on Tukey-Hanning$_2$ weights.

Table 8 show the results for the log version of the feasible theory based on $\hat{Z}_n^{\log}$ using $d = 10^{-6}$. The accuracy of the asymptotic predictions does not seem to change very much with ξ^2 and is much better than in the $\hat{Z}_n$ case. For small sample sizes we note some distortions in the tails, but generally the asymptotics results provide reasonably good approximations.

Table 8: **SV1F**: Finite sample properties of $\hat{Z}_n^{\log}$

ξ^2 = 0.01, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	8.01	-0.090	0.982	0.46	2.92	5.88	96.22	98.19	99.69
390	10.6	-0.074	0.984	0.51	2.70	5.51	95.91	98.01	99.65
780	14.3	-0.064	0.982	0.43	2.54	5.29	95.75	97.92	99.60
1560	19.8	-0.055	0.980	0.42	2.40	5.09	95.67	97.82	99.59
4680	33.6	-0.045	0.982	0.44	2.40	4.98	95.60	97.80	99.53
7800	43.1	-0.040	0.985	0.44	2.40	5.01	95.53	97.77	99.52
11700	52.7	-0.035	0.988	0.46	2.39	4.97	95.36	97.67	99.54
23400	74.2	-0.029	0.989	0.44	2.42	4.98	95.44	97.71	99.50

ξ^2 = 0.001, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	5.30	-0.121	0.972	0.95	3.60	6.42	97.10	98.75	99.85
390	5.94	-0.097	0.964	0.74	3.09	5.79	96.79	98.61	99.82
780	7.03	-0.078	0.963	0.64	2.74	5.44	96.50	98.52	99.78
1560	8.79	-0.060	0.967	0.56	2.67	5.18	96.26	98.27	99.72
4680	13.7	-0.044	0.976	0.53	2.60	5.10	95.91	98.07	99.67
7800	17.1	-0.037	0.981	0.52	2.53	5.02	95.75	97.92	99.60
11700	20.7	-0.035	0.982	0.52	2.61	5.16	95.73	97.93	99.60
23400	28.8	-0.028	0.983	0.52	2.56	5.00	95.57	97.86	99.59

ξ^2 = 0.0001, number of reps. = 100,000

No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	4.73	-0.123	0.962	0.91	3.42	6.26	97.28	98.86	99.86
390	4.82	-0.098	0.947	0.70	2.86	5.53	97.07	98.77	99.85
780	4.98	-0.075	0.942	0.61	2.60	5.06	96.95	98.74	99.82
1560	5.28	-0.054	0.945	0.51	2.47	4.94	96.68	98.52	99.82
4680	6.31	-0.034	0.959	0.50	2.44	4.86	96.21	98.27	99.69
7800	7.18	-0.027	0.966	0.46	2.31	4.78	95.89	98.06	99.68
11700	8.14	-0.024	0.970	0.47	2.37	4.80	95.87	98.02	99.63
23400	10.5	-0.017	0.974	0.47	2.34	4.79	95.65	97.92	99.62

Summary statistics and empirical quantiles for the feasible statistic, $\hat{Z}_n^{\log}$, that employs our estimate of the asymptotic variance, $\hat{\varpi}$, and $d = 10^{-6}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0,1)$. The simulation design is the one-factor model, **SV1F**, model and $K(X_\delta)$ based on Tukey-Hanning$_2$ weights.

To conserve space we only present one table with results for the two-factor model. Table 9 presents the results for the feasible theory based on our preferred $\hat{Z}_n^{\log}$ statistic. For the two-factor model we note a slower convergence to the asymptotic distribution. This result is not surprising because the integrated quarticity is harder to estimate in this design. In part because $t\int_0^t \sigma_s^4 \mathrm{d}s \Big/ \left(\int_0^t \sigma_s^2 \mathrm{d}s\right)^2$ tends to be larger in the two factor model than is the case for the one factor model. This makes the inequality $\widehat{IQ} \geq \{K(X_\delta)\}^2$ less valuable for the estimation of $t\int_0^t \sigma_s^4 \mathrm{d}s$.

Table 9: **SV2F**: Finite sample properties of $\hat{Z}_n^{\log}$

ξ^2 = 0.01, number of reps. = 100,000									
No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	8.50	-0.187	1.123	0.74	5.14	9.67	94.36	96.71	99.01
390	11.4	-0.159	1.114	1.01	4.99	9.01	94.32	96.80	99.07
780	15.5	-0.140	1.101	1.02	4.63	8.20	94.49	96.83	99.05
1560	21.5	-0.119	1.085	1.04	4.20	7.74	94.62	96.98	99.11
4680	36.6	-0.096	1.067	0.94	3.90	7.17	94.83	97.13	99.22
7800	47.0	-0.083	1.059	0.87	3.71	6.87	94.71	97.21	99.31
11700	57.4	-0.076	1.053	0.90	3.62	6.75	94.79	97.19	99.35
23400	80.9	-0.065	1.043	0.86	3.46	6.38	95.03	97.32	99.37
ξ^2 = 0.001, number of reps. = 100,000									
No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	5.53	-0.254	1.209	3.14	8.10	12.3	94.70	97.00	99.06
390	6.34	-0.203	1.188	2.71	7.12	11.0	94.49	96.91	99.07
780	7.67	-0.161	1.165	2.31	6.24	9.93	94.43	96.86	99.05
1560	9.77	-0.132	1.141	1.97	5.60	9.19	94.51	96.84	99.09
4680	15.5	-0.098	1.106	1.60	4.83	8.15	94.64	97.07	99.22
7800	19.5	-0.084	1.089	1.40	4.47	7.63	94.73	97.09	99.25
11700	23.6	-0.078	1.078	1.30	4.31	7.34	94.77	97.24	99.34
23400	32.9	-0.067	1.064	1.15	4.00	7.00	94.91	97.36	99.39
ξ^2 = 0.0001, number of reps. = 100,000									
No. obs	$\bar{H}^*$	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
195	4.82	-0.256	1.215	3.24	8.21	12.4	94.74	97.02	99.07
390	4.94	-0.201	1.194	2.84	7.17	11.1	94.61	96.90	99.02
780	5.15	-0.150	1.174	2.43	6.39	9.96	94.35	96.76	99.03
1560	5.54	-0.115	1.150	2.07	5.66	9.12	94.34	96.84	99.07
4680	6.83	-0.075	1.118	1.64	4.85	8.00	94.22	96.81	99.11
7800	7.91	-0.064	1.100	1.46	4.47	7.61	94.54	97.00	99.17
11700	9.06	-0.058	1.087	1.29	4.25	7.27	94.46	97.05	99.26
23400	11.8	-0.049	1.066	1.16	3.86	6.73	94.60	97.19	99.30

Summary statistics and empirical quantiles for the feasible statistic, $\hat{Z}_n^{\log}$, that employs our estimate of the asymptotic variance, $\hat{\varpi}$, and $d = 10^{-6}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0,1)$. The simulation design is the two-factor model, **SV2F**, and $K(X_\delta)$ based on Tukey-Hanning$_2$ weights.

7 Empirical study

7.1 Analysis of General Electric transactions in 2004

In this subsection we implement our efficient, feasible inference procedure for the daily increments of $[Y]$ for the realised kernel estimator on transaction log-prices of General Electric (GE) shares carried out on the New York Stock Exchange in 2004. A more detailed analysis, including a comparison with results based on data from 2000 and on 29 other major stocks, is provided in our Web Appendix. We should note that the variance of the noise was around 10 times higher in 2000 than in 2004 and so looking over both periods is instructive. This Appendix also details the cleaning we carried out on the data before it was analysed and the precise way we calculated all of our statistics.

