

DESIGNING REALIZED KERNELS TO MEASURE THE EX POST VARIATION OF EQUITY PRICES IN THE PRESENCE OF NOISE

BY OLE E. BARNDORFF-NIELSEN, PETER REINHARD HANSEN,
ASGER LUNDE, AND NEIL SHEPHARD¹

This paper shows how to use realized 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 equals that of the maximum likelihood estimator in the parametric version of this problem. Realized kernels can also be selected to (i) be analyzed using endogenously spaced data such as that in data bases on transactions, (ii) allow for market frictions which are endogenous, and (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, realized variance.

1. INTRODUCTION

IN THE LAST SIX YEARS the harnessing of high frequency financial data has led 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 realized variance (e.g., Andersen, Bollerslev, Diebold, and Labys (2001), Meddahi (2002), 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 realized 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), Andersen, Bollerslev, Diebold, and Labys (2000)). In this paper we study the class of *realized kernel* estimators of quadratic variation. We show how to design these estimators to be robust to certain types of frictions and to be efficient.

¹The second and fourth author are affiliated with CREATES, a research center funded by the Danish National Research Foundation. 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, anonymous referees, and the co-editor, Whitney Newey, for valuable comments. The Ox language of Doornik (2006) was used to perform the calculations reported here.

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 (2008), and the autoregressive filter of Bollen and Inder (2002) are estimators that use pre-whitening techniques; see also Bandi and Russell (2008). Aït-Sahalia, Mykland, and Zhang (2005) and Oomen (2005) proposed 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 owing 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 realized 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 realized kernel type. 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 realized kernels take on the form

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

Here the nonstochastic $k(x)$ for $x \in [0, 1]$ is a weight function and the h th realized autocovariance is

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

with $h = -H, \dots, -1, 0, 1, \dots, H$ 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 realized variance of X . Here $K(X_\delta) - \gamma_0(X_\delta)$ is the realized kernel correction to realized 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 an 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 minimize 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 realized 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 data bases on transactions (see Renault and Werker (2008) 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 realized kernels are related to so-called heteroskedastic autocorrelation (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 realized 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 realized 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) used realized 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. Analysis of the finite sample performance of realized kernels is provided by Bandi and Russell (2006).

In Section 2 we detail our notation and assumptions about the efficient price process, market frictions, and realized kernels. In Section 3 we give a central limit theory for $\gamma_h(X_\delta)$. Section 4 then looks at the corresponding properties of realized kernels. Here we also analyze the realized 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 have on our theory and extend the analysis of realized 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 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq T^*}, P)$, 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; 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

$$(1) \quad [Y]_t = \text{plim}_{n \rightarrow \infty} \sum_{j=1}^{t_j \leq t} (Y_{t_j} - Y_{t_{j-1}})^2$$

(e.g., [Protter \(2004, pp. 66–77\)](#) and [Jacod and Shiryaev \(2003, p. 51\)](#)) for any sequence of deterministic partitions $0 = t_0 < t_1 < \dots < t_n = T$ with $\sup_j \{t_{j+1} - t_j\} \rightarrow 0$ for $n \rightarrow \infty$. 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 ($Y \in \mathcal{BSM}$)

$$(2) \quad Y_t = \int_0^t a_u du + \int_0^t \sigma_u dW_u,$$

where a is a predictable locally bounded drift, σ is a cadlag 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

$$[Y]_t = \int_0^t \sigma_u^2 du.$$

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

$$(3) \quad \sigma_t = \sigma_0 + \int_0^t a_u^\# du + \int_0^t \sigma_u^\# dW_u + \int_0^t v_u^\# dV_u,$$

where $a^\#$, $\sigma^\#$, and $v^\#$ are adapted cadlag processes, with $a^\#$ also being predictable and locally bounded, and V is a Brownian motion independent of W . Moreover, σ^2 is assumed to be almost surely 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

$$(4) \quad X = Y + U,$$

and think of $Y \in \mathcal{BSM}$ as the efficient price. Our scientific interest will be in estimating $[Y]_t = \int_0^t \sigma_u^2 du$. 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 analyze data in thickly traded stocks recorded every minute. In [Section 5.5](#) we will show that realized 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

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

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 \(2005\)](#), [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 that our kernel can be made to be consistent 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 realized kernels.

2.3. Defining the Realized Autocovariation Process

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

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

$$h = -H, \dots, -1, 0, 1, 2, \dots, H.$$

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

$$(5) \quad \gamma_h(X_\delta) = \gamma_h(Y_\delta) + \gamma_h(Y_\delta, U_\delta) + \gamma_h(U_\delta, Y_\delta) + \gamma_h(U_\delta).$$

The daily increments of the realized QV, $\gamma_0(X_\delta)$, are called realized variances; their square roots are called the realized volatilities. Realized 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 realized 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 Realized Kernel

We study the realized kernel

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

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 superscript \top to denote a transpose:

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

2.5. Maximum Likelihood Estimator of QV

To put nonparametric results in context, it is helpful to have a parametric benchmark. In this subsection we recall the behavior 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 $\hat{\sigma}_{ML}^2$ and $\hat{\omega}_{ML}^2$ denote the ML estimators in this Gaussian model. Their asymptotic properties are given from classical results about the MA(1) process. By adopting the expression given in [Ait-Sahalia, Mykland, and Zhang \(2005, Proposition 1\)](#) to our notation, we have that for $\omega^2 > 0$,

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

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

Naturally, $\hat{\sigma}_{ML}^2$ may be interpreted as a quasi maximum likelihood estimator (QMLE) when the Gaussian model is misspecified. Interestingly, [Ait-Sahalia, Mykland, and Zhang \(2005\)](#) have shown that the asymptotic distribution of $\hat{\sigma}_{ML}^2$ does not depend on the actual distribution of U . Hence, if U has a non-Gaussian distribution, we continue to have $n^{1/4}(\hat{\sigma}_{ML}^2 - \sigma^2) \xrightarrow{L} N(0, 8\omega\sigma^3)$, even though $\hat{\sigma}_{ML}^2$ is derived under a Gaussian specification for U . 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 maximum likelihood 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 $\hat{\sigma}_{ML}^2$, since $n^{1/2}(\hat{\sigma}_{ML}^2 - \sigma^2) \xrightarrow{L} N(0, 6\sigma^4)$. When ω^2 is also known a priori to be zero, and so is not estimated, then

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

2.6. Notation and Jittering

To simplify the exposition of some results, we redefine the price measurements at the two endpoints, 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 discussion we consider kernel weight functions, $k(x)$, that are two times continuously differentiable on $[0, 1]$, and define

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

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 kernels as *smooth kernels*.

3. CENTRAL LIMIT THEORY FOR $\gamma(X_\delta)$

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

3.1. Background Result

Here we will study the large sample behavior 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{LY} will denote convergence in law stably with respect to the σ -field, $\sigma(Y)$, generated by the process Y , a type of convergence which will be discussed in the next subsection.

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*

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

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

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

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

$$\delta^{-1/2} \Gamma_{\delta,H} \xrightarrow{LY} \text{MN} \left(0, 2 \int_0^t \sigma_u^4 du \times A \right), \quad A = \text{diag}(1, 2, \dots, 2),$$

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

structure

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad B_{11} = \begin{pmatrix} 1 & \bullet \\ -1 & 2 \end{pmatrix},$$

$$B_{21} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \quad B_{22} = \begin{pmatrix} 2 & \bullet & \bullet & \bullet \\ -1 & 2 & \bullet & \bullet \\ \ddots & \ddots & \ddots & \bullet \\ \dots & 0 & -1 & 2 \end{pmatrix},$$

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

$$E\{\gamma(U_\delta)\} = 2\omega^2 n(1, -1, 0, 0, \dots, 0)^\top,$$

$$\text{Cov}\{\gamma(U_\delta)\} = 4\omega^4(nC + D + m^{-1}E),$$

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

$$C_{11} = \begin{pmatrix} 1 + \lambda^2 & \bullet \\ -2 - \lambda^2 & 5 + \lambda^2 \end{pmatrix},$$

$$C_{21} = \begin{pmatrix} 1 & -4 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad C_{22} = \begin{pmatrix} 6 & \bullet & \bullet & \bullet \\ -4 & 6 & \bullet & \bullet \\ 1 & -4 & 6 & \bullet \\ 0 & \ddots & \ddots & \ddots \end{pmatrix},$$

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

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

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

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

$$E_{22} = \begin{pmatrix} 8 & \bullet & \bullet & \cdots \\ -5 & 8 & \bullet & \ddots \\ 0 & -5 & 8 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

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

Using $\gamma_h(U_\delta) + \gamma_{-h}(U_\delta)$ in the construction of our realized kernels, rather than $2\gamma_h(U_\delta)$, say, is essential for obtaining a consistent estimator. The explanation is simple. Part of the realized 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 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 Appendix A) conceals a key property of stable convergence, which is a useful joint convergence. Let \mathcal{Y}_n denote a random variate on (Ω, \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 $\mathcal{Y}_n \xrightarrow{L_{\mathcal{G}}} \mathcal{Y}$, if and only if $(\mathcal{Y}_n, Z) \xrightarrow{L} (\mathcal{Y}, Z)$ for all \mathcal{G} -measurable random variables Z and some random variate \mathcal{Y} . When $\mathcal{G} = \sigma(X)$ we will write $\xrightarrow{L_X}$ in place of $\xrightarrow{L_{\mathcal{G}}}$.

Consider the simple example where

$$(10) \quad \mathcal{Y}_n = \delta^{-1/2} \left(\gamma_0(Y_\delta) - \int_0^t \sigma_u^2 du \right) \xrightarrow{L_{\mathcal{Y}}} \text{MN}(0, 2Z) \quad \text{and} \quad Z = \int_0^t \sigma_u^4 du.$$

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

$$(11) \quad \delta^{-1/2} \left(\gamma_0(Y_\delta) - \int_0^t \sigma_u^2 du \right) / \sqrt{\int_0^t \sigma_u^4 du} \xrightarrow{L} N(0, 2),$$

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 realized autocovariances is new. The first part of Theorem 1 implies that the simple kernel

$$\delta^{-1/2} \left(\gamma_0(Y_\delta) + \gamma_1(Y_\delta) + \gamma_{-1}(Y_\delta) - \int_0^t \sigma_u^2 du \right) \xrightarrow{L_Y} \text{MN} \left(0, 6 \int_0^t \sigma_u^4 du \right),$$

in the no-noise case. This increases the asymptotic variance by a factor of 3 relative to that in (10). So in the absence of noise, there will be no gains from realized 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 (2005), and Zhang, Mykland, and Ait-Sahalia (2005). Note that both the mean and variance of $\gamma_0(U_\delta)$ explode as $n \rightarrow \infty$. 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)) $E\{\gamma_0(U_\delta) + \gamma_1(U_\delta) + \gamma_{-1}(U_\delta)\} = 0$ and $\text{Var}\{\gamma_0(U_\delta) + \gamma_1(U_\delta) + \gamma_{-1}(U_\delta)\} = 4\omega^4(2n - 1.5)$.

The higher-order autocovariances, $h \geq 2$, are noisy estimates of zero as $E\{\gamma_h(U_\delta) + \gamma_{-h}(U_\delta)\} = 0$ and $\text{Var}\{\gamma_h(U_\delta) + \gamma_{-h}(U_\delta)\} \propto n$. 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 $\text{Var}\{K(U_\delta)\} \simeq (n/H^2)8\omega^4$ 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 realized kernel, and we have the following result.

THEOREM 2: *Write $w = (1, 1, k(\frac{1}{H}), \dots, k(\frac{H-1}{H}))^\top$. Then as H increases,*

$$\begin{aligned} w^\top Aw &= 2Hk_{\bullet}^{0,0} + O(1), \\ w^\top Bw &= H^{-1}k_{\bullet}^{1,1} + O(H^{-2}), \\ w^\top Cw &= \begin{cases} H^{-2}\{k'(0)^2 + k'(1)^2\} + O(H^{-3}), & \text{if } k'(0)^2 + k'(1)^2 \neq 0, \\ H^{-3}k_{\bullet}^{2,2} + O(H^{-4}), & \text{if } k'(0)^2 + k'(1)^2 = 0, \end{cases} \\ w^\top Dw &= -H^{-1}\frac{1}{2}k'(1)^2 + O(H^{-2}), \end{aligned}$$

$$w^\top Ew = H^{-1}k_{\bullet}^{1,1} + O(H^{-2}).$$

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

4. BEHAVIOR OF KERNELS

4.1. Core Result

In this section we derive the asymptotic behavior of arbitrary realized kernels. In Section 4.3 we derive a way to choose the number of terms to use in the kernel, which is indexed by ω^2 and $\int_0^t \sigma_u^4 du$. Subsequently, we provide estimators of these quantities, implying the feasible asymptotic distribution of the realized 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 the following result.

