Subsampling realised kernels*

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

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

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

Neil Shephard
Nuffield College, University of Oxford, Oxford OX1 1NF, UK
neil.shephard@economics.ox.ac.uk

August 2006

Abstract

In a recent paper we have introduced the class of realised kernel estimators of the increments of quadratic variation in the presence of noise. We showed that this estimator is consistent and derived its limit distribution under various assumptions on the kernel weights. In this paper we extend our analysis, looking at the class of subsampled realised kernels and we derive the limit theory for this class of estimators. We find that subsampling is highly advantageous for estimators based on discontinuous kernels, such as the truncated kernel. For *kinked kernels*, such as the Bartlett kernel, we show that subsampling is impotent, in the sense that subsampling has no effect on the asymptotic distribution. Perhaps surprisingly, for the efficient *smooth kernels*, such as the Parzen kernel, we show that subsampling is harmful as it increases the asymptotic variance. We also study the performance of subsampled realised kernels in simulations and in empirical work.

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

*Neil Shephard's research is supported by the UK's ESRC. The Ox language of Doornik (2001) was used to perform the calculations reported here.

1 Introduction

High frequency financial data allows us to try to measure the ex-post variation of asset prices by estimating the increments to quadratic variation (e.g. Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002)). Common estimators, such as the realised variance, can be sensitive to market frictions when applied to returns recorded over shorter time intervals such as 1 minute, or even more ambitiously, 1 second (e.g. Zhou (1996), Fang (1996) and Andersen, Bollerslev, Diebold, and Labys (2000)). In response two non-parametric generalisations have been proposed in the literature: *subsampling* and *realised kernels* by Zhang, Mykland, and Aït-Sahalia (2005b) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), respectively. In this paper we partially unify these approaches by studying the properties of *subsampled realised kernels*.

Our interest will be on inference for the ex-post variation of log-prices over some arbitrary fixed time period, such as a day, using estimators of the realised kernel type. We represent this period as the single interval $[0,t]$. For a continuous time log-price process X and time gap $\delta > 0$, the flat-top *realised kernels* of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) take on the following form

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

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

$$\gamma_h(X_\delta) = \sum_{j=1}^{n_\delta} x_j x_{j-h}, \quad x_j = X_{\delta j} - X_{\delta(j-1)},$$

with $h = -H, ..., -1, 0, 1, ..., H$ and $n_\delta = \lfloor t/\delta \rfloor$. We will think of δ as being small and so x_j represents the j-th high frequency return, while $\gamma_0(X_\delta)$ is the realised variance of X. Here $\tilde{K}(X_\delta) - \gamma_0(X_\delta)$ is the realised kernel correction to realised variance for market frictions. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) gave a relatively exhaustive treatment of $\tilde{K}(X_\delta)$ when X is a Brownian semimartingale plus noise, where the noise evolves in observation time. The non-flat-top kernel replaces the kernel weight $k\left(\frac{h-1}{H}\right)$ with $k\left(\frac{h}{H}\right)$, whose properties are also studied by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).

Realised kernels are based on returns that are computed on a time mesh which is started at time $t = 0$. Starting at $t = 0$ is an ad hoc choice and there may be efficiency gains possible by jittering the initial value many times and averaging the resulting collection of different realised kernel estimators. This point is made forcefully in the context of calculating realised variances by Zhang, Mykland, and Aït-Sahalia (2005b). The jittering of the initial value is illustrated in Figure 1.

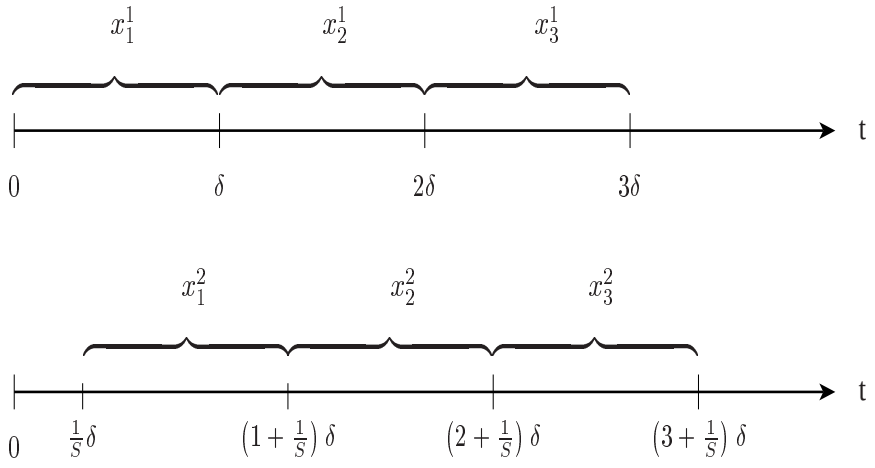

Figure 1: *Two sets of returns. The top series x_j^1 are the conventional ones. The bottom series are the offset returns x_j^s, $s = 2, \ldots, S$.* These are used to compute alternative realized autocovariances and subsampled realized kernels.