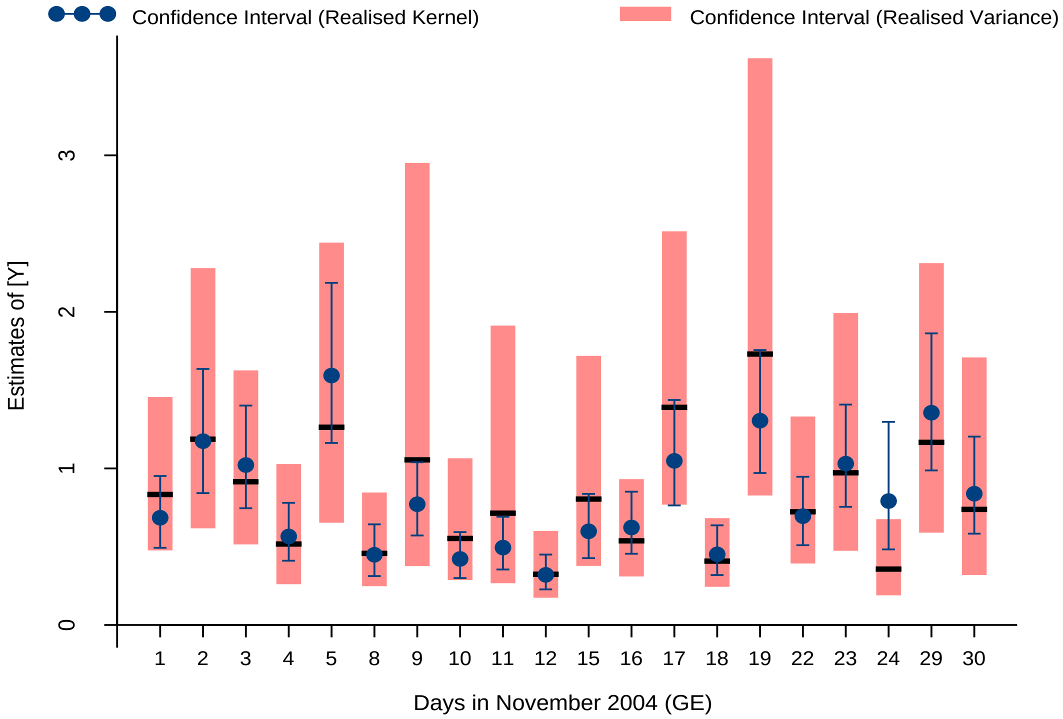

Figure 3: *Confidence intervals for the daily increments to* $[Y]$ *for General Electrics (GE) in November 2004. Shaded rectangles denote the 95% confidence intervals based on 20 minute returns using the Barndorff-Nielsen and Shephard (2002) feasible realised variance inference method. The shorter intervals are for* $K(X_\delta)$ *based on Tukey-Hanning*$_2$ *weights, sampling in tick times so the period over which returns are calculated is roughly 60 seconds.*

Our realised kernel will be implemented on returns recorded every S transactions, where S is selected each day so that there are approximately 360 observations a day.[7] This means that on average these returns are recorded every 60 seconds. This inference method will be compared to the feasible procedure of Barndorff-Nielsen and Shephard (2002), which ignores the presence of market microstructure effects, based on returns calculated over 20 minutes within each day. This baseline was chosen as Hansen and Lunde (2006) has suggested that the Barndorff-Nielsen and Shephard (2002) method was empirically sound when based on that type of interval for thickly traded stocks.

General Electric shares are traded very frequently on the NYSE. A typical day results in between $1,500$ and $6,000$ transactions. For this stock Hansen and Lunde (2006) have presented detailed work which suggests that over 60 second intervals it is empirically reasonable to assume that Y and

[7]As our sample size is quite large it is important to calculate it in tick time in order not to be influenced by the bias effect discussed by Renault and Werker (2005) caused by sampling in calender time.

Date	Trans	Lower	RV20m	Upper	n	Lower	KV60s	Upper	S	n	H	$\hat{\omega}^2$	$\check{\omega}^2$
1. Nov	4631	0.48	0.83	1.46	20	0.49	0.69	0.95	13	357	5	.0016	.0007
2. Nov	4974	0.62	1.19	2.28	20	0.84	1.17	1.64	14	356	6	.0025	.0011
3. Nov	4918	0.51	0.92	1.63	20	0.75	1.02	1.40	14	352	5	.0021	.0008
4. Nov	5493	0.26	0.52	1.03	20	0.41	0.57	0.78	16	344	5	.0013	.0005
5. Nov	5504	0.65	1.26	2.44	20	1.16	1.59	2.19	16	344	5	.0028	.0009
8. Nov	4686	0.25	0.46	0.85	20	0.31	0.45	0.64	14	335	6	.0014	.0008
9. Nov	4923	0.38	1.05	2.95	20	0.57	0.77	1.04	14	352	4	.0014	.0004
10. Nov	4970	0.29	0.55	1.07	20	0.30	0.42	0.59	14	355	6	.0013	.0007
11. Nov	4667	0.27	0.71	1.91	20	0.35	0.49	0.69	13	359	5	.0011	.0004
12. Nov	4822	0.17	0.32	0.60	20	0.23	0.32	0.45	14	345	6	.0009	.0005
15. Nov	4681	0.38	0.80	1.72	20	0.43	0.60	0.84	14	335	5	.0015	.0007
16. Nov	4526	0.31	0.54	0.93	20	0.45	0.62	0.85	13	349	5	.0011	.0004
17. Nov	5477	0.77	1.39	2.51	20	0.76	1.05	1.44	16	343	5	.0018	.0006
18. Nov	4738	0.24	0.41	0.68	20	0.32	0.45	0.64	14	339	6	.0014	.0007
19. Nov	5224	0.83	1.73	3.62	20	0.97	1.31	1.76	15	349	4	.0019	.0005
22. Nov	5359	0.39	0.72	1.33	20	0.51	0.69	0.95	15	358	5	.0012	.0004
23. Nov	5405	0.47	0.97	1.99	20	0.75	1.03	1.41	15	361	5	.0016	.0005
24. Nov	4626	0.19	0.36	0.68	20	0.48	0.79	1.30	13	356	5	.0013	.0004
29. Nov	4709	0.59	1.17	2.31	20	0.99	1.36	1.86	14	337	5	.0023	.0007
30. Nov	4719	0.32	0.74	1.71	20	0.58	0.84	1.20	14	338	6	.0018	.0007

Table 10: *Inference for General Electric (GE) volatility in November 2004. "Trans" denotes the number of transactions. RV20m is the realised variance based on 20-minute returns, and KV60s is $K(X_\delta)$ based on Tukey-Hanning$_2$ weights, and sampling of every S'th transaction price, so the period over which returns are calculated is roughly 60 seconds.*

U are uncorrelated and U is roughly a white noise process. Hence the main assumptions behind the inference procedure for our efficient kernel estimator are roughly satisfied and so we feel comfortable implementing the feasible limit theory on this dataset. We should note that on all the days in 2004 our realised kernel estimator of the daily increments of $[Y]$ was positive.

Figure 3 shows daily 95% confidence intervals (CIs) for the realised kernel for November 2004 using the modified Tukey-Hanning$_2$ weights (16) with $H = c^*\hat{\xi}n^{1/2}$. Also drawn are the corresponding results for the realised variance. We can see the realised kernel has much shorter CIs. The width of these intervals does change through time, with them tending to be slightly wider in high volatility periods. Over the entire year there are only 3 days when the CIs do not overlap.

Table 10 shows the details of these results for November 2004. The estimates of ω^2 are very small, ranging from about 0.0004 to 0.0011. These are in the range of the small to medium levels of noise set out in our Monte Carlo designs discussed in the previous Section. The Table shows the sample size for the realised kernel, which is between 335 and 361 intervals of roughly 60 seconds. Typically each interval corresponds to about 15 transactions. It records the daily selected value of H that ranges from 4 to 6, which is rather modest and is driven by the fact that ω^2 is quite small.

Table 11 provides summary statistics for some alternative estimators over the entire year. This suggests the other realised kernel estimators have roughly the same average value and that they are quite tightly correlated. The Table also records the summary statistics for the realised variance

δ	Average	Std. (HAC)	$\widehat{\text{Corr}}(\cdot, TH_2)$	acf(1)	acf(2)	acf(5)	acf(10)
Tukey-Hanning$_2$ kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.908	0.541 (1.168)	1.000	0.34	0.38	0.28	0.09
Parzen kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.914	0.546 (1.182)	0.999	0.35	0.37	0.28	0.09
Cubic kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.915	0.542 (1.172)	0.998	0.35	0.37	0.28	0.09
5th order kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.919	0.530 (1.160)	0.999	0.36	0.39	0.29	0.09
8th order kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.912	0.550 (1.185)	0.995	0.34	0.38	0.28	0.09
Bartlett kernel ($H = c\xi^2 n^{2/3}$)							
$\approx$ 1 minute	0.934	0.551 (1.192)	0.988	0.36	0.35	0.27	0.08
Simple Realised Variance $= \gamma_0(X_\delta)$							
20 minutes	0.879	0.524 (1.008)	0.794	0.28	0.24	0.26	0.06
5 minutes	0.948	0.518 (1.100)	0.954	0.36	0.34	0.26	0.10
1 minute	0.941	0.382 (0.919)	0.887	0.44	0.40	0.38	0.11
10 seconds	1.330	0.389 (1.142)	0.803	0.60	0.56	0.51	0.32
1 tick	2.183	0.569 (1.828)	0.733	0.69	0.66	0.57	0.48

Table 11: *Summary statistics for six realised kernels based on returns measured every S'th transaction, where S is selected such that returns span 60 seconds on average. The same statistics are computed for the realised variance using five different values for δ. The realised variance based on all tick-by-tick data is identical to the realised variance with $\delta = 1$ second. The empirical correlations between the realised Tukey-Hanning$_2$ kernel and each of the estimators are given in column 4 and some empirical autocorrelations are given in columns 5-8.*

computed using 20, 5, 1 minute and 10, and 1 second intervals. The last two of these estimators show a substantially higher mean. Interestingly, the realised QV based on 5 minute sampling is most correlated with the realized kernels, a result in line with the optimal sampling frequencies reported in Bandi and Russell (2005a). The realised kernels have a stronger degree of serial dependence than our benchmark realised variance that is based on a moderate sampling so the period over which returns are calculated is 20 minutes. This point suggests the realised kernel may be useful when it comes to forecasting, extending the exciting work of Andersen, Bollerslev, Diebold, and Labys (2001). The high serial dependence found in the realised variances based on the high sampling frequencies suggests a strong dependence in the bias components of these estimators.