THEOREM 3: *As $n, H \rightarrow \infty$ and $H/n \rightarrow 0$,*

$$\begin{aligned} \sqrt{\frac{n}{H}} \left\{ K(Y_\delta) - \int_0^t \sigma_u^2 du \right\} &\xrightarrow{L_Y} \text{MN} \left(0, 4k_{\bullet}^{0,0} t \int_0^t \sigma_u^4 du \right), \\ \sqrt{H} \{ K(Y_\delta, U_\delta) + K(U_\delta, Y_\delta) \} &\xrightarrow{L_Y} \text{MN} \left(0, k_{\bullet}^{1,1} 8\omega^2 \int_0^t \sigma_u^2 du \right), \\ \sqrt{\frac{H^2}{n}} \{ K(U_\delta) \} &\xrightarrow{L} N[0, 4\omega^4 \{ k'(0)^2 + k'(1)^2 \}]. \end{aligned}$$

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

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

It is useful to define

$$\xi^2 = \omega^2 / \sqrt{t \int_0^t \sigma_u^4 du} \quad \text{and} \quad \rho = \int_0^t \sigma_u^2 du / \sqrt{t \int_0^t \sigma_u^4 du}$$

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-Schwarz inequality, we have $\rho \leq 1$. It is worth noting that the proof of Theorem 3 can be adapted to the case where $\{U_{j\delta}\}$ is heteroskedastic. This will only affect the terms in the asymptotic variance that involve ω^4 .

The large n and large H asymptotic variance of $K(X_\delta) - \int_0^t \sigma_u^2 du$ is

$$(12) \quad 4t \int_0^t \sigma_u^4 du \times \left[\frac{H}{n} k_{\bullet}^{0,0} + 2 \frac{k_{\bullet}^{1,1}}{H} \rho \xi^2 + n \left\{ \frac{k'(0)^2 + k'(1)^2}{H^2} + \frac{k_{\bullet}^{2,2}}{H^3} \right\} \xi^4 + \frac{k_{\bullet}^{1,1}}{Hm} \xi^4 \right].$$

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*

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

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

$$(14) \quad n^{1/4} \left\{ K(X_\delta) - \int_0^t \sigma_u^2 du \right\} \xrightarrow{L_Y} \text{MN} \left\{ 0, 4t \int_0^t \sigma_u^4 du (c_0 k_{\bullet}^{0,0} + c_0^{-1} 2k_{\bullet}^{1,1} \rho \xi^2 + c_0^{-3} k_{\bullet}^{2,2} \xi^4) \right\}.$$

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 $m \rightarrow \infty$ 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 to “jitter” 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 minimizes the asymptotic variance in (13) is

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

TABLE I
 PROPERTIES OF SOME $n^{1/6}$ FLAT-TOP REALIZED KERNELS

	$k(x)$	$k'(0)$	$k'(1)$	$k_{\bullet}^{0,0}$	$k_{\bullet}^{1,1}$	$k_{\bullet}^{2,2}$	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

^aThe 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 realized kernels in this class.

and the lower bound for the asymptotic variance is

$$\begin{aligned}
 (15) \quad & 4c^* \omega^{4/3} \left(t \int_0^t \sigma_u^4 du \right)^{2/3} [k_{\bullet}^{0,0} + c^{*-3} \{k'(0)^2 + k'(1)^2\}] \\
 & = 6c^* k_{\bullet}^{0,0} \omega^{4/3} \left(t \int_0^t \sigma_u^4 du \right)^{2/3}.
 \end{aligned}$$

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

Three flat-top cases of this setup are analyzed in Table I. The flat-top Bartlett kernel puts $k(x) = 1 - x$, the 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.

4.3. Special Cases With $n^{1/4}$

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

$$\begin{aligned}
 & 4t \int_0^t \sigma_u^4 du (ck_{\bullet}^{0,0}\xi + 2c^{-1}k_{\bullet}^{1,1}\rho\xi + c^{-3}k_{\bullet}^{2,2}\xi) \\
 & = \omega \left(t \int_0^t \sigma_u^4 du \right)^{3/4} \underbrace{4(ck_{\bullet}^{0,0} + 2c^{-1}\rho k_{\bullet}^{1,1} + c^{-3}k_{\bullet}^{2,2})}_{g(c)}.
 \end{aligned}$$

To determine the c that minimizes the asymptotic variance we simply minimize $g(c)$. 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

$$(16) \quad c^* = \sqrt{\rho \frac{k_{\bullet}^{1,1}}{k_{\bullet}^{0,0}} \left\{ 1 + \sqrt{1 + \frac{3d}{\rho}} \right\}}, \quad \text{where} \quad d = \frac{k_{\bullet}^{0,0}k_{\bullet}^{2,2}}{(k_{\bullet}^{1,1})^2}.$$

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

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

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 analyzed in Table II, 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 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 minimizes $k_{\bullet}^{2,2}$ within the class of smooth kernels, thus in general we have $k_{\bullet}^{2,2} \geq 12$. It is noteworthy that the realized kernel based on the cubic kernel has the same asymptotic distribution as the multiscale estimator. Naturally, minimizing $k_{\bullet}^{2,2}$ need not minimize g , and a well known kernel that has a smaller asymptotic variance is the flat-top Parzen kernel, which

TABLE II
PROPERTIES OF SOME $n^{1/4}$ FLAT-TOP REALIZED KERNELS^a

	$k(x)$	$k_{\bullet}^{0,0}$	$k_{\bullet}^{1,1}$	$k_{\bullet}^{2,2}$	c^*	g
Cubic kernel	$1 - 3x^2 + 2x^3$	0.371	1.20	12.0	3.68	9.04
5th order kernel	$1 - 10x^3 + 15x^4 - 6x^5$	0.391	1.42	17.1	3.70	10.2
6th order kernel	$1 - 15x^4 + 24x^5 - 10x^6$	0.471	1.55	22.8	3.97	12.1
7th order kernel	$1 - 21x^5 + 35x^6 - 15x^7$	0.533	1.71	31.8	4.11	13.9
8th order kernel	$1 - 28x^6 + 48x^7 - 21x^8$	0.582	1.87	43.8	4.31	15.7
Parzen	$\begin{cases} 1 - 6x^2 + 6x^3, & 0 \leq x \leq 1/2 \\ 2(1 - x)^3, & 1/2 \leq x \leq 1 \end{cases}$	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

^aThe 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 realized kernels in this class—8 being the parametric efficiency bound.

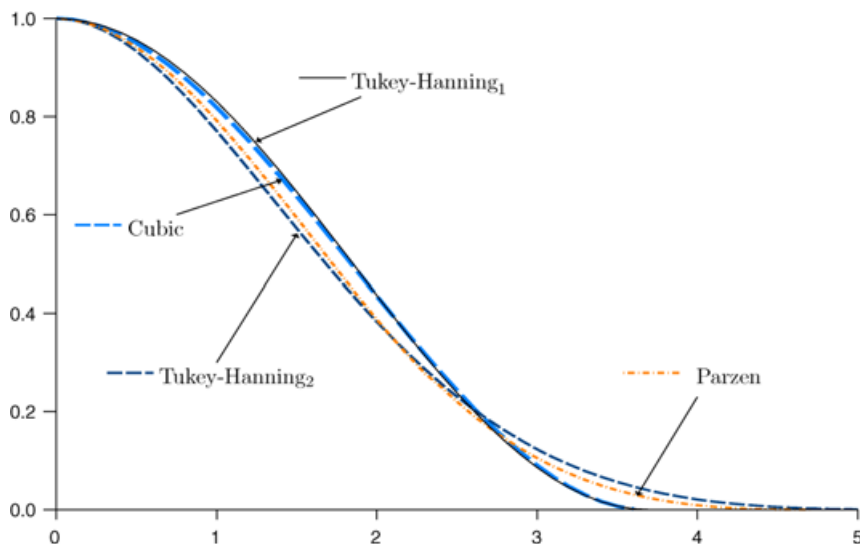


FIGURE 1.—Kernel functions, $k(x/c^*)$, scaled by their respective c^* to make them comparable.

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 the flat-top Tukey–Hanning $_p$ kernel, defined by

$$(17) \quad k(x) = \sin^2 \left\{ \frac{\pi}{2} (1-x)^p \right\}.$$

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

Table II 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, (17), when $p \geq 2$. While none of the standard kernels reaches the parametric efficiency bound, we see that the modified Tukey–Hanning kernel approaches the lower bound as we increase p . This kernel utilizes 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)).

4.4. Finite Sample Behavior

It is important to ask whether the approximation suggested by Theorem 3 and our special cases thereof provides a useful guide to finite sample behavior.

Table III gives

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

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 weights that were computed numerically.² More generally the table shows that the asymptotics provide 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.³ Of course, in that case the

TABLE III
FLAT-TOP REALIZED KERNELS^a

n	Opt.	TH ₂	Par	Cubic	Bart	Opt.	TH ₂	Par	Cubic	Bart
	$\xi^2 = 0.1$					$\xi^2 = 0.01$				
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
∞	8.00	8.29	8.54	9.04	∞	8.00	8.29	8.54	9.04	∞
	$\xi^2 = 0.001$					$\xi^2 = 0.0001$				
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
∞	8.00	8.29	8.54	9.04	∞	8.00	8.29	8.54	9.04	∞

^a $\text{Var}\{n^{1/4}K(X_\delta)\}\omega^{-1}(t \int_0^t \sigma_u^4 du)^{-3/4}$ listed against n . Asymptotic lower bound is 8. “Opt.” refers to kernel weights that were selected numerically to minimize the finite sample variance of a flat-top realized kernel. “Cubic” refers to $k(x) = 1 - 3x^2 + 2x^3$. “TH₂” denotes the modified Tukey–Hanning with $p = 2$; see (17).

²The weights were computed as $w^* = (1, 1, v)'$ where $v = -(n^{-1}A_{22} + 2\xi^2\rho B_{22} + n\xi^4 C_{22})^{-1}(2\xi^2\rho B_{21} + n\xi^4 C_{21})(1, 1)'$. The matrices A_{22} , B_{22} , C_{22} , B_{21} , and C_{21} are given in Theorem 1. These weights will, as $H \rightarrow \infty$, converge to those of the optimal kernel function that will be introduced in the next section.

³See Bandi and Russell (2006) for an analysis and comparison of the finite sample properties of a variety of estimators including the realised kernels.

realized 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 $w^\top Bw \simeq H^{-1}k_{\bullet,1}^{1,1}$, are inaccurate when H is small. So in our simulations and empirical analysis we compute the variance of $K(X_\delta)$ using the matrix expressions directly, rather than the asymptotic expressions of Theorem 2. This greatly improves the finite sample behavior 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 generalizes the theory in various directions and can be skipped during a first reading of the paper.

4.5. Realized 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 realized kernels can be expressed as

$$(18) \quad K(X_\delta) = \gamma_0(X_\delta) + \sum_{h=1}^{n-1} k\left(\frac{h-1}{H}\right) \{\gamma_h(X_\delta) + \gamma_{-h}(X_\delta)\},$$

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 nonzero weight to all autocovariances.⁴ So we replace the requirement “ $k(x) = 0$ for $x > 1$ ” with “ $k(x) \rightarrow 0$ as $x \rightarrow \infty$ 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 $k_{\bullet,0}^{0,0}$, $k_{\bullet,1}^{1,1}$, and $k_{\bullet,2}^{2,2}$ to represent integrals over the whole positive axis, for example, $k_{\bullet,0}^{0,0} = \int_0^\infty k(x)^2 dx$. We have the following optimality result within the class of realized kernels defined by (18).

PROPOSITION 1—Optimal Kernel Function: *The infinite-lag realized kernel with $H^* = \xi n^{1/2}$ and*

$$(19) \quad k(x) = (1+x)e^{-x}$$

achieves the parametric efficiency bound.

We will refer to $k(x) = (1+x)e^{-x}$ as the *optimal kernel function*. Proposition 1 shows that the realized kernel based on (19) is asymptotically first-order equivalent to the maximum likelihood estimator in the parametric version of the problem. That is, when $\sigma_u^2 = \sigma^2$ (constant volatility) they both have the

⁴It is only necessary to compute the first \tilde{H} realised autocovariances for some \tilde{H} with $H/\tilde{H} = o(1)$ because $k(x) \rightarrow 0$ exponentially fast as $x \rightarrow \infty$.