7.2 Speculative analysis

The analysis in the previous subsection does not use all of the available data efficiently, for the realised kernel is computed only on every 15 or so transactions. This was carried out so that the empirical reality of the GE data matched the assumptions of our feasible central limit theory,

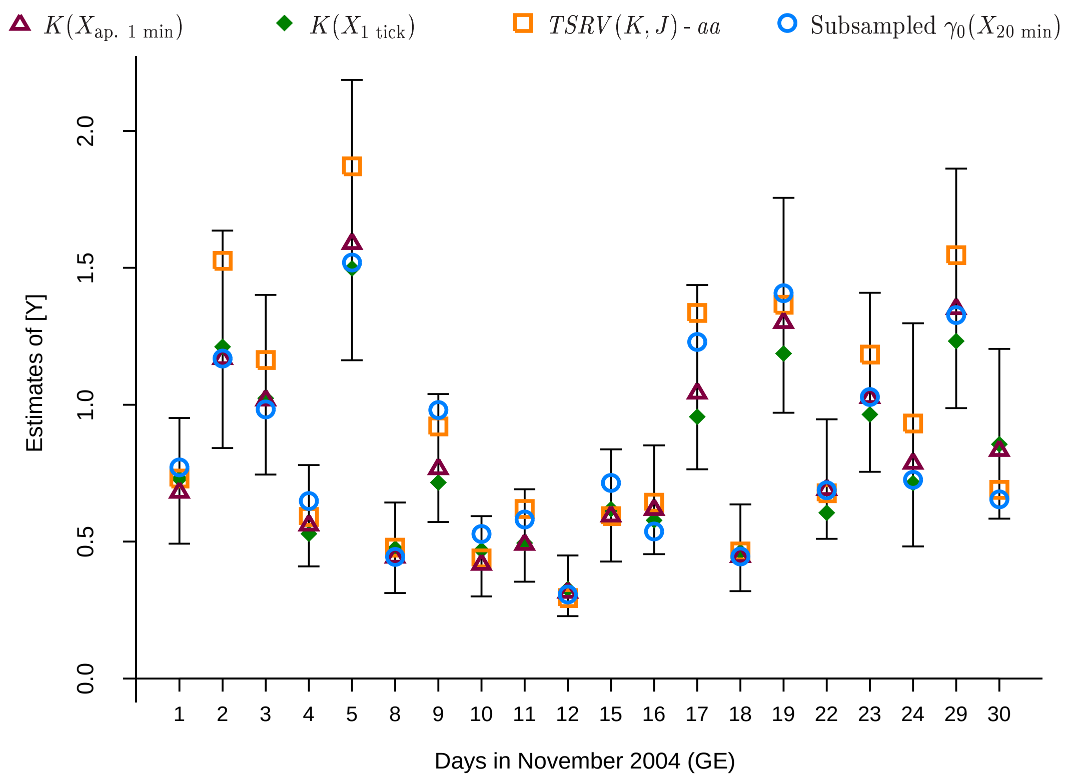

Figure 4: *Four estimators for the daily increments to* $[Y]$ *for General Electrics in November 2004. The intervals are the confidence intervals for* $K(X_\delta)$ *using on Tukey-Hanning*$_2$ *weights and returns sampled roughly every 60 seconds. Triangles denote averages of S distinct realised kernels, where each of the estimators is based on returns that span S transactions. Diamonds represent* $K(X_\delta)$ *using* $\delta = 1$ *tick. Circles are averages of 1200 realised variances using* $\delta = 20$ *minutes, where the individual RVs are obtained by shifting the times prices are recorded by one second at a time. Squares represent the bias adjusted two scale estimator, see Aït-Sahalia, Mykland, and Zhang (2006, eq. 4.22).*

allowing us to calculate daily confidence intervals. In this subsection, we give up on the goal of carrying out inference and simply focus on estimating $[Y]$ by employing all of the data. The results in Section 5.3 suggest our efficient realised kernel can do this, even though the $U \in \mathcal{WN}$ and $Y \perp\!\!\!\perp U$ are no longer empirically well-grounded assumptions. For these robust estimators we select $H = c_0 n^{2/3}$, where we use the same values for $c_0 = c^*\xi$ as in the previous subsection. Inevitably then, the results in this subsection will be more speculative than those given in the previous analysis.

We calculate the realised kernel using every transactions on each day, based on returns sampled roughly every 60 seconds, or by applying the kernel weights to returns sampled every transactions. The time series of these estimators are drawn in Figure 4, together with the corresponding bias corrected two scale estimator and a subsampled version of the realised variance estimator using 5 minute returns, where the degree of subsampling was selected to exhaust the available data. For the sake of comparison, we also include the confidence intervals from Figure 3.

δ	Average	Std. (HAC)	$\widehat{\text{Corr}}(\cdot, TH_2)$	acf(1)	acf(2)	acf(5)	acf(10)
Tukey-Hanning$_2$ kernel ($H = c\xi n^{1/2}$)							
$\approx$ 1 minute	0.908	0.541 (1.168)	1.000	0.34	0.38	0.28	0.09
Tukey-Hanning$_2$ kernel (inefficient rate $H = c\xi n^{2/3}$)							
1 tick	0.894	0.497 (1.104)	0.991	0.37	0.38	0.32	0.09
Parzen kernel (inefficient rate $H = c\xi n^{2/3}$)							
1 tick	0.901	0.502 (1.111)	0.990	0.37	0.37	0.31	0.09
Cubic kernel (inefficient rate $H = c\xi n^{2/3}$)							
1 tick	0.907	0.505 (1.115)	0.991	0.37	0.37	0.30	0.09
5th order kernel (inefficient rate $H = c\xi n^{2/3}$)							
1 tick	0.910	0.507 (1.121)	0.989	0.37	0.37	0.30	0.09
8th order kernel (inefficient rate $H = c\xi n^{2/3}$)							
1 tick	0.908	0.549 (1.183)	0.996	0.34	0.36	0.28	0.09
Subsampled Realised Variance							
20 minutes	0.885	0.516 (1.036)	0.933	0.27	0.27	0.27	0.08
5 minutes	0.943	0.503 (1.088)	0.984	0.37	0.32	0.30	0.08
1 minute	0.942	0.376 (0.921)	0.899	0.46	0.43	0.38	0.12
ZMA (2005) TSRV$(K, 1)$							
1 tick	0.544	0.321 (0.711)	0.842	0.40	0.34	0.29	0.05
1 tick (*adj*)	0.596	0.353 (0.784)	0.854	0.40	0.34	0.29	0.04
AMZ (2006) TSRV(K, J)							
1 tick	0.736	0.436 (0.929)	0.944	0.33	0.35	0.28	0.11
1 tick (*aa*)	0.946	0.560 (1.194)	0.944	0.33	0.35	0.28	0.11

Table 12: *Twelve estimators that make use of all available (tick-by-tick) data benchmarked against the realised kernel in the first row, which is based on returns that span 1 minute on average. The realised kernels using $\delta = 1$ tick use a large H in order to counteract possible dependence in U. The subsampled realised variances make use of all available data by changing the initial place at which prices are recorded and average the resulting estimates. E.g. when $\delta = 5$ minutes the subsampled RV is an average of 300 estimates, obtained by shifting the times at which prices are recorded by one second at a time. The four two scale estimators are those of Zhang, Mykland, and Aït-Sahalia (2005, equations 55 and 64) and Aït-Sahalia, Mykland, and Zhang (2006, equations 4.4 and 4.22), where the last two are designed to be robust to dependence in U. The estimators identified by (adj) and (aa) involve finite-sample bias corrections.*

Figure 4 shows that realised kernels give very similar estimates – on some days the estimates are almost identical. The two scale estimators by Zhang, Mykland, and Aït-Sahalia (2005) are quite biased because they rely on the white noise, which is at odds with tick-by-tick data. The two scale estimators by Aït-Sahalia, Mykland, and Zhang (2006) are designed to be robust to serial dependence in U. The bias observed in the unadjusted estimator is ascribed to the bias of $\hat{\omega}^2$, as we discussion in Section 5.1. The bias is overcome by the "area adjusted" version of this estimator, which seems in line with the results for the realised kernels and the subsampled RV estimators.[8]

[8] In empirical work we found this estimator to be sensitive to the choice of K. To be consistent with our empirical

Table 12 provides summary statistics of these estimators. The realised kernels are pretty robust to choice of the design of the weights.

8 Conclusions

In this paper we have provided a detailed analysis of the accuracy of realised kernels as estimators of quadratic variation when an efficient price is obscured by simple market frictions. We show how to make these estimators consistent and derive central limit theorems for the estimators under various assumptions about the kernel weights. Such estimators can be made to converge at the fastest possible rate and are very close to being efficient. They can be made robust to dynamics in the noise process, robust to endogenous market frictions and robust to endogenous spacing in the timing of the data.

Our efficient feasible central limit theory for our estimators performed satisfactorily in Monte Carlo experiments designed to assess finite sample behaviour. Our kernel was shown to be consistent under rather broad assumptions on the dynamics of the noise term. We have applied the estimator empirically, using 60 second return data on General Electric transaction data for 2004. Feasible inference for our realised kernel is compared with that for a simpler realised variance estimator based on 20 minute returns. The empirical results suggest that the realized kernel estimator is more accurate. Its serial correlation suggests that the realized kernel may be useful for forecasting, following Andersen, Bollerslev, Diebold, and Labys (2001).

There are many possible extensions to this work, e.g. multivariate versions of these results which deal with the scrambling effects discussed by, for example, Hayashi and Yoshida (2005), Bandi and Russell (2005b), Zhang (2005), Sheppard (2006), Voev and Lunde (2007) and Griffin and Oomen (2006) and derive an asymptotically efficient choice of kernel under temporal dependence in U.

References

Aït-Sahalia, Y., P. A. 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*, 351–416.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2006). Ultra high frequency volatility estimation with dependent microstructure noise. Unpublished paper: Department of Economics, Princeton University.

Aldous, D. J. and G. K. Eagleson (1978). On mixing and stability of limit theorems. *Annals of Probability 6*, 325–331.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000). Great realizations. *Risk 13*, 105–108.

findings, J has to be about 15 (yielding returns measured roughly over 1 minute). The estimator in Aït-Sahalia, Mykland, and Zhang (2006) needs $K > J$ and the theory relies on $K/J \to \infty$. Unfortunately, their formula for choosing K, typically selects a value smaller than J. So we imposed $K \geq 5J$, which worked well in our empirical analysis.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55. Correction published in 2003, volume 98, page 501.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics 20*, 371–395.

Bandi, F. M. and J. R. Russell (2005a). Microstructure noise, realized variance, and optimal sampling. Unpublished paper, Graduate School of Business, University of Chicago.

Bandi, F. M. and J. R. Russell (2005b). Realized covariation, realized beta and microstructure noise. Unpublished paper, Graduate School of Business, University of Chicago.

Bandi, F. M. and J. R. Russell (2006). Market microstructure noise, integrated variance estimators, and the limitations of asymptotic approximations: A solution. Unpublished paper, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In Y. Kabanov, R. Lipster, and J. Stoyanov (Eds.), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, pp. 33–68. Springer.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, and N. Shephard (2006). Limit theorems for realised bipower variation in econometrics. *Econometric Theory 22*, 677–719.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica 72*, 885–925.

Barndorff-Nielsen, O. E. and N. Shephard (2005). Power variation and time change. *Theory of Probability and Its Applications 50*, 1–15.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*, 1–30.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society, Supplement 8*, 27–41.

Billingsley, P. (1995). *Probability and Measure* (3 ed.). New York: Wiley.

Bollen, B. and B. Inder (2002). Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance 9*, 551–562.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Delattre, S. and J. Jacod (1997). A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round off errors. *Bernoulli 3*, 1–28.

Fang, Y. (1996). Volatility modeling and estimation of high-frequency data with Gaussian noise. Unpublished Ph.D. thesis, Sloan School of Management, MIT.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). Expected stock returns and volatility. *Journal of Financial Economics 19*, 3–29.

Gallant, A. R. (1987). *Nonlinear Statistical Models.* New York: John Wiley.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Gloter, A. and J. Jacod (2001a). Diffusions with measurement errors. I — local asymptotic normality. *ESAIM: Probability and Statistics 5*, 225–242.

Gloter, A. and J. Jacod (2001b). Diffusions with measurement errors. II — measurement errors. *ESAIM: Probability and Statistics 5*, 243–260.

Goncalves, S. and N. Meddahi (2004). Bootstrapping realized volatility. Unpublished paper, CIRANO, Montreal.

Griffin, J. E. and R. C. A. Oomen (2006). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Unpublished: Dept. of Statistics, University of Warwick.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications.* San Diego: Academic Press.

Hansen, P. R., J. Large, and A. Lunde (2006). Moving average-based estimators of integrated variance. *Econometric Reviews*. Forthcoming.

Hansen, P. R. and A. Lunde (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics 3*, 525–554.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics 24*, 127–218.

Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli 11*, 359–379.

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

Jacod, J. (1994). Limit of random measures associated with the increments of a Brownian semimartingale. Preprint number 120, Laboratoire de Probabilitiés, Université Pierre et Marie Curie, Paris.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In J. Azema, M. Emery, and M. Yor (Eds.), *Séminaire Probability XXXI, Lecture Notes in Mathematics*, Volume 1655, pp. 232–246. Berlin: Springer Verlag.

Jacod, J. (2006). Asymptotic properties of realized power variations and related functionals of semimartingales. Unpublished paper: Laboratoire de Probabilites, Universite Paris VI.

Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability 26*, 267–307.

Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes* (2 ed.). Springer: Berlin.

Kalnina, I. and O. Linton (2006). Estimating quadratic variation consistently in the presence of correlated measurement error. Unpublished paper: Department of Economics, LSE.

Large, J. (2005). Theoretical and empirical studies of financial markets. Unpublished D.Phil. thesis, University of Oxford.

Malliavin, P. and M. E. Mancino (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics 6*(1), 49–61.

Mancino, M. E. and S. Sanfelici (2007). Stepwise multiple testing as formalized data snooping. *Fortcoming in Computational Statistics and Data Analysis*.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics 17*, 479–508.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*, 323–361.

Mykland, P. A. (2006). A Gaussian calculus for inference from high frequency data. Unpublished paper: Department of Statistics, University of Chicago.

Mykland, P. A. and L. Zhang (2006). ANOVA for diffusions and Ito processes. *Annals of Statistics 34*, 1931–1963.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Oomen, R. A. A. (2005). Properties of bias corrected realized variance in calender time and business time. *Journal of Financial Econometrics 3*, 555–577.

Park, S. K. and R. A. Schowengerdt (1983). Image reconstruction by parametric cubic convolution. *Computer Vision, Graphics, and Image Processing 3*, 258–272.

Phillips, P. C. B. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica 58*, 165–193.

Phillips, P. C. B., Y. Sun, and S. Jin (2003). Long run variance estimation using steep origin kernels without truncation. Unpublished paper: Cowles Foundation, Yale University.

Phillips, P. C. B. and J. Yu (2006). Comment of Hansen and Lunde (2006). *Journal of Business and Economic Statistics 24*, 202–208.

Politis, D. N. (2005). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. Unpublished paper: Department of Mathematics, USCD.

Politis, D. N. and J. P. Romano (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis 16*, 67–103.

Protter, P. (2004). *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

Renault, E. and B. Werker (2005). Stochastic volatility models with transaction time risk. Department of Economics, University of North Carolina.

Rényi, A. (1963). On stable sequences of events. *Sankya, Series A 25*, 293–302.

Revuz, D. and M. Yor (1999). *Continuous Martingales and Brownian motion* (3 ed.). Heidelberg: Springer.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.

Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security prices. Working paper 11, Graduate School of Business Administration, University of California, Berkeley. Reprinted in Shephard (2005).

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Sheppard, K. (2006). Measuring realized covariance. Unpublished paper: Department of Economics, University of Oxford.

Stein, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association 82*, 765–772.

Voev, V. and A. Lunde (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics 5*, 68–104.

Zhang, L. (2005). Estimating covariation: Epps effect and microstructure noise. Unpublished paper, Department of Finance, University of Illinois, Chicago.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli 12*, 1019–1043.

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.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics 14*, 45–52.

Appendix A: Stable Convergence

The concise mathematical definition of stable convergence is as follows. Let $\{\mathcal{X}_n\}$ denote a sequence of random variates defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in a Polish space $(E, \mathcal{E})$,[9] and let $\mathcal{G}$ be a sub-σ–field of $\mathcal{F}$. Then X_n is said to converge $\mathcal{G}$-stably in law if there exists a probability measure μ on $(\Omega \times E, \mathcal{F} \times \mathcal{E})$ such that for every bounded G-measurable random variable[10] V on $(\Omega, \mathcal{F}, P)$ and every bounded and continuous function f on E we have that, for $n \to \infty$,

$$\mathrm{E}\{Vf(\mathcal{X}_n)\} \to \int V(\omega) f(x) \mu(\mathrm{d}\omega, \mathrm{d}x). \tag{A.1}$$

If $\mathcal{X}_n$ converges stably in law then, in particular, it converges in distribution (or in law or weak convergence), the limiting law being $\mu(\Omega, \cdot)$. Accordingly, one says that $\mathcal{X}_n$ converges stably to some E-valued random variate $\mathcal{X}$ on $\Omega \times E$, written $\mathcal{X}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{X}$ provided $\mathcal{X}$ has law $\mu(\Omega, \cdot)$. Things can always be set up so that such a random variate $\mathcal{X}$ exists.