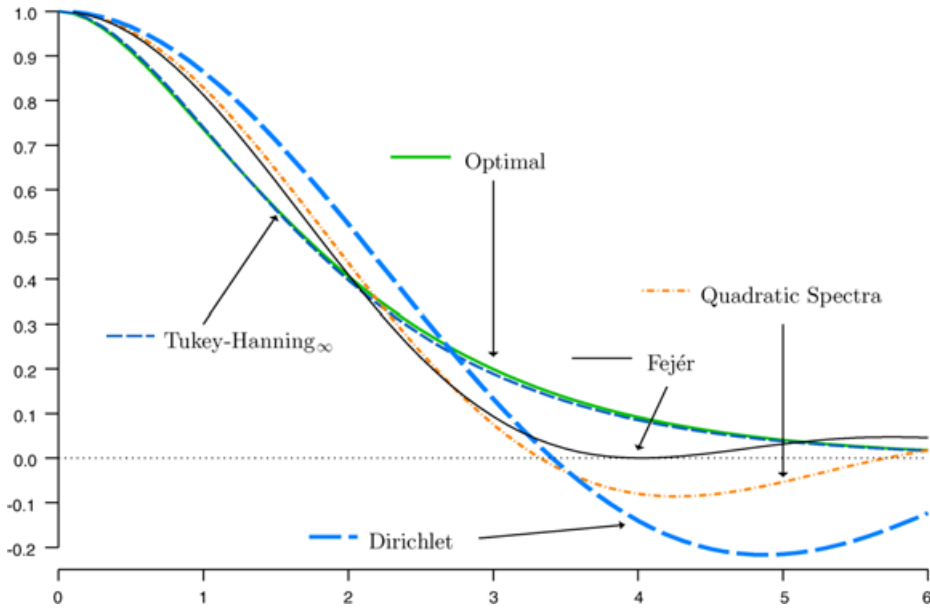


FIGURE 2.—Kernel functions, $k(x/c^*)$, scaled by their respective c^* to make them comparable.

asymptotic variance $8\omega\sigma^3$ (see (7)) and they both converge at rate $n^{1/4}$. An elegant feature of this kernel function, $k(x) = (1 + x)e^{-x}$, is that its optimal bandwidth is simply $H^* = \xi n^{1/2}$. So $c^* = 1$ for this kernel function. The proof of Proposition 1 (given in the Appendix) essentially amounts to solving a calculus of variation problem.

Five kernel functions for infinite-lag realized kernels are shown in Figure 2 and key statistics for these kernels functions are given in Table IV: first, the optimal kernel which achieves the parametric lower bound; second, the Tukey–Hanning $_\infty$ kernel that is given as the limit of (17) as $p \rightarrow \infty$ (this kernel is very similar to the optimal kernel); third, 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 realized kernel using the Dirichlet kernel weights,

$$\begin{aligned}
 k_N(z) &= \frac{1}{2N + 1} \frac{\sin\left(\left(N + \frac{1}{2}\right)z\right)}{\sin\left(\frac{1}{2}z\right)} \rightarrow k(x) \\
 &= \frac{\sin(x)}{x} \quad \text{as } N \rightarrow \infty, \quad x = \left(N + \frac{1}{2}\right)z.
 \end{aligned}$$

TABLE IV
 PROPERTIES OF SOME $n^{1/4}$ FLAT-TOP INFINITE-LAG REALIZED KERNELS^a

	$k(x)$	$k_{\bullet}^{0,0}$	$k_{\bullet}^{1,1}$	$k_{\bullet}^{2,2}$	c^*	g
Optimal	$(1+x)e^{-x}$	$\frac{5}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1.0000	8.0000
Tukey–Hanning _∞	$\sin^2\{\frac{\pi}{2}\exp(-x)\}$	0.52	$\frac{\pi^2}{16}$	$\frac{\pi^2(1+\pi^2)}{32}$	2.3970	8.0124
Quadratic spectral	$\frac{3}{x^2}(\frac{\sin x}{x} - \cos x)$	$\frac{3\pi}{5}$	$\frac{3\pi}{35}$	$\frac{\pi}{35}$	0.7395	9.3766
Dirichlet _∞	$\frac{\sin x}{x}$	$\frac{\pi}{2}$	$\frac{\pi}{6}$	$\frac{\pi}{10}$	1.0847	11.662
Fejér	$(\frac{\sin x}{x})^2$	$\frac{\pi}{3}$	$\frac{2\pi}{15}$	$\frac{16\pi}{105}$	1.2797	8.8927

^aThe g measures the relative asymptotic efficiency of the realized kernels in this class—8 being the parametric efficiency bound.

Mancino and Sanfelici (2007) introduced another variant of the Fourier estimator, which also has a realized 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 infinite-lag realized kernels is that they require a very large number of out-of-period intraday returns. The reason is that the h th autocovariance estimator needs h intraday returns before time 0 and after time t . Because the gains in precision from these estimators is relatively small, we will not utilize these “infinite-lag” estimators in our simulations and empirical analysis.

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 realized kernel becomes

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

where we assume $k(0) = 1$ and $k(1) = 0$. Now we have $E\{K(U_{\delta})\} = 2\omega^2 n\{1 - k(\frac{1}{H})\}$ and the bias is proportional to $(n/H)k'(0) + (n/H^2)k''(0)/2 + O(n/H^3)$. So the bias in the Bartlett case $k(x) = 1 - x$ is $O(n/H) = O(n^{1/3})$, and 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, $k(x) = 1 + ax^j + bx^{j+1} + cx^{j+2}$, $j = 3, 4, \dots$, yield $c = -(j + j^2)/2$, $b = 2j + j^2$,

and $a = -1 - 3j/2 - j^2/2$. Examples of this include

$$(20) \quad k(x) = \begin{cases} 1 - 10x^3 + 15x^4 - 6x^5, & j = 3, \\ 1 - 15x^4 + 24x^5 - 10x^6, & j = 4, \\ 1 - 21x^5 + 35x^6 - 15x^7, & j = 5, \\ 1 - 28x^6 + 48x^7 - 21x^8, & j = 6. \end{cases}$$

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 II shows that these estimators are less efficient than the realized kernels produced by (17).

Table V shows the corresponding finite sample behavior for some j th order realized kernels. In addition to the scaled variance, we also report the scaled squared bias

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

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

4.7. Small- ω^2 Asymptotic Analysis

Given that ω^2 is estimated to be small relative to the integrated variance, $\int_0^t \sigma_u^2 du$, it becomes interesting to analyze the realized 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 $0 \leq \alpha < 1$,⁵ and define $\xi_0^2 = \omega_0^2 / \sqrt{t \int_0^t \sigma_u^4 du}$.⁶ In this situation, the asymptotic variance is

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

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

⁵When $\alpha > 1$, the asymptotic analysis is essentially the same as $\omega^2 = 0$ —the case without noise. Note that as $\alpha < 1$, then $[U] = \infty$ so $U \notin \mathcal{SM}$.

⁶A similar local-to-zero asymptotics was used by Kalnina and Linton (2006) in the context of the two scale estimator.

TABLE V
FINITE SAMPLE VALUE OF $\text{Var}\{n^{1/4}K(X_\delta)\}\omega^{-1}(\int_0^t \sigma_u^4 du)^{-3/4}$ LISTED AGAINST n AND SCALED SQUARED BIAS FOR VARIOUS ORDER CASES WHEN $\rho = 1^a$

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
	5th Order Kernel						6th 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

^aIn the $n = 256$ case, when ξ^2 is very small H is selected to be zero and so the realized kernel becomes the RV.

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 $O(n^{-(1+\alpha)/2})$ 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 $O(n^{-(1+\alpha)/2-\alpha}m^{-1})$, which is lower than that of the other terms, even if m is constant. So jittering ($m \rightarrow \infty$) is not needed under this form of asymptotics. The rate of convergence is now slightly faster, for setting $H = c\xi_0 n^{(1-\alpha)/2}$ we have

$$(21) \quad n^{(1+\alpha)/4} \left\{ K(X_\delta) - \int_0^t \sigma_u^2 du \right\} \xrightarrow{L_Y} \text{MN} \left\{ 0, \omega_0 \left(4t \int_0^t \sigma_u^4 du \right)^{3/4} (ck^{0,0} + c^{-1}2\rho k^{1,1} + c^{-3}k^{2,2}) \right\},$$

which implies that 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 (21) it follows that the optimal c is the same as under the fixed- ω^2 asymptotics. Our estimator of $t \int_0^t \sigma_u^4 du$ 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 $\hat{\xi}^2 = \hat{\omega}^2/\widehat{IQ}_{\delta,S}$ 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 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

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

so that $n^{(1+\alpha)/2} \hat{\omega}$ is consistent for the appropriate asymptotic variance given in (21).

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

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

When $k'(0) = 0$ and we use the optimal rate, $H = c\xi_0 n^{(1-\alpha)/2}$, 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-toppedness or some other remedy.

When the kernel is “kinked,” in the sense that $k'(0)^2 + k'(1)^2 \neq 0$, one can show $H \propto n^{2/3(1-\alpha)}$ is optimal and that the best rate of convergence is $n^{(1+2\alpha)/6}$. 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 $\alpha > 0$, for example, 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 $\int_0^t \sigma_u^4 du$ 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 $\tilde{\omega}^2 = -\{\gamma_1(X_\delta) + \gamma_{-1}(X_\delta)\}/2n$, while, for example, Bandi and Russell (2008) suggested $\hat{\omega}^2 = \gamma_0(X_\delta)/2n$ which has a bias of $\int_0^t \sigma_u^2 du/2n$. Both estimators have their shortcoming, as $\tilde{\omega}^2$ may be negative and $\hat{\omega}^2$ can be severely biased because $\int_0^t \sigma_u^2 du/2n$ may be large relative to ω^2 . Using Theorem 1 we have that

$$\begin{aligned} \text{Var}\{n^{1/2}(\tilde{\omega}^2 - \omega^2)\} &= \omega^4(5 + \lambda^2), \\ \text{Var}\{n^{1/2}(\hat{\omega}^2 - \omega^2)\} &= \omega^4(1 + \lambda^2). \end{aligned}$$

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 nonnegative estimate. We have that

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

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

$$(22) \quad \check{\omega}^2 = \exp\{\log \hat{\omega}^2 - K(X_\delta)/\gamma_0(X_\delta)\}.$$

Note that $K(X_\delta)/\gamma_0(X_\delta)$ is an estimate of the relative bias of $\hat{\omega}^2$, which vanishes as $n \rightarrow \infty$, so that $\check{\omega}^2 - \hat{\omega}^2 \xrightarrow{p} 0$.

5.2. Estimation of Integrated Quarticity, $\int_0^t \sigma_u^4 du$

Estimating integrated quarticity reasonably efficiently is a tougher problem than estimating $[Y]$. When we wrote the first draft of this paper we did not know of any research which had solved this problem in the context with noise, so we introduced the method given below. However, we would like to point out a recent paper by Jacod, Li, Mykland, Podolskij, and Vetter (2007), who have introduced a novel pre-averaging method which could also be used here instead of our method.

Define the subsampled squared returns, for some $\tilde{\delta} > 0$, $x_{j,\cdot}^2 = \frac{1}{S} \times \sum_{s=0}^{S-1} (X_{\tilde{\delta}(j+s/S)} - X_{\tilde{\delta}(j-1+s/S)})^2$, $j = 1, 2, \dots, \tilde{n}$, where $\tilde{n} = \lfloor t/\tilde{\delta} \rfloor$. This allows us to define a bipower variation type estimator of integrated quarticity

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

The no-noise case of this statistic with no subsampling was introduced by Barndorff-Nielsen and Shephard (2004, 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 $\{X_{\tilde{\delta}}, \omega^2; S\}^{[2,2]}$ is approximately $72\omega^8\tilde{n}^3/S^2$, so that $\tilde{\delta}^{3/2}S \rightarrow \infty$ leads to consistency.⁷ 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 $t \int_0^t \sigma_u^4 du \geq (\int_0^t \sigma_u^2 du)^2$. This is useful as we have a very efficient estimator of $\int_0^t \sigma_u^2 du$. Thus our preferred way to estimate $t \int_0^t \sigma_u^4 du$ is

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

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 spelled out in, for example, Protter (2004, pp. 66–77) and Jacod and Shiryaev (2003, p. 51). It is important, likewise, to

⁷Let $\varepsilon_j = \frac{1}{S} \sum_{s=0}^{S-1} [(U_{\delta(j+s/S)} - U_{\delta(j-1+s/S)})^2 - 2\omega^2 + 2(U_{\delta(j+s/S)} - U_{\delta(j-1+s/S)})(Y_{\delta(j+s/S)} - Y_{\delta(j-1+s/S)})]$ and $R = [\sum_{j=1}^n (x_{j,\cdot}^2 - 2\omega^2)(x_{j-2,\cdot}^2 - 2\omega^2)] - \sum_{j=1}^n y_{j,\cdot}^2 y_{j-2,\cdot}^2 = \sum_{j=1}^n y_{j,\cdot}^2 \varepsilon_{j-2} + \sum_{j=1}^n \varepsilon_j y_{j-2,\cdot}^2 + \sum_{j=1}^n \varepsilon_j \varepsilon_{j-2}$. Then $R \simeq \sum_{j=1}^n \varepsilon_j (y_{j-2,\cdot}^2 + y_{j+2,\cdot}^2) + \sum_{j=1}^n \varepsilon_j \varepsilon_{j-2}$. Now $\text{Var}(\sum_{j=1}^n \varepsilon_j (y_{j-2,\cdot}^2 + y_{j+2,\cdot}^2) | Y) \simeq (12\omega^4/S) \sum_{j=1}^n (y_{j-2,\cdot}^2 + y_{j+2,\cdot}^2)^2 = O(n^{-1}S^{-1})$, so

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

This is enough to ensure the estimator converges to integrated quarticity as we know $\sum_j y_j^2 y_{j-2}^2$ will deliver the required quantity.

be able to derive central limit theorems (CLTs) 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 emphasized by Renault and Werker (2008) in both their theoretical and empirical work.