This concept and its extension to stable convergence of processes is discussed in Jacod and Shiryaev (2003, pp. 512-518). For earlier discussions see, for example, Rényi (1963), Aldous and Eagleson (1978), Hall and Heyde (1980, pp. 56-58) and Jacod (1997). An early use of this concept in econometrics was Phillips and Ouliaris (1990). It is used extensively in Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006).

However, the above formalisation does not reveal the key property of stable convergence which is that convergence $\mathcal{G}$-stably in law to X is equivalent[11] to the statement that for any G-measurable random variable W, the pair $(W, \mathcal{X}_n)$ converges in law to $(W, \mathcal{X})$.

The following results are helpful in the use we wish to make of this concept.

Let $\{\mathcal{Y}_n\}$ and $\{\mathcal{Z}_n\}$ denote two sequences of random variates on $(\Omega, \mathcal{F}, P)$ and with values in $(E, \mathcal{E})$ and suppose that $\mathcal{Y}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y}$ for some sub-σ–field $\mathcal{G}$ of $\mathcal{F}$

Lemma 1 *If $\mathcal{Y}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y}$ and $\{W_n\}$ is a sequence of positive random variables on $(\Omega, \mathcal{F}, P)$ tending in probability to a positive $\mathcal{G}$-measurable random variable W such that $W_n/W \overset{p}{\to} 1$ then $W_n\mathcal{Y}_n \overset{L}{\to} W\mathcal{Y}$.*

Proof. By the definition of $\mathcal{G}$-stable convergence we have that $(W, \mathcal{Y}_n) \overset{L}{\to} (W, \mathcal{Y})$ and therefore that $W\mathcal{Y}_n \overset{L}{\to} W\mathcal{Y}$. Since $W_n\mathcal{Y}_n = (W_n/W) W\mathcal{Y}_n$ and $W_n/W \overset{p}{\to} 1$ we have $W_n\mathcal{Y}_n \overset{L}{\to} W\mathcal{Y}$. $\square$

Lemma 2 *If E is a normed vector space and if $\mathcal{Z}_n \overset{p}{\to} 0$ then $\mathcal{Y}_n + \mathcal{Z}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y}$.*

Proof. This follows simply from the defining condition (A.1), using the tightness of $\{\mathcal{Y}_n\}$. $\square$

Lemma 3 *If $\mathcal{Z}_n \overset{L}{\to} \mathcal{Z}$ and if $\{\mathcal{Y}_n\}$ and $\{\mathcal{Z}_n\}$ are independent then $\mathcal{Y}_n + \mathcal{Z}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y} + \mathcal{Z}$*

Proof. We have $\mathcal{Y}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y}$ if and only if $(W, \mathcal{Y}_n) \overset{L}{\to} (W, \mathcal{Y})$ for all $\mathcal{G}$-measurable random variables W on Ω. Since $(W, \mathcal{Y}_n) \perp\!\!\!\perp \mathcal{Z}_n$ and $\mathcal{Z}_n \overset{L}{\to} \mathcal{Z}$ it follows that $(W, \mathcal{Y}_n, \mathcal{Z}_n) \overset{L}{\to} (W, \mathcal{Y}, \mathcal{Z})$. This implies $(W, \mathcal{Y}_n + \mathcal{Z}_n) \overset{L}{\to} (W, \mathcal{Y} + \mathcal{Z})$ for all $\mathcal{G}$-measurable W which is equivalent to $\mathcal{Y}_n + \mathcal{Z}_n \overset{L_{\mathcal{G}}}{\to} \mathcal{Y} + \mathcal{Z}$. $\square$

[9] i.e. E is a complete separable metric space and $\mathcal{E}$ denotes the Borel σ-algebra of E.

[10] As is common, we take random variable to mean a real-valued random variate. We use capital Roman letters to denote random variables or vectors and capital Frakturs to denote general random variates.

[11] See Jacod and Shiryaev (2003, pp. 512-518) Proposition 5.33.

Definition 5 *Let $\{Z_n\}$ be a sequence of d-dimensional random vectors on $(\Omega, \mathcal{F}, P)$. We say that the conditional law of Z_n given $\mathcal{G}$ converges in probability provided there exists a $\mathcal{G}$-measurable characteristic function $\phi(\zeta|\mathcal{G})$ (possibly defined on an extension of $(\Omega, \mathcal{F}, P)$) such that for all $\zeta \in \mathbb{R}^d$*

$$\phi_n(\zeta|\mathcal{G}) \xrightarrow{p} \phi(\zeta|\mathcal{G}) \tag{A.2}$$

where $\phi_n(\zeta|\mathcal{G}) = \mathrm{E}\left\{e^{i\zeta^\top Z_n}|\mathcal{G}\right\}$.

Remark. We can, without restriction, assume that there exists a d-dimensional random vector Z, defined on $(\Omega, \mathcal{F}, P)$ or an extension thereof, such that $\phi(\zeta|\mathcal{G})$ is the conditional characteristic function of Z given $\mathcal{G}$. Then we also write (A.2) as

$$\mathcal{L}(Z_n|\mathcal{G}) \xrightarrow{p} \mathcal{L}(Z|\mathcal{G}),$$

where $\mathcal{L}\{\cdot|\mathcal{G}\}$ means conditional law given $\mathcal{G}$.

Proposition 4 *Let $\{Y_n\}$ and $\{Z_n\}$ be sequences of random vectors. Suppose $Y_n \xrightarrow{L_\mathcal{G}} Y$ and $\mathcal{L}\{Z_n|\mathcal{G}\} \xrightarrow{p} \mathcal{L}\{Z|\mathcal{G}\}$. Then $(Y_n, Z_n) \xrightarrow{L_\mathcal{G}} (Y, Z)$.*

Proof. Let W be an arbitrary $\mathcal{G}$-measurable random variable. For all $\eta, \zeta, \psi \in \mathbb{R}$ and $n \to \infty$, it must be verified that $\mathrm{E}\left\{e^{i\eta^\top Y_n + i\zeta^\top Z_n + i\psi W}\right\} \to \mathrm{E}\left\{e^{i\eta^\top Y + i\zeta^\top Z + i\psi W}\right\}$. Now,

$$\begin{aligned}
\mathrm{E}\left(e^{i\eta^\top Y_n + i\zeta^\top Z_n + i\psi W}\right) - \mathrm{E}\left(e^{i\eta^\top Y + i\zeta^\top Z + i\psi W}\right) &= \mathrm{E}\left(e^{i\eta^\top Y_n + i\zeta^\top Z_n + i\psi W}\right) \\
- \mathrm{E}\left(e^{i\eta^\top Y_n + i\zeta^\top Z + i\psi W}\right) &+ \mathrm{E}\left(e^{i\eta^\top Y_n + i\zeta^\top Z + i\psi W}\right) - \mathrm{E}\left(e^{i\eta^\top Y + i\zeta^\top Z + i\psi W}\right) \\
= \mathrm{E}\left[e^{i\eta^\top Y_n + i\psi W}\left\{\phi_n(\zeta|\mathcal{G}) - \phi(\zeta|\mathcal{G})\right\}\right] &+ \mathrm{E}\left\{e^{i\eta^\top Y_n + i\psi W}\phi(\zeta|\mathcal{G})\right\} - \mathrm{E}\left\{e^{i\eta^\top Y + i\psi W}\phi(\zeta|\mathcal{G})\right\}.
\end{aligned}$$

By (A.2),

$$\left|\mathrm{E}\left[e^{i\eta^\top Y_n + i\psi W}\left\{\phi_n(\zeta|\mathcal{G}) - \phi(\zeta|\mathcal{G})\right\}\right]\right| \leq \mathrm{E}\left\{|\phi_n(\zeta|\mathcal{G}) - \phi(\zeta|\mathcal{G})|\right\} \to 0.$$

Moreover, since $Y_n \xrightarrow{L_\mathcal{G}} Y$, it holds that $\mathrm{E}\left\{e^{i\eta^\top Y_n + i\psi W}\phi(\zeta|\mathcal{G})\right\} - \mathrm{E}\left\{e^{i\eta^\top Y + i\psi W}\phi(\zeta|\mathcal{G})\right\} \to 0$. $\square$

Remark. The conclusion of the Proposition holds not only for random variables Y_n and Z_n but extends to general random variates. In the present paper such an extension is, however, not required.

Appendix B: Proofs

Proof of Theorem 1. We first show the stated separate limit results for $\delta^{-1/2}\Gamma_{\delta,H}$, $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ and $\delta^{1/2}\{\gamma(U_\delta) - E\gamma(U_\delta)\}$. Throughout we take $t = 1$ and so $\delta = 1/n$ for the other cases follow trivially. Consider first $\gamma(Y_\delta)$. Write $y_j = Y_{j/n} - Y_{(j-1)/n}$, then the terms we need to study are $\sum_{j=1}^n y_j^2$, $\sum_{j=1}^n y_j y_{j+1}$, $\ldots$, $\sum_{j=1}^n y_j y_{j+H}$. This can be written in the form of a set of multipower variation statistics (e.g. Barndorff-Nielsen, Graversen, Jacod, and Shephard (2006))

$$\sum_{j=1}^n \prod_{k=0}^H g_{l,k}(y_{j+k}) = \sum_{j=1}^n g_l(y_j, y_{j+1}, ..., y_{j+H}), \quad l = 0, 1, 2, ..., H,$$

by selecting the functions $g_{l,k}$ appropriately. In particular, writing $g_{l,k}$ into a matrix form

$$g(x)=\begin{pmatrix} x_0^2 & 1 & 1 & \cdots & 1 \\ x_0 & x_1 & 1 & \cdots & 1 \\ x_0 & 1 & x_2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \\ x_0 & 1 & 1 & \cdots & x_H \end{pmatrix}, \quad x\in\mathbb{R}^{H+1}.$$

We satisfy all the conditions in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006) except there the $g_{l,k}(x)$ are assumed to be even functions. To see that for our specific form of g this assumption of evenness does not matter, we will look solely at the $\sum_{j=1}^{n} y_j y_{j+1}$ statistic. The other terms then follow immediately by the same argument.

We think of the bipower variation statistic

$$\frac{1}{n}\sum_{j=1}^{n} g_{2,1}\left(\sqrt{n}y_j\right) g_{2,2}\left(\sqrt{n}y_{j+1}\right),$$

where $g_{2,1}(x_0)=x_0$ and $g_{2,2}(x_1)=x_1$. Then using the notation of Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006) that $\rho_\sigma(h)=\mathrm{E}\left\{h\left(x\right)|\sigma^2\right\}$, $x|\sigma^2\sim N(0,\sigma^2)$ we note that $\rho_\sigma(g_{2,1})=\rho_\sigma(g_{2.2})=0$, which enormously simplifies the task. Inspection of their proof shows two steps use this assumption. It is used on page 67, where various features of their z_i^n are defined and studied. In our case $z_i^n=0$ and so they follow immediately. The only non-trivial step involves their equation (4.12) applied to the bipower case which is presented in the first equation of their Proposition 4.2. This corresponds to checking condition (7.29) in Jacod and Shiryaev (2003). The sole task then is satisfied if we can show $\sum_{j=1}^{n}\mathrm{E}(\zeta_j^n w_j|\mathcal{F}_{\frac{j-1}{n}})\overset{p}{\to}0$, (actually converging to a continuous process would be enough) where we define

$$\zeta_j^n=\frac{1}{\sqrt{n}}g_{2,1}(\beta_{j-1})g_{2,2}(\beta_{j-1}^{'}),$$

with $\beta_j=\sqrt{n}\sigma_{\frac{j-1}{n}}w_j$, $\beta_j'=\sqrt{n}\sigma_{\frac{j-1}{n}}w_{j+1}$, and $w_j=W_{\frac{j}{n}}-W_{\frac{j-1}{n}}$. Thus $\zeta_j^n=\sqrt{n}\sigma_{\frac{i-2}{n}}^2 w_{j-1}w_j$. Clearly

$$\sum_{j=1}^{n}\mathrm{E}\left(\zeta_j^n\Delta_j^n W|\mathcal{F}_{\frac{j-1}{n}}\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sigma_{\frac{j-2}{n}}^2\left(\Delta_{j-1}^n W\right)=\frac{1}{\sqrt{n}}\int_0^1\sigma_u^2\mathrm{d}W_u+o_p(n^{-1/2})\overset{p}{\to}0.$$

Hence the result holds. This implies then that Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006) result holds and so

$$\delta^{-1/2}\left\{\ \gamma_0(Y_\delta)-\textstyle\int_0^t\sigma_u^2\mathrm{d}u,\gamma_1(Y_\delta),\ldots,\gamma_H(Y_\delta)\ \right\}^\top\overset{L_Y}{\to}MN\left(0,\int_0^t A(\sigma_u,g)\mathrm{d}u\right),$$

where $A_{l,k}(\sigma,g)=\mathrm{Cov}\left\{g_l(x),g_k(x)|\sigma^2\right\}=\sigma^4\mathrm{Cov}\left\{g_l(x),g_k(x)\right\}$, $x|\sigma^2\sim N(0,\sigma^2 I)$, and $x\sim N(0,I)$. Simple calculations based on the normal distribution delivers the result immediately.

In considering the cross term $\gamma\left(Y_\delta,U_\delta\right)+\gamma\left(U_\delta,Y_\delta\right)$, let $\bar{\gamma}\left(Y_\delta,U_\delta\right)=[Y_\delta]_t^{-1/2}\gamma\left(Y_\delta,U_\delta\right)$ and $\bar{y}_j=[Y_\delta]_t^{-1/2}y_j$. Then

$$\bar{\gamma}\left(Y_\delta,U_\delta\right)=\sum_{j=1}^{n}\bar{y}_j\begin{pmatrix} U_{j\delta}-U_{(j-1)\delta} \\ U_{(j+1)\delta}-U_{j\delta}+U_{(j-1)\delta}-U_{(j-2)\delta} \\ \vdots \\ U_{(j+H)\delta}-U_{(j+H-1)\delta}+U_{(j-H)\delta}-U_{(j-H-1)\delta}\end{pmatrix}$$

and the coefficients $\bar{y}_j$ are uniformly asymptotically negligible for almost all realisations of Y. Standard central limit theory then yields that, conditionally on Y, $\bar{\gamma}(Y_\delta, U_\delta) \xrightarrow{L} N\left(0, 2\omega^2 B\right)$. Since $\gamma(Y_\delta, U_\delta) - \gamma(U_\delta, Y_\delta) = o_p(1)$ and $[Y_\delta]_t \xrightarrow{p} [Y]_t$ this implies that a.s.

$$[Y_\delta]_t^{-1/2} \left(\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)\right) \xrightarrow{L} N\left(0, 2\omega^2 B\right).$$

Finally we verify that jointly $\delta^{-1/2}\Gamma_{\delta,H}$, $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ and $\delta^{1/2}\left\{\gamma(U_\delta) - E\gamma(U_\delta)\right\}$ converge $\sigma(Y)$-stably. On account of Proposition 4, we can conclude that the joint law of $\delta^{-1/2}\Gamma_{\delta,H}$ and $[Y_\delta]_t^{-1/2}\left(\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)\right)$ converges $\sigma(Y)$-stably. By Lemma 1 we then obtain that $\delta^{-1/2}\Gamma_{\delta,H}$ and $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ jointly converges $\sigma(Y)$-stably. Finally, since $\delta^{1/2}\left\{\gamma(U_\delta) - E\gamma(U_\delta)\right\} \xrightarrow{L} N\left(0, 4\omega^2 C\right)$, a further application of Proposition 4 gives that $\delta^{-1/2}\Gamma_{\delta,H}$, $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ and $\delta^{1/2}\left\{\gamma(U_\delta) - E\gamma(U_\delta)\right\}$ are jointly $\sigma(Y)$-stably convergent.

Next we consider the pure noise term, $\gamma(U_\delta)$. Define $V_h = \sum_{j=1}^{n-h-1} U_{j\delta}U_{(j+h)\delta}$, $h \geq 0$, $W_h = \sum_{j=1}^{h-1} U_{(j-h)\delta}U_{j\delta} + \sum_{j=n-h+1}^{n} U_{j\delta}U_{(j+h)\delta}$, $h \geq 1$, and $Z_h = U_0 U_{h\delta} + U_t U_{(n-h)\delta}$, $h \in \mathbb{Z}$, where all terms are mutually uncorrelated. Then $\gamma_0(U_\delta) = (2V_0 - 2V_1) + (Z_0 - 2Z_1)$ and

$$\begin{aligned}
\gamma_1(U_\delta) + \gamma_{-1}(U_\delta) &= (-2V_0 + 4V_1 - 2V_2) + (-W_2) + (Z_{-1} - Z_0 + 3Z_1 - 2Z_2), \\
\gamma_h(U_\delta) + \gamma_{-h}(U_\delta) &= (-2V_{h-1} + 4V_h - 2V_{h+1}) + (-W_{h-1} + 2W_h - W_{h+1}) \\
&\qquad + (Z_{-h} - Z_{-h+1} - Z_{h-1} + 3Z_h - 2Z_{h+1}), \quad h \geq 2.
\end{aligned}$$