Let $Y \in \mathcal{BSM}$ and assume we have measurements at times $t_j = T_{\delta j}$, $j = 1, 2, \dots, n$, where $0 = t_0 < t_1 < \dots < t_n = T_1$ and where T is a stochastic process of the form $T_t = \int_0^t \tau_u^2 du$, with τ having strictly positive, cadlag sample paths. Then we can construct a new process $Z_t = Y_{T_t}$, so at the measurement times $Z_{\delta j} = Y_{T_{\delta j}}$, $j = 1, 2, \dots, n$. Performing the analysis on observations of Z made at equally spaced times then allows one to analyze irregularly spaced data on Y . The following argument shows that $Z \in \mathcal{BSM}$ with spot volatility $\sigma_{T_t} \tau_t$ and so the analysis is straightforward when we replace σ_t with $\sigma_{T_t} \tau_t$ in all our arguments. This is consistent with the central limit results in the realized variance case obtained by Mykland and Zhang (2006). In particular, the feasible CLT is implemented by recording data every five transactions, say, but then analyzing it as if the spacing had been equidistant.

Write $Z = Y \circ T$ and $S_t = \int_0^t \sigma_u^2 du$. 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 2: *Let $v_t = \sigma_{T_t} \tau_t$ and*

$$(23) \quad Y_t = \int_0^t v_u^2 du.$$

Then v is a cadlag process and $Y = S \circ T$.

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

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

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

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 extent that

$$(24) \quad \sum_{h=1}^H a_{h,H} U_{h\delta} = O_p(1) \quad \text{for any} \quad \sum_{h=1}^H a_{h,H}^2 = O(1).$$

So we have in mind a situation where the serial dependence is tied to the sampling frequency, δ , as opposed to calendar time. Thus, the asymptotic experiment is one where the dependence between U_s and U_t ($s \neq t$) vanishes as $\delta \rightarrow 0$, while $U_{h\delta}$ and $U_{(h-j)\delta}$ with j fixed may be highly dependent for any δ . A simple example of this is where $U_{j\delta}$ is a first-order moving average process in j with a temporal dependence parameter which is unaltered as δ changes.

PROPOSITION 3: *Suppose (24) holds. If $k'(0) = k'(1) = 0$, then*

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

With dependent noise, it is no longer true that $H^{-1/2} \sum_{h=1}^H k''(\frac{h}{H}) n^{-1/2} \times \sum_{i=1}^n U_{i\delta} U_{(i-h)\delta} \xrightarrow{L} N(0, k''^2 \omega^4)$. 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 4: *We assume $k''(0) = 0$, that $|k'''(0)| < \infty$, and that $U_{j\delta}$, $j = \dots, 0, 1, 2, \dots$, is an AR(1) process with persistence parameter φ , $|\varphi| < 1$. Then*

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

This means that

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

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

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

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 realized 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 realized kernels continue to be consistent when U is endogenous. We do this under a simple linear model of endogeneity,

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

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 realized kernels under exogenous noise can be used to study the impact of endogenous noise on realized 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 realized 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 explore the impact of jumps under some strong assumptions. Our arguments here are heuristic. 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 $D \perp\!\!\!\perp Y \perp\!\!\!\perp U$. 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 $h = \pm 1, \dots, \pm H$. We have established the asymptotic properties of $K(X_\delta) - \int_0^t \sigma_s^2 ds$, so the asymptotic properties of $K(R_\delta) - \int_0^t \sigma_s^2 ds - \sum_{j=1}^n d_j^2$ hinge on those of $L_\delta + M_\delta$. We assume $\sigma \perp\!\!\!\perp W$ and $a = 0$ to ease the exposition.

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

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

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

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

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

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

Then

$$\begin{aligned} \text{Var}(\sqrt{H} M_\delta | d_{\lfloor n\tau \rfloor}) &= H \omega^2 d_{\lfloor n\tau \rfloor}^2 \sum_{h=-H-1}^H \left\{ k\left(\frac{h-1}{H}\right) - k\left(\frac{h}{H}\right) \right\}^2 \\ &\rightarrow \omega^2 d_{\lfloor n\tau \rfloor}^2 2k_{\bullet}^{1,1}, \end{aligned}$$

so if $H \propto n^{1/2}$, the noise contributes to the asymptotic distribution, while if $H \propto n^{2/3}$, it does not. This suggests the realized 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 of this argument 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 realized kernel. A much more thorough analysis is provided in the working paper version of this paper (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2007)).

Before we turn our attention to feasible asymptotic distributions, we note that the working paper also reports on the accuracy of $K(X_\delta)$ as an estimator of $\int_0^t \sigma_u^2 du$ and $\widehat{Q}_{\delta,s}$ as an estimator of $t \int_0^t \sigma_u^4 du$. 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 working paper shows that this occurrence is extremely rare. In our simulations we generated millions of artificial samples and less than 25 of them resulted in a negative value for $K(X_\delta)$, using the Tukey–Hanning₂ weights.

In this short section our focus will be assessing the infeasible and feasible central limit theories for $K(X_\delta) - \int_0^t \sigma_u^2 du$. 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

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

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))

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

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 positive.

In the infeasible case, our simple rule-of-thumb for the choice of H is $H = c^* \omega \sqrt{n / \int_0^t \sigma_u^2 du}$, which immediately gives us ϖ . In practice this is less interesting than the feasible version, which puts $\hat{H} = c^* \hat{\omega} \sqrt{n / \gamma_0(\bar{X}_\delta)}$, where $\gamma_0(\bar{X}_\delta)$ is the realized 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 $K(\bar{X}_\delta)$, $\hat{\omega}^2$, and $\widehat{IQ}_{\delta,S}$ into our expression for ϖ , replacing $\int_0^1 \sigma_u^2 du$, ω^2 , and $\int_0^1 \sigma_u^4 du$, respectively. Monte Carlo results reported in the working paper suggest taking $S = \sqrt{n}$ in computing $\widehat{IQ}_{\delta,S}$.

6.2. Simulation Design

We focus on the Tukey–Hanning₂ 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 1 second to be $1/23,400$, so that $[0, 1]$ is thought to span 6.5 hours. The X process is generated using a Euler scheme based on $N = 23,400$ 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, for example, [Huang and Tauchen \(2005\)](#) and [Goncalves and Meddahi \(2004\)](#). The first is a 1-factor model (SV1F):

$$\begin{aligned} dY_t &= \mu dt + \sigma_t dW_t, & \sigma_t &= \exp(\beta_0 + \beta_1 \tau_t), \\ d\tau_t &= \alpha \tau_t dt + dB_t, & \text{corr}(dW_t, dB_t) &= \varphi. \end{aligned}$$

Here φ is a leverage parameter. To make the results comparable to our constant volatility simulations reported in our working paper we impose that $E(\sigma_t^2) = 1$ by setting $\beta_0 = \beta_1^2 / (2\alpha)$. We utilize the fact that the stationary distribution $\tau_t \sim N(0, -\frac{1}{2\alpha})$ 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, for example, [Hansen and Lunde \(2005\)](#).

We also consider a 2-factor SV model (SV2F):⁸

$$\begin{aligned} dY_t &= \mu dt + \sigma_t dW_t, & \sigma_t &= \text{s-exp}(\beta_0 + \beta_1 \tau_{1t} + \beta_2 \tau_{2t}), \\ d\tau_{1t} &= \alpha_1 \tau_{1t} dt + dB_{1t}, & d\tau_{2t} &= \alpha_2 \tau_{2t} dt + (1 + \phi \tau_{2t}) dB_{2t}, \\ \text{corr}(dW_t, dB_{1t}) &= \varphi_1, & \text{corr}(dW_t, dB_{2t}) &= \varphi_2. \end{aligned}$$

⁸The function $\text{s-exp}[x]$ is defined in the working paper ([Barndorff-Nielsen et al. \(2007\)](#)), which also gives a detailed description of our discretization scheme for the models.

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

Table VI 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 table 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.

Table VII 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 realized volatility where the same phenomenon is observed.

Table VIII shows 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.

To conserve space we only present one table with results for the 2-factor model. Table IX presents the results for the feasible theory based on our preferred \hat{Z}_n^{\log} statistic. For the 2-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 $t \int_0^t \sigma_s^4 ds / (\int_0^t \sigma_s^2 ds)^2$ tends to be larger in the 2-factor model than is the case for the 1-factor model. This makes the inequality $\widehat{\text{IQ}} \geq \{K(X_\delta)\}^2$ less valuable for the estimation of $t \int_0^t \sigma_s^4 ds$.

TABLE VI
SV1F: FINITE SAMPLE PROPERTIES OF Z_n ^a

No. Obs.	\bar{H}^*	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
$\xi^2 = 0.01$, number of reps. = 100,000									
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
1,560	23.0	-0.003	1.015	0.26	1.94	4.49	94.18	96.70	99.09
4,680	40.0	-0.006	1.009	0.31	2.12	4.68	94.49	96.99	99.20
7,800	51.0	-0.005	1.009	0.34	2.23	4.70	94.57	97.05	99.26
11,700	63.0	-0.005	1.010	0.36	2.22	4.76	94.57	97.06	99.28
23,400	88.0	-0.002	1.007	0.40	2.30	4.84	94.66	97.19	99.34
$\xi^2 = 0.001$, number of reps. = 100,000									
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
1,560	8.00	-0.005	1.002	0.33	2.10	4.59	94.54	97.09	99.28
4,680	13.0	-0.005	1.003	0.39	2.32	4.82	94.72	97.27	99.40
7,800	17.0	-0.003	1.005	0.39	2.31	4.80	94.75	97.21	99.37
11,700	20.0	-0.004	1.004	0.43	2.38	4.93	94.81	97.36	99.39
23,400	28.0	-0.002	1.001	0.46	2.42	4.88	94.87	97.39	99.41
$\xi^2 = 0.0001$, number of reps. = 100,000									
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
1,560	3.00	-0.005	1.002	0.38	2.28	4.74	94.73	97.22	99.39
4,680	4.00	-0.004	1.005	0.45	2.45	5.02	94.83	97.31	99.40
7,800	6.00	-0.002	1.003	0.44	2.33	4.84	94.81	97.31	99.41
11,700	7.00	-0.002	1.001	0.45	2.39	4.84	94.85	97.37	99.44
23,400	9.00	0.000	0.999	0.46	2.37	4.83	94.90	97.43	99.47

^aSummary 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 1-factor model (SV1F) and $K(X_\delta)$ is based on Tukey–Hanning₂ 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 realized kernel estimator on transaction log prices of General Electric (GE) shares carried out on the New York Stock Exchange (NYSE) 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 working paper (Barndorff-Nielsen et al. (2007)). We should note that the variance of the noise was around 10 times higher in 2000 than in 2004, so looking over both periods is instructive. The working paper also details the

TABLE VII
SV1F: FINITE SAMPLE PROPERTIES OF \hat{Z}_n^a

No. Obs.	\bar{H}^*	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
$\xi^2 = 0.01$, number of reps. = 100,000									
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
1,560	19.8	-0.136	1.015	1.60	4.67	7.65	97.38	99.10	99.94
4,680	33.6	-0.104	1.002	1.22	3.97	6.84	96.98	98.78	99.88
7,800	43.1	-0.091	1.000	1.11	3.72	6.58	96.73	98.67	99.83
11,700	52.7	-0.082	1.000	1.02	3.62	6.41	96.52	98.54	99.82
23,400	74.2	-0.067	0.997	0.87	3.39	6.20	96.44	98.41	99.78
$\xi^2 = 0.001$, number of reps. = 100,000									
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
1,560	8.79	-0.108	0.986	1.27	3.88	6.66	97.36	99.03	99.93
4,680	13.7	-0.078	0.985	0.97	3.46	6.10	96.73	98.66	99.84
7,800	17.1	-0.066	0.988	0.86	3.27	5.91	96.45	98.45	99.78
11,700	20.7	-0.061	0.988	0.82	3.22	5.95	96.40	98.44	99.77
23,400	28.8	-0.049	0.987	0.76	3.08	5.65	96.16	98.27	99.74
$\xi^2 = 0.0001$, number of reps. = 100,000									
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
1,560	5.28	-0.091	0.958	1.03	3.43	6.04	97.53	99.11	99.94
4,680	6.31	-0.058	0.965	0.75	3.00	5.57	96.80	98.68	99.84
7,800	7.18	-0.046	0.969	0.68	2.79	5.30	96.42	98.42	99.78
11,700	8.14	-0.040	0.972	0.65	2.76	5.28	96.29	98.34	99.75
23,400	10.5	-0.030	0.976	0.62	2.66	5.19	96.01	98.19	99.72

^aSummary statistics and empirical quantiles for the feasible statistic, \hat{Z}_n , that employs our estimate of the asymptotic variance, $\hat{\sigma}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0, 1)$. The simulation design is the 1-factor model (SV1F) and $K(X_\delta)$ is based on Tukey-Hanning₂ weights.

cleaning we carried out on the data before it was analyzed and the precise way we calculated all of our statistics.

Our realized 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.⁹ 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

⁹As our sample size is quite large, it is important to calculate it in tick time so as not to be influenced by the bias effect discussed by [Renault and Werker \(2008\)](#) caused by sampling in calendar time.

TABLE VIII
SV1F: FINITE SAMPLE PROPERTIES OF $\hat{Z}_n^{\log a}$

No. Obs.	\bar{H}^*	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
$\xi^2 = 0.01$, number of reps. = 100,000									
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
1,560	19.8	-0.055	0.980	0.42	2.40	5.09	95.67	97.82	99.59
4,680	33.6	-0.045	0.982	0.44	2.40	4.98	95.60	97.80	99.53
7,800	43.1	-0.040	0.985	0.44	2.40	5.01	95.53	97.77	99.52
11,700	52.7	-0.035	0.988	0.46	2.39	4.97	95.36	97.67	99.54
23,400	74.2	-0.029	0.989	0.44	2.42	4.98	95.44	97.71	99.50
$\xi^2 = 0.001$, number of reps. = 100,000									
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
1,560	8.79	-0.060	0.967	0.56	2.67	5.18	96.26	98.27	99.72
4,680	13.7	-0.044	0.976	0.53	2.60	5.10	95.91	98.07	99.67
7,800	17.1	-0.037	0.981	0.52	2.53	5.02	95.75	97.92	99.60
11,700	20.7	-0.035	0.982	0.52	2.61	5.16	95.73	97.93	99.60
23,400	28.8	-0.028	0.983	0.52	2.56	5.00	95.57	97.86	99.59
$\xi^2 = 0.0001$, number of reps. = 100,000									
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
1,560	5.28	-0.054	0.945	0.51	2.47	4.94	96.68	98.52	99.82
4,680	6.31	-0.034	0.959	0.50	2.44	4.86	96.21	98.27	99.69
7,800	7.18	-0.027	0.966	0.46	2.31	4.78	95.89	98.06	99.68
11,700	8.14	-0.024	0.970	0.47	2.37	4.80	95.87	98.02	99.63
23,400	10.5	-0.017	0.974	0.47	2.34	4.79	95.65	97.92	99.62

^aSummary statistics and empirical quantiles for the feasible statistic, \hat{Z}_n^{\log} , that employs our estimate of the asymptotic variance, $\hat{\omega}$, and $d = 10^{-6}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0, 1)$. The simulation design is the 1-factor model (SV1F) and $K(X_\delta)$ is based on Tukey–Hanning₂ weights.

microstructure effects, based on returns calculated over 20 minutes within each day. This baseline was chosen as Hansen and Lunde (2006) 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 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 data

TABLE IX
SV2F: FINITE SAMPLE PROPERTIES OF $\hat{Z}_n^{\log a}$

No. Obs.	\hat{H}^*	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
$\xi^2 = 0.01$, number of reps. = 100,000									
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
1,560	21.5	-0.119	1.085	1.04	4.20	7.74	94.62	96.98	99.11
4,680	36.6	-0.096	1.067	0.94	3.90	7.17	94.83	97.13	99.22
7,800	47.0	-0.083	1.059	0.87	3.71	6.87	94.71	97.21	99.31
11,700	57.4	-0.076	1.053	0.90	3.62	6.75	94.79	97.19	99.35
23,400	80.9	-0.065	1.043	0.86	3.46	6.38	95.03	97.32	99.37
$\xi^2 = 0.001$, number of reps. = 100,000									
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
1,560	9.77	-0.132	1.141	1.97	5.60	9.19	94.51	96.84	99.09
4,680	15.5	-0.098	1.106	1.60	4.83	8.15	94.64	97.07	99.22
7,800	19.5	-0.084	1.089	1.40	4.47	7.63	94.73	97.09	99.25
11,700	23.6	-0.078	1.078	1.30	4.31	7.34	94.77	97.24	99.34
23,400	32.9	-0.067	1.064	1.15	4.00	7.00	94.91	97.36	99.39
$\xi^2 = 0.0001$, number of reps. = 100,000									
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
1,560	5.54	-0.115	1.150	2.07	5.66	9.12	94.34	96.84	99.07
4,680	6.83	-0.075	1.118	1.64	4.85	8.00	94.22	96.81	99.11
7,800	7.91	-0.064	1.100	1.46	4.47	7.61	94.54	97.00	99.17
11,700	9.06	-0.058	1.087	1.29	4.25	7.27	94.46	97.05	99.26
23,400	11.8	-0.049	1.066	1.16	3.86	6.73	94.60	97.19	99.30

^aSummary statistics and empirical quantiles for the feasible statistic, $\hat{Z}_n^{\log a}$, that employs our estimate of the asymptotic variance, $\hat{\omega}$, and $d = 10^{-6}$. The empirical quantiles are benchmarked against those for the limit distribution, $N(0, 1)$. The simulation design is the 2-factor model (SV2F) and $K(X_\delta)$ is based on Tukey–Hanning₂ weights.

set. We should note that on all the days in 2004 our realized kernel estimator of the daily increments of $[Y]$ was positive.

Figure 3 shows daily 95% confidence intervals (CIs) for the realized kernel for November 2004 using the modified Tukey–Hanning₂ weights (17) with $H = c^* \hat{\xi}_n^{1/2}$. Also drawn are the corresponding results for the realized variance using the limit distribution for the logarithm of the realized variance, $\log\{\gamma_0(X_{20\min})\}$; see Barndorff-Nielsen and Shephard (2002). We can see the realized kernel has much shorter CIs. The width of these intervals does change through time, tending to be slightly wider in high volatility periods. Over the entire year there are only 3 days when the CIs do not overlap.

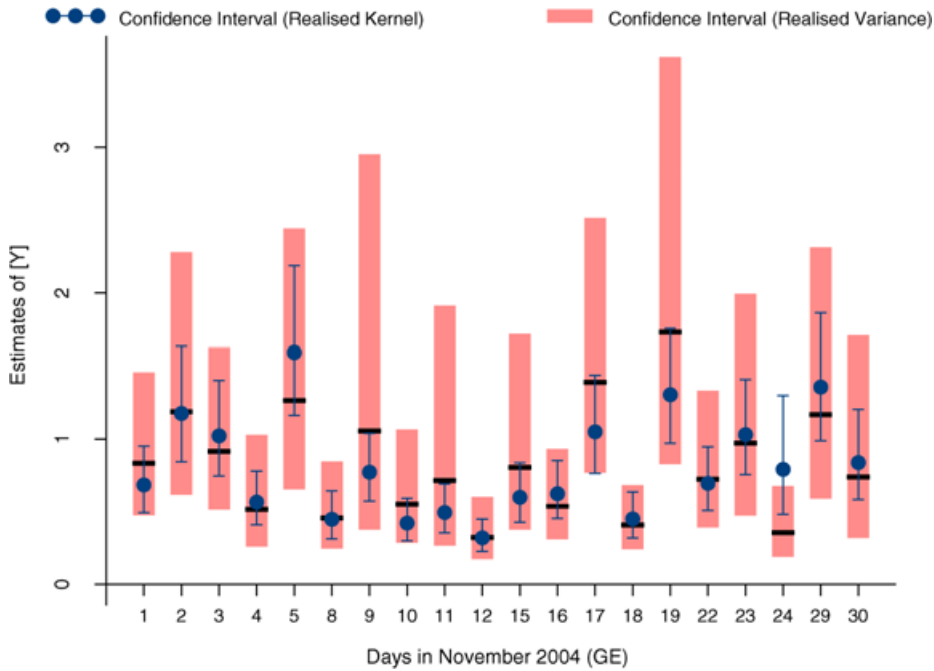


FIGURE 3.—Confidence intervals for the daily increments to $[Y]$ for General Electric (GE) in November 2004. Shaded rectangles denote the 95% confidence intervals based on 20 minute returns using the Barndorff-Nielsen and Shephard (2002) feasible realized variance inference method. The shorter intervals are for $K(X_\delta)$ based on Tukey–Hanning₂ weights, sampling in tick times so the period over which returns are calculated is roughly 60 seconds.

Table X 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 realized 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 XI provides summary statistics for some alternative estimators over the entire year. This suggests that the other realized kernel estimators have roughly the same average value and that they are quite tightly correlated. The table also records the summary statistics for the realized variance computed using 20, 5, and 1 minute and 10 and 1 second intervals. The last two of these estimators show a substantially higher mean. Interestingly, the realized 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 (2008). The realized kernels have a stronger degree of serial dependence

TABLE X
INFERENCE FOR GENERAL ELECTRIC (GE) VOLATILITY IN NOVEMBER 2004^a

Date	Trans	Lower	RV20m	Upper	n	Lower	KV60s	Upper	S	n	H	$\hat{\omega}^2$	$\tilde{\omega}^2$
1 Nov	4,631	0.48	0.83	1.46	20	0.49	0.69	0.95	13	357	5	0.0016	0.0007
2 Nov	4,974	0.62	1.19	2.28	20	0.84	1.17	1.64	14	356	6	0.0025	0.0011
3 Nov	4,918	0.51	0.92	1.63	20	0.75	1.02	1.40	14	352	5	0.0021	0.0008
4 Nov	5,493	0.26	0.52	1.03	20	0.41	0.57	0.78	16	344	5	0.0013	0.0005
5 Nov	5,504	0.65	1.26	2.44	20	1.16	1.59	2.19	16	344	5	0.0028	0.0009
8 Nov	4,686	0.25	0.46	0.85	20	0.31	0.45	0.64	14	335	6	0.0014	0.0008
9 Nov	4,923	0.38	1.05	2.95	20	0.57	0.77	1.04	14	352	4	0.0014	0.0004
10 Nov	4,970	0.29	0.55	1.07	20	0.30	0.42	0.59	14	355	6	0.0013	0.0007
11 Nov	4,667	0.27	0.71	1.91	20	0.35	0.49	0.69	13	359	5	0.0011	0.0004
12 Nov	4,822	0.17	0.32	0.60	20	0.23	0.32	0.45	14	345	6	0.0009	0.0005
15 Nov	4,681	0.38	0.80	1.72	20	0.43	0.60	0.84	14	335	5	0.0015	0.0007
16 Nov	4,526	0.31	0.54	0.93	20	0.45	0.62	0.85	13	349	5	0.0011	0.0004
17 Nov	5,477	0.77	1.39	2.51	20	0.76	1.05	1.44	16	343	5	0.0018	0.0006
18 Nov	4,738	0.24	0.41	0.68	20	0.32	0.45	0.64	14	339	6	0.0014	0.0007
19 Nov	5,224	0.83	1.73	3.62	20	0.97	1.31	1.76	15	349	4	0.0019	0.0005
22 Nov	5,359	0.39	0.72	1.33	20	0.51	0.69	0.95	15	358	5	0.0012	0.0004
23 Nov	5,405	0.47	0.97	1.99	20	0.75	1.03	1.41	15	361	5	0.0016	0.0005
24 Nov	4,626	0.19	0.36	0.68	20	0.48	0.79	1.30	13	356	5	0.0013	0.0004
29 Nov	4,709	0.59	1.17	2.31	20	0.99	1.36	1.86	14	337	5	0.0023	0.0007
30 Nov	4,719	0.32	0.74	1.71	20	0.58	0.84	1.20	14	338	6	0.0018	0.0007

^a“Trans” denotes the number of transactions. RV20m is the realized variance based on 20 minute returns and KV60s is $K(X_\delta)$ based on Tukey–Hanning₂ weights, and sampling of every S th transaction price, so the period over which returns are calculated is roughly 60 seconds.

than our benchmark realized variance that is based on a moderate sampling, so the period over which returns are calculated is 20 minutes. This point suggests the realized 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 realized 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 realized 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, 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 realized 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

TABLE XI

SUMMARY STATISTICS FOR SIX REALIZED KERNELS BASED ON RETURNS MEASURED EVERY S TH TRANSACTION, WHERE S IS SELECTED SUCH THAT RETURNS SPAN 60 SECONDS ON AVERAGE^a

δ	Average	Std. (HAC)	$\widehat{\text{Corr}}(\cdot, \text{TH}_2)$	acf(1)	acf(2)	acf(5)	acf(10)
<i>Tukey–Hanning₂ kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.908	0.541 (1.168)	1.000	0.34	0.38	0.28	0.09
<i>Parzen kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.914	0.546 (1.182)	0.999	0.35	0.37	0.28	0.09
<i>Cubic kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.915	0.542 (1.172)	0.998	0.35	0.37	0.28	0.09
<i>5th order kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.919	0.530 (1.160)	0.999	0.36	0.39	0.29	0.09
<i>8th order kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.912	0.550 (1.185)	0.995	0.34	0.38	0.28	0.09
<i>Bartlett kernel</i> ($H = c\xi^2 n^{2/3}$)							
≈ 1 minute	0.934	0.551 (1.192)	0.988	0.36	0.35	0.27	0.08
<i>Simple realized variance</i> = $\gamma_0(X_\delta)$							
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

^aThe same statistics are computed for the realized variance using five different values for δ . The realized variance based on all tick-by-tick data is identical to the realized variance with $\delta = 1$ second. The empirical correlations between the realized Tukey–Hanning₂ kernel and each of the estimators are given in column 4 and some empirical autocorrelations are given in columns 5–8.