So that $\gamma(U_\delta) = \gamma_V(U_\delta) + \gamma_W(U_\delta) + \gamma_Z(U_\delta)$, where

$$\begin{aligned}
\gamma_V(U_\delta) &= 2(V_0 - V_1, -V_0 + 2V_1 - V_2, \ldots, -V_{H-1} + 2V_H - V_{H+1})^\top, \\
\gamma_W(U_\delta) &= (0, -W_0 + 2W_1 - W_2, \ldots, -W_{H-1} + 2W_H - W_{H+1})^\top,
\end{aligned}$$

(using the convention $W_0 = W_1 = 0$) and

$$\gamma_Z(U_\delta) = \begin{pmatrix} Z_0 - 2Z_1 \\ Z_{-1} - Z_0 + 3Z_1 - 2Z_2 \\ Z_{-2} - Z_{-1} - Z_1 + 3Z_2 - 2Z_3 \\ \vdots \\ Z_{-H} - Z_{-H+1} - Z_{H-1} + 3Z_H - 2Z_{H+1} \end{pmatrix}.$$

Since $\mathrm{Var}(V_0) = (n-1)\lambda^2\omega^4$ and $\mathrm{Var}(V_h) = (n-h-1)\omega^4$ for $h \geq 1$. Also, $\mathrm{Var}(W_0) = 0$ and $\mathrm{Var}(W_h) = 2\omega^4(h-1)$. It is now straight forward to show that

$$\mathrm{Var}\left\{\gamma_V(U_\delta)\right\} = 4\omega^4\left(nC + D^{(V)}\right) \quad \text{and} \quad \mathrm{Var}\left\{\gamma_W(U_\delta)\right\} = 4\omega^4 D^{(W)}, \quad \text{where}$$

$$D_{11}^V = \begin{pmatrix} -\lambda^2 - 2 & \bullet \\ \lambda^2 + 4 & -\lambda^2 - 11 \end{pmatrix}, D_{11}^W = \begin{pmatrix} 0 & \bullet \\ 0 & \frac{1}{2} \end{pmatrix}, D_{21}^V = \begin{pmatrix} -2 & 10 \\ 0 & -3 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}, D_{21}^W = \begin{pmatrix} 0 & -1 \\ 0 & \frac{1}{2} \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix},$$

$$D_{22}^V = \begin{pmatrix} -18 & \bullet & \bullet & \bullet & \bullet \\ 14 & -24 & \bullet & \bullet & \bullet \\ -4 & 18 & -30 & \bullet & \bullet \\ 0 & -5 & 22 & -36 & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \text{ and } D_{22}^W = \begin{pmatrix} 3 & \bullet & \bullet & \bullet & \bullet \\ -3 & 6 & \bullet & \bullet & \bullet \\ \frac{2}{2} & -5 & 9 & \bullet & \bullet \\ 0 & \frac{3}{2} & -7 & 12 & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Finally U_0 and U_t are both averages of m independent noise terms, so that $\mathrm{Var}(U_0^2) = m^{-3}\lambda^2\omega^4 + 4m^{-4}\frac{(m-1)m}{2}\omega^4 = m^{-3}(\lambda^2-2)\omega^4 + 2m^{-2}\omega^4$, and $\mathrm{Var}(U_0U_{h\delta}) = m^{-1}\omega^4$ for $j \neq 0$, and similarly for $U_tU_{t-h\delta}$ terms. So

$$\mathrm{Var}(Z_h) = \mathrm{Var}(U_0U_{h\delta} + U_tU_{t-h\delta}) = \begin{cases} 4m^{-2}\omega^4\left\{m^{-1}(\frac{\lambda^2}{2}-1)+\omega^4\right\} & h=0, \\ 2m^{-1}\omega^4 & h\neq 0, \end{cases}$$

whereby $\mathrm{Var}\left\{\gamma_Z(U_\delta)\right\} = 4\omega^4 m^{-1}E$ follows. Using that the various terms are uncorrelated we have $\mathrm{Var}\left\{\gamma(U_\delta)\right\} = 4\omega^4\left(nC + D + m^{-1}E\right)$, where $D = D^V + D^W$. □

Proof of Theorem 2. $w^\intercal Aw = 2k_\bullet^{0,0}$ follows from the diagonal structure of A. Next $w^\intercal Bw = \sum_{h=1}^{H-1}\left\{k(\frac{H}{H}) - k(\frac{H-1}{H})\right\}^2$ gives the second result. The third results follows from

$$w^\intercal Cw = \left\{k(\tfrac{0}{H}) - 2k(\tfrac{1}{H}) + k(\tfrac{1}{H})^2\right\} + \sum_{h=1}^{H-1}\left\{k(\tfrac{h-1}{H}) - 2k(\tfrac{h}{H}) + k(\tfrac{h+1}{H})\right\}^2 + \left\{k(\tfrac{H}{H}) - (k(\tfrac{H-1}{H})\right\}^2.$$

With D^V and D^W defined in the proof of Theorem 1 and $D = D^V + D^W$, the fourth result follows from

$$-w^\intercal D^V w = 0 + 2\left\{k(\tfrac{1}{H}) - k(\tfrac{0}{H})\right\}^2 + (H+2)\left\{k(\tfrac{H}{H}) - k(\tfrac{H-1}{H})\right\}^2 + \sum_{h=1}^{H-1}(h+2)\left\{k(\tfrac{h-1}{H}) - 2k(\tfrac{h}{H}) + k(\tfrac{h+1}{H})\right\}^2,$$

$$w^\intercal D^W w = \frac{1}{2}\sum_{h=1}^{H-1} h\left\{k(\tfrac{h-1}{H}) - 2k(\tfrac{h}{H}) + k(\tfrac{h+1}{H})\right\}^2 + \frac{1}{2}H\left\{k(\tfrac{H}{H}) - k(\tfrac{H-1}{H})\right\}^2.$$

The last result follows from, $w^\intercal Ew = \left\{k(\frac{1}{H}) - k(\frac{0}{H})\right\}^2 + \sum_{h=1}^{H}\left\{k(\frac{h}{H}) - k(\frac{h-1}{H})\right\}^2 + \sum_{h=1}^{H-1}\left\{k(\frac{h-1}{H}) - 2k(\frac{h}{H}) + k(\frac{h+1}{H})\right\}^2 + \left\{k(\frac{H}{H}) - k(\frac{H-1}{H})\right\}^2$. □

Lemma 4 *Let $a_{n,h}$ be a non-stochastic array with $\sum_{h=1}^{H} a_{n,h}^2 \neq 0$, where H may depend on n. Then as $n \to \infty$*

$$\sum_{j=1}^{n}\zeta_{n,j} \xrightarrow{L} N(0,\omega^4), \quad \textit{where } \zeta_{n,j} = n^{-1/2}U_{j\delta}\sum_{h=1}^{H} a_{n,h}U_{(j-h)\delta} \Big/ \sqrt{\textstyle\sum_{h=1}^{H} a_{n,h}^2}.$$

Proof of Lemma 4. The array, $\left\{\zeta_{n,j}, \mathcal{F}_{n,j}\right\}$, is martingale difference when we set $\mathcal{F}_{n,j} = \sigma(U_{j\delta}, U_{(j-1)\delta}, \ldots)$. It can now be shown that

$$\sum_{j=1}^{n}\mathrm{E}\left(\zeta_{n,j}^2|\mathcal{F}_{n,j-1}\right) = \omega^2 n^{-1}\sum_{j=1}^{n}\left(\sum_{h=1}^{H} a_{n,h}^2 U_{(j-h)\delta}^2\right)\Big/\textstyle\sum_{h=1}^{H} a_{n,h}^2 \xrightarrow{p} \omega^4,$$

$$\sum_{j=1}^{n}\mathrm{E}\left\{\zeta_{n,j}^2\mathbf{1}(|\zeta_{n,j}| \geq \varepsilon)\right\} \leq \varepsilon^{-2}\sum_{j=1}^{n}\mathrm{E}|\zeta_{n,j}|^4 = O(n^{-1}),$$

where we used Minkowski's inequality. This verifies the two conditions of Billingsley (1995, theorem 35.12) and the result follows. □

Proof of Theorem 3. The results for $K(Y_\delta)$ and $K(Y_\delta, U_\delta) + K(U_\delta, Y_\delta)$ follows from Theorems 1 and 2. The only thing left to be shown is the asymptotic result for $K(U_\delta)$. We have

$$K(U_\delta) = -\sum_{h=1}^{H}\left(w_{h+1} - 2w_h + w_{h-1}\right)V_{h,n} - \sum_{h=1}^{H}(w_{h+1} - w_{h-1})R_{h,n}, \tag{A.3}$$

where $w_h = k(\frac{h-1}{H})$, $V_{h,n} = \frac{1}{2}\sum_{j=1}^{n}\left(U_{j\delta}U_{(j-h)\delta} + U_{j\delta}U_{(j+h)\delta} + U_{(j-1)\delta}U_{(j-1-h)\delta} + U_{(j-1)\delta}U_{(j-1+h)\delta}\right)$, and $R_{h,n} = \frac{1}{2}\left\{U_0\left(U_{-h\delta} - U_{h\delta}\right) + U_t\left(U_{t+h\delta} - U_{t-h\delta}\right)\right\}$. The last term in (A.3) is due to end-effects, and we have

$$\begin{aligned} H^{1/2}\sum_{h=1}^{H}(w_{h+1} - w_{h-1})R_{h,n} &= \tfrac{U_0}{H^{1/2}}\sum_{h=1}^{H}\tfrac{w_{h+1}-w_{h-1}}{2/H}\left(U_{-h\delta} - U_{h\delta}\right) \\ &\quad + \tfrac{U_t}{H^{1/2}}\sum_{h=1}^{H}\tfrac{w_{h+1}-w_{h-1}}{2/H}(U_{t+h\delta} - U_{t-h\delta}), \end{aligned} \tag{A.4}$$

which, conditionally on U_0 and U_t, converges in law to $MN\left\{0, \left(U_0^2 + U_t^2\right)2\omega^2 k_\bullet^{1,1}\right\}$.