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 realized kernel using every transaction on each day, based on returns sampled roughly every 60 seconds, or by applying the kernel weights to returns sampled every transaction. 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 realized 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.

Figure 4 shows that realized kernels give very similar estimates—on some days the estimates are almost identical. The two scale estimators by Zhang, Mykland, and Ait-Sahalia (2005) are quite biased because they rely on the white noise assumption, which is at odds with tick-by-tick data. The two scale estimators by Ait-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 discussed in Section 5.1. The bias is over-

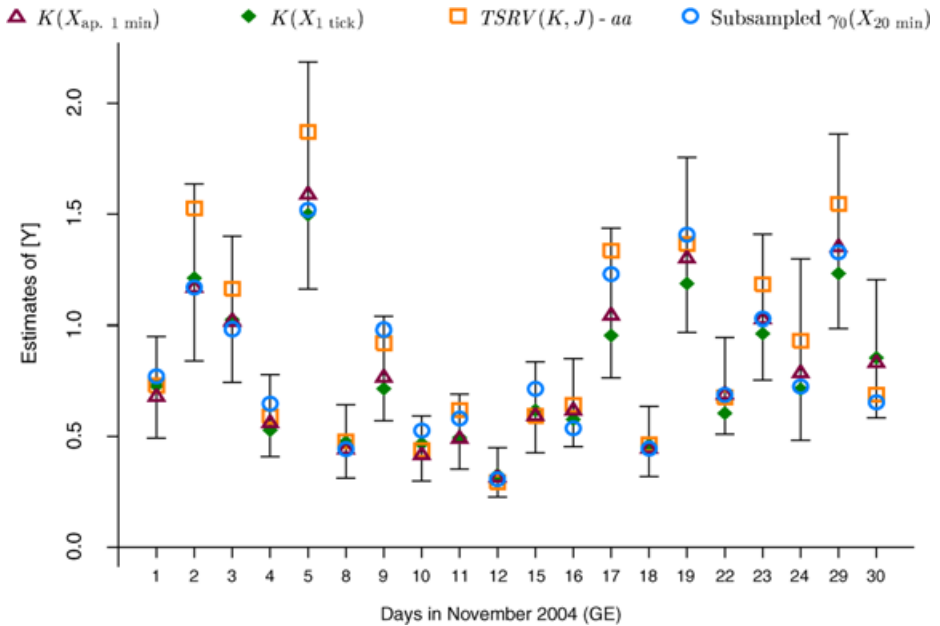


FIGURE 4.—Four estimators for the daily increments to $[Y]$ for General Electric in November 2004. The intervals are the confidence intervals for $K(X_\delta)$ using Tukey–Hanning₂ weights and returns sampled roughly every 60 seconds. Triangles denote averages of S distinct realized kernels, where each of the estimators is based on returns that span S transactions. Diamonds represent $K(X_\delta)$ using $\delta = 1$ tick. Circles are averages of 1,200 realized variances using $\delta = 20$ minutes, where the individual RVs are obtained by shifting the times prices are recorded by 1 second at a time. Squares represent the bias adjusted two scale estimator; see Ait-Sahalia, Mykland, and Zhang (2006, eq. 4.22).

come by the “area adjusted” version of this estimator, which seems in line with the results for the realized kernels and the subsampled random variable (RV) estimators.¹⁰ Table XII provides summary statistics of these estimators. The realized 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 realized kernels as estimators of quadratic variation when an efficient price is obscured by simple market frictions. We show how to make these estimators consistent

¹⁰In empirical work we found this estimator to be sensitive to the choice of K . To be consistent with our empirical findings, J has to be about 15 (yielding returns measured roughly over 1 minute). The estimator in Ait-Sahalia, Mykland, and Zhang (2006) needs $K > J$ and the theory relies on $K/J \rightarrow \infty$. 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.

TABLE XII

TWELVE ESTIMATORS THAT MAKE USE OF ALL AVAILABLE (TICK-BY-TICK) DATA BENCHMARKED AGAINST THE REALIZED KERNEL IN THE FIRST ROW, WHICH IS BASED ON RETURNS THAT SPAN 1 MINUTE ON AVERAGE^a

δ	Average	Std. (HAC)	$\widehat{\text{Corr}}(\cdot, \text{TH}_2)$	acf(1)	acf(2)	acf(5)	acf(10)
<i>Tukey–Hanning₂ kernel</i> ($H = c\xi n^{1/2}$)							
≈ 1 minute	0.908	0.541 (1.168)	1.000	0.34	0.38	0.28	0.09
<i>Tukey–Hanning₂ kernel</i> (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
<i>Parzen kernel</i> (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
<i>Cubic kernel</i> (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
<i>5th order kernel</i> (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
<i>8th order kernel</i> (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
<i>Subsampled realized variance</i>							
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
<i>ZMA 2005 TSRV</i> ($K, 1$)							
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
<i>AMZ 2006 TSRV</i> (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

^aThe realized kernels using $\delta = 1$ tick apply a large H to counteract possible dependence in U . The subsampled realized variances make use of all available data by changing the initial place at which prices are recorded and average the resulting estimates. For example, when $\delta = 5$ minutes, the subsampled RV is an average of 300 estimates, obtained by shifting the times at which prices are recorded by 1 second at a time. The four two scale estimators are those of Zhang, Mykland, and Ait-Sahalia (2005, equations 55 and 64) and Ait-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.

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 be 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 behavior. The realized 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 realized kernel is compared with that for a

simpler realized 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). The economic value of such forecast improvements has been documented in Bandi, Russell, and Yang (2007).

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

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 (Ω, \mathcal{F}, P) and taking values in a Polish space (E, \mathcal{E}) ,¹¹ 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 \mathcal{G} -measurable random variable¹² V on (Ω, \mathcal{F}, P) and every bounded and continuous function f on E we have that, for $n \rightarrow \infty$,

$$(A.1) \quad E\{Vf(\mathcal{X}_n)\} \rightarrow \int V(\omega)f(x)\mu(d\omega, dx).$$

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 $\mathcal{X}_n \xrightarrow{L_{\mathcal{G}}} \mathcal{X}$ 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 formalization does not reveal the key property of stable convergence which is that convergence \mathcal{G} -stably in law to X is equivalent¹³ to

¹¹That is, E is a complete separable metric space and \mathcal{E} denotes the Borel σ -algebra of E .

¹²As 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 Fraktur to denote general random variates.

¹³See Jacod and Shiryaev (2003, Proposition 5.33, pp. 512–518).

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 (Ω, \mathcal{F}, P) and with values in (E, \mathcal{E}) , and suppose that $\mathcal{Y}_n \xrightarrow{L\mathcal{G}} \mathcal{Y}$ for some sub- σ -field \mathcal{G} of \mathcal{F} .

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

PROOF: By the definition of \mathcal{G} -stable convergence, we have that $(W, \mathcal{Y}_n) \xrightarrow{L} (W, \mathcal{Y})$ and therefore that $W\mathcal{Y}_n \xrightarrow{L} W\mathcal{Y}$. Since $W_n\mathcal{Y}_n = (W_n/W)W\mathcal{Y}_n$ and $W_n/W \xrightarrow{P} 1$, we have $W_n\mathcal{Y}_n \xrightarrow{L} W\mathcal{Y}$. Q.E.D.

LEMMA 2: *If E is a normed vector space and if $\mathcal{Z}_n \xrightarrow{P} 0$, then $\mathcal{Y}_n + \mathcal{Z}_n \xrightarrow{L\mathcal{G}} \mathcal{Y}$.*

PROOF: The proof follows simply from the defining condition (A.1), using the tightness of $\{\mathcal{Y}_n\}$. Q.E.D.

LEMMA 3: *If $\mathcal{Z}_n \xrightarrow{L} \mathcal{Z}$ and if $\{\mathcal{Y}_n\}$ and $\{\mathcal{Z}_n\}$ are independent, then $\mathcal{Y}_n + \mathcal{Z}_n \xrightarrow{L\mathcal{G}} \mathcal{Y} + \mathcal{Z}$.*

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

DEFINITION 5: Let $\{Z_n\}$ be a sequence of d -dimensional random vectors on (Ω, \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 (Ω, \mathcal{F}, P)) such that for all $\zeta \in \mathbb{R}^d$,

$$(A.2) \quad \phi_n(\zeta|\mathcal{G}) \xrightarrow{P} \phi(\zeta|\mathcal{G}),$$

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

REMARK: We can, without restriction, assume that there exists a d -dimensional random vector Z , defined on (Ω, \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 5: 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 $n \rightarrow \infty$, it must be verified that $E\{e^{i\eta^\top Y_n + i\zeta^\top Z_n + i\psi W}\} \rightarrow E\{e^{i\eta^\top Y + i\zeta^\top Z + i\psi W}\}$. Now,

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

By (A.2),

$$\begin{aligned} |E[e^{i\eta^\top Y_n + i\psi W} \{\phi_n(\zeta|\mathcal{G}) - \phi(\zeta|\mathcal{G})\}]| &\leq E\{|\phi_n(\zeta|\mathcal{G}) - \phi(\zeta|\mathcal{G})|\} \\ &\rightarrow 0. \end{aligned}$$

Moreover, since $Y_n \xrightarrow{L\mathcal{G}} Y$, it holds that $E\{e^{i\eta^\top Y_n + i\psi W} \phi(\zeta|\mathcal{G})\} - E\{e^{i\eta^\top Y + i\psi W} \phi(\zeta|\mathcal{G})\} \rightarrow 0$. Q.E.D.

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 $\delta^{1/2} \{\gamma(U_\delta) - E\gamma(U_\delta)\}$. Consider $\gamma(Y_\delta) = \delta \sum_{j=1}^n f(\Delta_{j-H}^n Y / \sqrt{\delta}, \dots, \Delta_{j+H}^n Y / \sqrt{\delta})$, where $\Delta_j^n Y = Y_{j\delta} - Y_{(j-1)\delta}$ and $f(y_{j-H}, \dots, y_j, \dots, y_{j+H}) = \{y_j^2, y_j(y_{j-1} + y_{j+1}), \dots, y_j(y_{j-H} + y_{j+H})\}^\top$. Since $f(a, \dots, z) = f(-a, \dots, -z)$, we can apply Jacod (2007, Theorem 7.1.a) to

establish the stable convergence for $\gamma(Y_\delta)$; see also [Kinnebrock and Podolskij \(2008\)](#). The asymptotic variance is deduced from $\sum_{l=-2H}^{2H} E(\mathbf{f}_0 \mathbf{f}_l^\top) - (4H + 1)E(\mathbf{f}_0)E(\mathbf{f}_0^\top) = \sigma^4 \text{diag}(2, 4, \dots, 4)$, where we define $\mathbf{f}_l = f(Z_l, \dots, Z_{2H+l})$ with $\{Z_j\} \sim \text{i.i.d. } N(0, \sigma^2)$. This result and [Jacod \(2007, eq. 7.1 and 7.3\)](#) yield the asymptotic variance, $2 \int_0^t \sigma_u^4 du \times A$.

In considering the cross-term $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$, let $\bar{\gamma}(Y_\delta, U_\delta) = \gamma_0(Y_\delta)^{-1/2} \gamma(Y_\delta, U_\delta)$ and $\bar{y}_j = \gamma_0(Y_\delta)^{-1/2} \Delta_j^n Y$, where

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

and the coefficients \bar{y}_j are uniformly asymptotically negligible for almost all realizations of Y . Standard central limit theory then yields that, conditionally on Y , $\bar{\gamma}(Y_\delta, U_\delta) \xrightarrow{L} N(0, 2\omega^2 B)$. Since $\gamma(Y_\delta, U_\delta) - \gamma(U_\delta, Y_\delta) = o_p(1)$ and $\gamma_0(Y_\delta) \xrightarrow{p} [Y]_t$, this implies that almost surely

$$\gamma_0(Y_\delta)^{-1/2} (\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)) \xrightarrow{L} N(0, 8\omega^2 B).$$

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

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

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

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

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

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

We have $\text{Var}(V_0) = (n - 1)\lambda^2\omega^4$ and $\text{Var}(V_h) = (n - h - 1)\omega^4$ for $h \geq 1$, and similarly $\text{Var}(W_0) = 0$ and $\text{Var}(W_h) = 2\omega^4(h - 1)$ for $h \geq 1$. It is now straightforward to show that

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

where

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

$$D_{21}^V = \begin{pmatrix} -2 & 10 \\ 0 & -3 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}, \quad D_{21}^W = \begin{pmatrix} 0 & -1 \\ 0 & \frac{1}{2} \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix},$$

$$D_{22}^V = \begin{pmatrix} -18 & \bullet & \bullet & \bullet & \bullet \\ 14 & -24 & \bullet & \bullet & \bullet \\ -4 & 18 & -30 & \bullet & \bullet \\ 0 & -5 & 22 & -36 & \bullet \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

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 $\text{Var}(U_0^2) = m^{-3}\lambda^2\omega^4 + 4m^{-4}\frac{(m-1)m}{2}\omega^4 = m^{-3}(\lambda^2 - 2)\omega^4 + 2m^{-2}\omega^4$ and

$\text{Var}(U_0U_{h\delta}) = m^{-1}\omega^4$ for $j \neq 0$, and similarly for $U_tU_{t-h\delta}$ terms. So

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

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

Finally we verify that jointly $\delta^{-1/2}\Gamma_{\delta,H}$, $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ and $\delta^{1/2}\{\gamma(U_\delta) - E\gamma(U_\delta)\}$ converge $\sigma(Y)$ -stably. On account of Proposition 5, we can conclude that the joint law of $\delta^{-1/2}\Gamma_{\delta,H}$ and $[Y_\delta]_t^{-1/2}(\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta))$ 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 converge $\sigma(Y)$ -stably. Finally, since $\delta^{1/2}\{\gamma(U_\delta) - E\gamma(U_\delta)\} \xrightarrow{L} N(0, 4\omega^2C)$, a further application of Proposition 5 gives that $\delta^{-1/2}\Gamma_{\delta,H}$, $\gamma(Y_\delta, U_\delta) + \gamma(U_\delta, Y_\delta)$ and $\delta^{1/2}\{\gamma(U_\delta) - E\gamma(U_\delta)\}$ are jointly $\sigma(Y)$ -stably convergent. Q.E.D.

PROOF OF THEOREM 2: $w^\top Aw = 2k^{0,0}$ follows from the diagonal structure of A . Next $w^\top Bw = \sum_{h=1}^{H-1} \left\{ k\left(\frac{h}{H}\right) - k\left(\frac{h-1}{H}\right) \right\}^2$ gives the second result. The third result follows from

$$\begin{aligned} w^\top Cw &= \left\{ k\left(\frac{0}{H}\right) - 2k\left(\frac{1}{H}\right) + k\left(\frac{1}{H}\right)^2 \right\} \\ &\quad + \sum_{h=1}^{H-1} \left\{ k\left(\frac{h-1}{H}\right) - 2k\left(\frac{h}{H}\right) + k\left(\frac{h+1}{H}\right) \right\}^2 \\ &\quad + \left\{ k\left(\frac{H}{H}\right) - \left(k\left(\frac{H-1}{H}\right) \right)^2 \right\}. \end{aligned}$$

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

$$\begin{aligned} -w^\top D^V w &= 0 + 2 \left\{ k\left(\frac{1}{H}\right) - k\left(\frac{0}{H}\right) \right\}^2 \\ &\quad + (H+2) \left\{ k\left(\frac{H}{H}\right) - k\left(\frac{H-1}{H}\right) \right\}^2 \\ &\quad + \sum_{h=1}^{H-1} (h+2) \left\{ k\left(\frac{h-1}{H}\right) - 2k\left(\frac{h}{H}\right) + k\left(\frac{h+1}{H}\right) \right\}^2, \end{aligned}$$

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

The last result follows from $w^\top E w = \{k(\frac{1}{H}) - k(\frac{0}{H})\}^2 + \sum_{h=1}^H \{k(\frac{h}{H}) - k(\frac{h-1}{H})\}^2 + \sum_{h=1}^{H-1} \{k(\frac{h-1}{H}) - 2k(\frac{h}{H}) + k(\frac{h+1}{H})\}^2 + \{k(\frac{H}{H}) - k(\frac{H-1}{H})\}^2$. *Q.E.D.*

LEMMA 4: Let $a_{n,h}$ be a nonstochastic array with $\sum_{h=1}^H a_{n,h}^2 \neq 0$, where H may depend on n . Then as $n \rightarrow \infty$,

$$\sum_{j=1}^n \zeta_{n,j} \xrightarrow{L} N(0, \omega^4),$$

where

$$\zeta_{n,j} = n^{-1/2} U_{j\delta} \sum_{h=1}^H a_{n,h} U_{(j-h)\delta} / \sqrt{\sum_{h=1}^H a_{n,h}^2}.$$

PROOF: The array $\{\zeta_{n,j}, \mathcal{F}_{n,j}\}$ is a martingale difference when we set $\mathcal{F}_{n,j}$ of $\sigma(U_{j\delta}, U_{(j-1)\delta}, \dots)$. It can now be shown that

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

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

where we used Minkowski’s inequality. This verifies the two conditions of Billingsley (1995, Theorem 35.12) and the result follows. *Q.E.D.*

PROOF OF THEOREM 3: The result for $K(Y_\delta)$ follows from Theorems 1 and 2 extended to the case where $H \propto n^\gamma$ for $\gamma = 1/2$ and $\gamma = 2/3$. See Jacod (2008) for a general treatment of results of this kind. Next, for the cross-term, $K(Y_\delta, U_\delta) + K(U_\delta, Y_\delta)$, we can use the same trick as in the proof of Theorem 1, by deriving the limit distribution $\gamma_0(Y_\delta)^{-1/2} H^{1/2} K(Y_\delta, U_\delta) \xrightarrow{L} N(0, 2k^{1,1} \omega^2)$ (for almost all Y) and using $K(Y_\delta, U_\delta) - K(U_\delta, Y_\delta) = o_p(H^{-1/2})$. The last term is more involved. We have

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

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

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

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

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

$$\begin{aligned} (A.5) \quad & \frac{w_2 - w_1}{1/H} n^{-1/2} V_{1,n} - \frac{w_H - w_{H-1}}{1/H} n^{-1/2} V_{H,n} \\ &+ \sum_{h=2}^{H-1} \left(\frac{w_{h+1} - w_h}{1/H} - \frac{w_h - w_{h-1}}{1/H} \right) n^{-1/2} V_{h,n}. \end{aligned}$$

The result now follows from $n^{-1/2}V_{h,n} = 2n^{-1/2} \sum_{j=1}^n U_{j\delta}U_{(j-h)\delta} + o_p(1) \xrightarrow{L} N(0, 4\omega^4)$, $H(w_2 - w_1) \rightarrow k'(0)$, $H(w_H - w_{H-1}) \rightarrow k'(1)$, $\text{Cov}(n^{-1/2}V_{h,n}, n^{-1/2}V_{l,n}) = 0$ 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 proportional to $U_0^2 + U_t^2$. Since $E\{(U_0^2 + U_t^2)^2\} = O(m^{-2})$, this term vanishes when $m \rightarrow \infty$. Next, set $a_{h,n} = 2H^{3/2}(w_{h+1} - 2w_h + w_{h-1})$ so that

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

The result now follows by Lemma 4 and the fact that $\sum_{h=1}^n a_{n,h}^2 \rightarrow 4k^{\bullet,2,2}$.
Q.E.D.

PROOF OF THEOREM 4: The convergence of the individual terms follow from Theorem 3, and the stable convergence for the sum of the three terms follows by Lemma 1 and Proposition 5, using the same arguments as those in our proof of Theorem 1. Q.E.D.

PROOF OF PROPOSITION 1: To achieve the fastest rate of convergence, we need $k'(0) = 0$. With $H = c\xi\sqrt{n}$, the asymptotic variance of the resulting realized kernel estimator is that in (14), $\omega(t \int_0^t \sigma_u^4 du)^{3/4} (ck_{\bullet}^{0,0} + 2c^{-1}\rho k_{\bullet}^{1,1} + c^{-3}k_{\bullet}^{2,2})$. In the parametric version of the problem, $\rho = 1$, integration by parts gives us,

$$4(ck_{\bullet}^{0,0} + 2c^{-1}k_{\bullet}^{1,1} + c^{-3}k_{\bullet}^{2,2}) = 4c \int_0^\infty \{k(x) - c^{-2}k''(x)\}^2 dx.$$

So determining the optimal kernel function for this case amounts to solving the calculus of variation problem

$$\min_k \int_0^\infty F(x, k(x), k'(x), k''(x)) dx \quad \text{s.t.} \quad k(0) = 1 \text{ and } k'(0) = 0,$$

where $F(x, k(x), k'(x), k''(x)) = \{k(x) - c^{-2}k''(x)\}^2$. The generalized Euler equation for this problem is

$$\begin{aligned} 0 &= \frac{\partial F}{\partial k} - \frac{d}{dx} \frac{\partial F}{\partial \dot{k}} + \frac{d^2}{dx^2} \frac{\partial F}{\partial \ddot{k}} \\ &= 2\{k(x) - c^{-2}k''(x)\} - 0 - 2c^{-2} \frac{d^2}{dx^2} \{k(x) - c^{-2}k''(x)\} \\ &= 2k(x) - 2c^{-2}k''(x) - 2c^{-2}k''(x) + 2c^{-4}k^{(4)}(x) \\ &= 2k(x) - 4c^{-2}k''(x) + 2c^{-4}k^{(4)}(x), \end{aligned}$$

where $\dot{k} = k'(x)$ and $\ddot{k} = k''(x)$. The four roots of the corresponding characteristic polynomial, $1 - 2\lambda^2/c^2 + \lambda^4/c^4 = 0$, are $\lambda_{1,2} = \pm c$ and $\lambda_{3,4} = \pm ic$. So we seek a solution in the form $C_1e^{cx} + C_2xe^{cx} + C_3e^{-cx} + C_4xe^{-cx}$. We need $C_1 = C_2 = 0$ to rule out the explosive solutions, and $k(0) = 1$ implies $C_3 = 1$. Finally, from $k'(0) = 0$, we find $\partial(e^{-cx} + C_4xe^{-cx})/\partial x|_{x=0} = C_4 - c = 0$, so that $C_4 = c$. Hence the relevant solution is $k_c(x) = e^{-cx} + cxe^{-cx} = (1 + cx)e^{-cx}$. We note that $4c \int_0^\infty \{k_c(x) - c^{-2}k_c''(x)\}^2 dx = 8$ does not depend on c , so we are free to choose the scale. We choose $c = 1$, and the resulting kernel function, $k(x) = (1 + x)e^{-x}$, has $k_{\bullet}^{0,0} = \frac{5}{4}$, $k_{\bullet}^{1,1} = \frac{1}{4}$, and $k_{\bullet}^{2,2} = \frac{1}{4}$. Consistent with $c = 1$, we find that the expression (16) leads to

$$c^* = \sqrt{\frac{k_{\bullet}^{1,1}}{k_{\bullet}^{0,0}} \left\{ 1 + \sqrt{1 + \frac{3k_{\bullet}^{0,0}k_{\bullet}^{2,2}}{(k_{\bullet}^{1,1})^2}} \right\}} = \sqrt{\frac{1}{5} \{1 + \sqrt{16}\}} = 1,$$

and, finally, we find that

$$\begin{aligned}
 g &= \frac{16}{3} \sqrt{k_{\bullet}^{0,0} k_{\bullet}^{1,1}} \left\{ \frac{1}{\sqrt{1 + \sqrt{1 + 3k_{\bullet}^{0,0} k_{\bullet}^{2,2} / (k_{\bullet}^{1,1})^2}}} \right. \\
 &\quad \left. + \sqrt{1 + \sqrt{1 + \frac{3k_{\bullet}^{0,0} k_{\bullet}^{2,2}}{(k_{\bullet}^{1,1})^2}}} \right\} \\
 &= \frac{16}{3} \sqrt{\frac{5}{16}} \left\{ \frac{1}{\sqrt{1 + \sqrt{1 + 15}}} + \sqrt{1 + \sqrt{1 + 15}} \right\} = 8,
 \end{aligned}$$

so that the realized kernel based on this kernel function achieves the parametric efficiency bound. Q.E.D.