Case: $k'(0)^2 + k'(1)^2 \neq 0$. With $H \propto n^{2/3}$ the end effect, (A.4), vanishes. Now rewrite $Hn^{-1/2}\sum_{h=1}^{H}\left(w_{h+1} - 2w_h + w_{h-1}\right)V_{h,n}$ as

$$\tfrac{w_2-w_1}{1/H}n^{-1/2}V_{1,n} - \tfrac{w_H-w_{H-1}}{1/H}n^{-1/2}V_{H,n} + \sum_{h=2}^{H-1}\left(\tfrac{w_{h+1}-w_h}{1/H} - \tfrac{w_h-w_{h-1}}{1/H}\right)n^{-1/2}V_{h,n}. \tag{A.5}$$

The result now follows from $n^{-1/2}V_{h,n} = 2n^{-1/2}\sum_{j=1}^{n}U_{j\delta}U_{(j-h)\delta} + o_p(1) \xrightarrow{L} N(0, 4\omega^4)$, $H(w_2 - w_1) \to k'(0)$, $H(w_H - w_{H-1}) \to k'(1)$, $\mathrm{Cov}(n^{-1/2}V_{h,n}, n^{-1/2}V_{l,n}) = 0$ for $h \neq l$, and the fact that last term of (A.5) vanishes in probability.

Case: $k'(0)^2 + k'(1)^2 = 0$. The contribution to the asymptotic variance from (A.4) is proportion $U_0^2 + U_t^2$. Since $\mathrm{E}\left\{\left(U_0^2 + U_t^2\right)^2\right\} = O(m^{-2})$ this term vanishes when $m \to \infty$. Next, set $a_{h,n} = 2H^{3/2}\left(w_{h+1} - 2w_h + w_{h-1}\right)$ so that

$$\left(n/H^3\right)^{-1/2}\sum_{h=1}^{H}\left(w_{h+1} - 2w_h + w_{h-1}\right)V_{h,n} = \sum_{j=1}^{n}n^{-1/2}U_{j\delta}\sum_{h=1}^{H}a_{n,h}U_{(j-h)\delta} + o_p(1).$$

The result now follows by Lemma 4 and the fact that $\sum_{h=1}^{n}a_{n,h}^2 \to 4k_\bullet^{2,2}$. □

Proof of Theorem 4. The convergence by the individual terms follow from Theorem 3, and the stable convergence for the sum of the three terms follows by Theorem 1. □

Proof of Proposition 1. The càdlàg property of υ follows by direct argument. Further, by Lebesgue's Theorem, the integral (20) is the same whether interpreted as a Riemann integral or a Lebesgue integral. With the latter interpretation we find $\Upsilon_t = \int_0^t \sigma_{T_u}^2\tau_u^2 \mathrm{d}u = \int_0^t \sigma_{T_u}^2 \mathrm{d}T_u = \int_0^{T_t}\sigma_u^2 \mathrm{d}u = S \circ T_t$. □

Proof of Proposition 2. With

$$a_{h,H} = H^{-1/2}\frac{k\left(\frac{h+1}{H}\right) - k\left(\frac{h-1}{H}\right)}{2/H}, \quad \text{so that } \sum_{h=1}^{H}a_{h,H}^2 \to k_\bullet^{1,1},$$

we see from (A.4), that the second term in the kernel representation (A.3) is $O_p(H^{-1/2})$. Consider now the first term in (A.3). We have $V_{h,n} = 2\sum_{i=1}^{n}U_{j\delta}U_{(j-h)\delta} + O_p(1)$, so the first term is ($-1$ times)

$$\sum_{h=1}^{H}\left\{k\left(\tfrac{h}{H}\right) - 2k\left(\tfrac{h-1}{H}\right) + k\left(\tfrac{h-2}{H}\right)\right\}V_{h,n} \;=\; \sum_{h=1}^{H}\left\{k\left(\tfrac{h}{H}\right) - 2k\left(\tfrac{h-1}{H}\right) + k\left(\tfrac{h-2}{H}\right)\right\}\left\{2n\bar{\gamma}_h + O_p(1)\right\}$$

$$= \frac{n}{H^2}\sum_{h=1}^{H} k''\left(\tfrac{h}{H}\right)\bar{\gamma}_h + O_p(\frac{n}{H^3}) + O_p(H^{-1}),$$

where we used the notation $\bar{\gamma}_h = n^{-1}\sum_{i=1}^{n} U_{i\delta}U_{(i-h)\delta}$. $\square$

Proof of Proposition 3. Let $\bar{\gamma}_h = n^{-1}\sum_{i=1}^{n} U_{i\delta}U_{(i-h)\delta}$. Since $k''(0) = 0$ we have

$$-\frac{n}{H^2}\sum_{|h|\le \bar{H}} k''\left(\frac{h}{H}\right)\bar{\gamma}_h = -\frac{n}{H^3}\sum_{|h|\le \bar{H}} k'''(0)\,|h|\,\bar{\gamma}_h + O_p\left(\frac{n}{H^3}\right) = O_p\left(\frac{n}{H^3}\right).$$

This leaves us thinking of $\sum_{H\ge|h|>\bar{H}} k''\left(\frac{h}{H}\right)\bar{\gamma}_h$. From Bartlett (1946) we know that for $k>h$

$$\sqrt{n}\left(\begin{array}{c}\bar{\gamma}_h - \mathrm{E}\bar{\gamma}_h\\ \bar{\gamma}_k - \mathrm{E}\bar{\gamma}_k\end{array}\right) \xrightarrow{L} N\left\{\left(\begin{array}{c}0\\0\end{array}\right), \omega^4\sum_{j=-\infty}^{\infty}\left(\begin{array}{cc}\rho_h^2 + \rho_{j+h}\rho_{j-h} & \rho_j\rho_{j+(k-h)} + \rho_{j+k}\rho_{j-h}\\ \rho_j\rho_{j+(k-h)} + \rho_{j+k}\rho_{j-h} & \rho_j^2 + \rho_{j+k}\rho_{j-k}\end{array}\right)\right\}$$

where ρ_j denotes the population autocorrelation. In the AR(1) case, with persistence parameter $|\varphi|<1$ then it is well known that this simplifies to

$$\sqrt{n}\left(\begin{array}{c}\bar{\gamma}_h - \varphi^h\omega^2\\ \bar{\gamma}_k - \varphi^k\omega^2\end{array}\right) \xrightarrow{L} N\left\{\left(\begin{array}{c}0\\0\end{array}\right), 2\omega^4\tfrac{1+\varphi^2}{1-\varphi^2}\left(\begin{array}{cc}1 & \varphi^{k-h}\\ \varphi^{k-h} & 1\end{array}\right)\right\},$$

noting $\sum_{j=-\infty}^{\infty}\varphi^{2j} = \left(1+\varphi^2\right)/\left(1-\varphi^2\right)$. Since $\lim_{H\to\infty}\sum_{H\ge h>\bar{H}} k''\left(\frac{h}{H}\right)\varphi^h = 0$, the impact of the serial dependence is that

$$\sqrt{\frac{n}{H}}\left\{\sum_{H\ge h>\bar{H}} 2k''\left(\tfrac{h}{H}\right)\bar{\gamma}_h\right\} \xrightarrow{L} N\left(0, 4\omega^4\tfrac{1+\varphi^2}{1-\varphi^2}k_\bullet^{2,2}\right).$$

This implies $-\frac{n}{H^2}\sum_{|h|\le\bar{H}} k''\left(\frac{h}{H}\right)\bar{\gamma}_h = O_p\left(\frac{n^{1/2}}{H^{3/2}}\right)$. Overall we have $O_p\left(n^{1/2}H^{-3/2}\right) + O_p(nH^{-3}) + O_p\left(H^{-1/2}\right)$. Placing $H \propto n^{1/2}$ delivers a term which is $O_p\left(n^{-1/4}\right)$. Since $|\varphi|<1$ we continue to have $U_0, U_t = O_p(m^{-1/2})$ with jittering, so the end-effects vanishes at the proper rate. $\square$