PROOF OF PROPOSITION 2: The cadlag property of v follows by direct argument. Further, by Lebesgue’s theorem, the integral (23) is the same whether interpreted as a Riemann integral or a Lebesgue integral. With the latter interpretation, we find $Y_t = \int_0^t \sigma_{T_u}^2 \tau_u^2 du = \int_0^t \sigma_{T_u}^2 dT_u = \int_0^{T_t} \sigma_u^2 du = S \circ T_t$. Q.E.D.

PROOF OF PROPOSITION 3: With

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

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 $V_{h,n} = 2 \sum_{i=1}^n U_{j\delta} \times U_{(j-h)\delta} + O_p(1)$, so the first term is (–1 times)

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

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

PROOF OF PROPOSITION 4: Let $\bar{\gamma}_h = n^{-1} \sum_{i=1}^n U_{i\delta} U_{(i-h)\delta}$. Since $k''(0) = 0$, we have

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

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

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

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

$$\sqrt{n} \begin{pmatrix} \bar{\gamma}_h - \varphi^h \omega^2 \\ \bar{\gamma}_k - \varphi^k \omega^2 \end{pmatrix} \xrightarrow{L} N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, 2\omega^4 \frac{1 + \varphi^2}{1 - \varphi^2} \begin{pmatrix} 1 & \varphi^{k-h} \\ \varphi^{k-h} & 1 \end{pmatrix} \right\},$$

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

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

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

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. [1482,1487]
 ——— (2006): "Ultra High Frequency Volatility Estimation With Dependent Microstructure Noise," Unpublished Paper, Dept. of Economics, Princeton University. [1485,1519-1521]

- ALDOUS, D. J., AND G. K. EAGLESON (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331. [1522]
- 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. [1482]
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000): "Great Realizations," *Risk*, 13, 105–108. [1481]
- (2001): "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55; Correction (2003), 98, 501. [1481,1486,1518,1522]
- ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [1483,1496]
- BACK, K. (1991): "Asset Pricing for General Processes," *Journal of Mathematical Economics*, 20, 371–395. [1484]
- BANDI, F. M., AND J. R. RUSSELL (2005): "Realized Covariation, Realized Beta and Microstructure Noise," Unpublished Paper, Graduate School of Business, University of Chicago. [1485,1491,1522]
- (2006): "Market Microstructure Noise, Integrated Variance Estimators, and the Limitations of Asymptotic Approximations: A Solution," Unpublished Paper, Graduate School of Business, University of Chicago. [1483,1497]
- (2008): "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369. [1482,1504,1517]
- BANDI, F. M., J. R. RUSSELL, AND C. YANG (2007): "Realized Volatility Forecasting and Option Pricing," *Journal of Econometrics*, forthcoming. [1522]
- 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. [1481,1486,1491,1510,1514–1517]
- (2004): "Econometric Analysis of Realised Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics," *Econometrica*, 72, 885–925. [1505]
- (2005): "Power Variation and Time Change," *Theory of Probability and Its Applications*, 50, 1–15. [1506]
- (2006): "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30. [1505]
- 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. [1485,1505,1522]
- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2007): Working Paper, available at <http://www.hha.dk/~alunde/BNHLS/BNHLS.htm>. [1510,1511,1513]
- 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. [1532]
- BILLINGSLEY, P. (1995): *Probability and Measure* (Third Ed.). New York: Wiley. [1528]
- BOLLEN, B., AND B. INDER (2002): "Estimating Daily Volatility in Financial Markets Utilizing Intraday Data," *Journal of Empirical Finance*, 9, 551–562. [1482]
- COMTE, F., AND E. RENAULT (1998): "Long Memory in Continuous-Time Stochastic Volatility Models," *Mathematical Finance*, 8, 291–323. [1486]
- DELATTE, 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. [1482]
- DOORNIK, J. A. (2006): *An Object-Oriented Matrix Language—Ox 4*. London: Timberlake Consultants Press.
- FANG, Y. (1996): "Volatility Modeling and Estimation of High-Frequency Data With Gaussian Noise," Unpublished Ph.D. Thesis, Sloan School of Management, MIT. [1481,1491]
- FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29. [1486]
- GALLANT, A. R. (1987): *Nonlinear Statistical Models*. New York: Wiley. [1483]

- GHYSELS, E., A. C. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," in *Statistical Methods in Finance*, ed. by C. R. Rao and G. S. Maddala. Amsterdam: North-Holland, 119–191. [1484]
- GLOTER, A., AND J. JACOD (2001a): "Diffusions With Measurement Errors. I—Local Asymptotic Normality," *ESAIM: Probability and Statistics*, 5, 225–242. [1487]
- (2001b): "Diffusions With Measurement Errors. II—Measurement Errors," *ESAIM: Probability and Statistics*, 5, 243–260. [1487]
- GONCALVES, S., AND N. MEDDAHI (2004): "Bootstrapping Realized Volatility," Unpublished Paper, CIRANO, Montreal. [1510,1511]
- GRIFFIN, J. E., AND R. C. A. OOMEN (2006): "Covariance Measurement in the Presence of Non-Synchronous Trading and Market Microstructure Noise," Unpublished Paper, Dept. of Statistics, University of Warwick. [1522]
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. San Diego: Academic Press. [1522]
- 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. [1511]
- (2006): "Realized Variance and Market Microstructure Noise" (With Discussion), *Journal of Business and Economic Statistics*, 24, 127–218. [1483,1485,1512,1515]
- HANSEN, P. R., J. LARGE, AND A. LUNDE (2008): "Moving Average-Based Estimators of Integrated Variance," *Econometric Reviews*, 27, 79–111. [1482]
- HAYASHI, T., AND N. YOSHIDA (2005): "On Covariance Estimation of Non-Synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379. [1522]
- HUANG, X., AND G. TAUCHEN (2005): "The Relative Contribution of Jumps to Total Price Variation," *Journal of Financial Econometrics*, 3, 456–499. [1511,1512]
- JACOD, J. (1994): "Limit of Random Measures Associated With the Increments of a Brownian Semimartingale," Preprint 120, Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris. [1491]
- (1997): "On Continuous Conditional Gaussian Martingales and Stable Convergence in Law," in *Séminaire Probability XXXI*, Lecture Notes in Mathematics, Vol. 1655, ed. by J. Azema, M. Emery, and M. Yor. Berlin: Springer Verlag, 232–246. [1522]
- (2006): "Asymptotic Properties of Realized Power Variations and Related Functionals of Semimartingales," Unpublished Paper, Laboratoire de Probabilités, Université Paris VI. [1510]
- (2007): "Statistics and High Frequency Data," Unpublished Lecture Notes, Séminaire Européen de Statistique 2007: Statistics for Stochastic Differential Equations Models. [1524, 1525]
- (2008): "A New Type of Limit Theorems for Discretized Processes," Unpublished Paper, Laboratoire de Probabilités, Université Paris VI. [1528]
- JACOD, J., AND P. PROTTER (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307. [1491]
- JACOD, J., AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes* (Second Ed.). Berlin: Springer Verlag. [1484,1505,1522]
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2007): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," Unpublished Paper, Dept. of Statistics, University of Chicago. [1504]
- KALNINA, I., AND O. LINTON (2006): "Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error," Unpublished Paper, Dept. of Economics, London School of Economics. [1485,1501,1508]
- KINNEBROCK, S., AND M. PODOLSKIJ (2008): "A Note on the Central Limit Theorem for Bipower Variation of General Functions," *Stochastic Processes and Their Applications*, 118, 1056–1070. [1525]
- LARGE, J. (2005): "Theoretical and Empirical Studies of Financial Markets," Unpublished Ph.D. Thesis, University of Oxford. [1482]

- MALLIAVIN, P., AND M. E. MANCINO (2002): "Fourier Series Method for Measurement of Multivariate Volatilities," *Finance and Stochastics*, 6, 49–61. [1499]
- MANCINO, M. E., AND S. SANFELICI (2007): "Robustness of Fourier Estimator of Integrated Volatility in the Presence of Microstructure Noise," *Computational Statistics and Data Analysis*, 52, 2966–2989. [1500]
- MEDDAHI, N. (2002): "A Theoretical Comparison Between Integrated and Realized Volatilities," *Journal of Applied Econometrics*, 17, 479–508. [1481]
- MERTON, R. C. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361. [1486]
- MYKLAND, P. A. (2006): "A Gaussian Calculus for Inference From High Frequency Data," Unpublished Paper, Dept. of Statistics, University of Chicago. [1505]
- MYKLAND, P. A., AND L. ZHANG (2006): "ANOVA for Diffusions and Ito Processes," *Annals of Statistics*, 34, 1931–1963. [1506]
- NEWBY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [1483]
- OOMEN, R. A. A. (2005): "Properties of Bias Corrected Realized Variance in Calendar Time and Business Time," *Journal of Financial Econometrics*, 3, 555–577. [1482,1504]
- PARK, S. K., AND R. A. SCHWENGERDT (1983): "Image Reconstruction by Parametric Cubic Convolution," *Computer Vision, Graphics, and Image Processing*, 3, 258–272. [1495]
- PHILLIPS, P. C. B., AND S. OULIARIS (1990): "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193. [1522]
- PHILLIPS, P. C. B., AND J. YU (2006): "Comment on 'Realized Variance and Market Microstructure Noise'," *Journal of Business and Economic Statistics*, 24, 202–208. [1485,1506]
- 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. [1495]
- POLITIS, D. N. (2005): "Higher-Order Accurate, Positive Semi-Definite Estimation of Large-Sample Covariance and Spectral Density Matrices," Unpublished Paper, Dept. of Mathematics, USCD. [1483]
- POLITIS, D. N., AND J. P. ROMANO (1995): "Bias-Corrected Nonparametric Spectral Estimation," *Journal of Time Series Analysis*, 16, 67–103. [1483]
- PROTTER, P. (2004): *Stochastic Integration and Differential Equations*. New York: Springer Verlag. [1484,1505]
- RENAULT, E., AND B. WERKER (2008): "Causality Effects in Return Volatility Measures With Random Times," Unpublished Paper, Dept. of Economics, University of North Carolina. [1483, 1506,1514]
- RÉNYI, A. (1963): "On Stable Sequences of Events," *Sankyā, Series A*, 25, 293–302. [1522]
- REVUZ, D., AND M. YOR (1999): *Continuous Martingales and Brownian Motion* (Third Ed.). Heidelberg: Springer Verlag. [1506]
- RIPLEY, B. D. (1987): *Stochastic Simulation*. New York: Wiley. [1491]
- 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 (2005) *Stochastic Volatility: Selected Readings*. London: Oxford University Press. [1486]
- SHEPHARD, N. (2005): *Stochastic Volatility: Selected Readings*. London: Oxford University Press. [1484]
- SHEPPARD, K. (2006): "Measuring Realized Covariance," Unpublished Paper, Dept. of Economics, University of Oxford. [1522]
- STEIN, M. L. (1987): "Minimum Norm Quadratic Estimation of Spatial Variograms," *Journal of the American Statistical Association*, 82, 765–772. [1487]
- 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. [1522]
- ZHANG, L. (2005): "Estimating Covariation: Epps Effect and Microstructure Noise," Unpublished Paper, Dept. of Finance, University of Illinois, Chicago. [1522]

- (2006): “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” *Bernoulli*, 12, 1019–1043. [1482]
- 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. [1482,1485,1491,1519,1521]
- ZHOU, B. (1996): “High-Frequency Data and Volatility in Foreign-Exchange Rates,” *Journal of Business and Economic Statistics*, 14, 45–52. [1481-1483,1491]

The T.N. Thiele Centre for Mathematics in Natural Science, Dept. of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark and CREATES, University of Aarhus, DK-8000, Aarhus, Denmark; oebn@imf.au.dk,

Dept. of Economics, Stanford University, Landau Economics Building, 579 Serra Mall, Stanford, CA 94305-6072, U.S.A.; peter.hansen@stanford.edu,

Dept. of Marketing and Statistics, Aarhus School of Business, University of Aarhus, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark and CREATES, University of Aarhus, DK-8000 Aarhus, Denmark; alunde@asb.dk,

and

Oxford-Man Institute, University of Oxford, Blue Boar Court, Alfred Road, Oxford OX1 4EH, U.K. and Dept. of Economics, University of Oxford, Oxford OX1 4EH, U.K.; neil.shephard@economics.ox.ac.uk.

Manuscript received May, 2006; final revision received April, 2008.