For the analysis of subsampled realised kernels it is helpful to distinguish between three types of kernels functions, $k(x)$, with $k(0) = 1$ and $k(1) = 0$. We label the three types of kernel functions as *smooth*, *kinked*, and *discontinuous* kernels. Representative members of these three classes of weight functions are the Parzen, the Bartlett, and the truncated kernel, respectively. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) have shown that the class of smooth weights, which satisfy $k'(0) = k'(1) = 0$, lead to realised kernels that converges at the efficient rate, $n^{1/4}$. Whereas the kinked kernels, which do not satisfy $k'(0) = k'(1) = 0$, lead to realised kernels that convergence at the slower rate, $n^{1/6}$. The discontinuous kernels lead to inconsistent estimators as we show in Section 6.

In this paper we show that subsampling is very useful for the class of discontinuous kernels, because subsampling makes these estimators consistent and converge in distribution at rate $n^{1/6}$. In his pioneering paper, Zhou (1996) used a simple truncated kernel and gave a brief discussion of the subsampled version of his realised kernel. His estimator belongs to the class of discontinuous kernels. We will see that his estimator can be made consistent by allowing $S \to \infty$ as $n \to \infty$, a result which is implicit in his paper, but one he did not explicitly draw out. For the class of kinked kernels, we show that subsampling is impotent, in the sense that the asymptotic distribution is the same whether subsampling is used or not. Finally, we show that subsampling is harmful when applied to smooth kernels. In fact, if the number of subsamples, S, increases with the sample size, n, the best rate of convergence is reduced to less than the efficient one, $n^{1/4}$.

Still, subsampling does provide a simple way to make use of all available data while making valid inference using realistic assumptions about the noise in tick-by-tick data. We discuss this

aspect in Section 7 and make recommendations on how to implement subsampled realised kernels in empirical work.

Our analysis is based on equally spaced data. By applying the time-change argument of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), it follows that our results also applies to irregularly spaced data. For instance, the case where δ corresponds to the time between every fifth transaction.

This paper has the following structure. We present the basic framework in Section 2 along with some known results. In Section 3 we derive the limit theory for subsampled realised autocovariances. We present our main results in Section 4. Here we derive the limit theory for subsampled realised kernels and show that subsampling cannot improve realised kernels within a very broad class of estimators. Section 5 presents some intuition for our theoretical results. In Section 6 we characterize some poorly designed kernels and show that subsampling improves upon such estimators. In Section 7, we given some specific recommendations on empirical implementation of subsampled realised kernels and how to conduct valid inference in this context. We present results from a small simulation study in Section 8 and an empirical application in Section 9. We conclude in Section 10 and present all proofs in an appendix.

2 Notation, definitions and background

2.1 Semimartingales and quadratic variation

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

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

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

(e.g. Protter (2004, p. 66–77) and Jacod and Shiryaev (2003, p. 51)) for any sequence of deterministic partitions $0 = t_0 < t_1 < ... < t_N = t$ with $\sup_j \{t_{j+1} - t_j\} \to 0$ for $N \to \infty$.

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

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

where a is a predictable locally bounded drift, σ is a càdlàg volatility process and W is a Brownian motion. For reviews of the econometrics of this type of process see, for example, Ghysels, Harvey, and Renault (1996) and Shephard (2005). If $Y \in \mathcal{BSM}$ then

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

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

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

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

2.2 Assumptions about noise

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

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

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

We make a white noise assumption about the U process ($U \in \mathcal{WN}$) which we assume has

$$\mathrm{E}(U_t) = 0, \quad \mathrm{Var}(U_t) = \omega^2, \quad \mathrm{Var}(U_t^2) = \lambda^2 \omega^4, \quad U_t \perp\!\!\!\perp U_s \tag{5}$$

for any $t \neq s$, where $\lambda \in \mathbb{R}^+$. This white noise assumption is unsatisfactory from a number of viewpoints (e.g. Phillips and Yu (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 (2005b) 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.

2.3 Some known results

Analogous to the realised autocovariances we define

$$\gamma_h(Y_\delta, U_\delta) = \sum_{j=1}^{n_\delta} y_j u_{j-h}, \quad y_j = Y_{\delta j} - Y_{\delta(j-1)} \quad \text{and} \quad u_j = U_{\delta j} - U_{\delta(j-1)}.$$

From (4) we have that

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

It will be useful to have the following notation $\widetilde{\gamma}(X_\delta) = \{\gamma_0(X_\delta), \widetilde{\gamma}_1(X_\delta), ..., \widetilde{\gamma}_H(X_\delta)\}^\intercal$, where $\widetilde{\gamma}_h(X_\delta) = \gamma_h(X_\delta) + \gamma_{-h}(X_\delta)$, and introduce the analogous definitions of $\widetilde{\gamma}(Y_\delta)$, $\widetilde{\gamma}(U_\delta)$, and $\widetilde{\gamma}(Y_\delta, U_\delta)$.

In the non-subsampling case Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) derived the following helpful results.

Theorem 1 *We study properties as $\delta \downarrow 0$, implying $n_\delta \to \infty$. Writing MN to denote a mixed normal distribution. Suppose that $Y \in \mathcal{BSM}$ and (3) holds, then*

$$n_\delta^{1/2} \begin{pmatrix} \gamma_0(Y_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u \\ \tilde{\gamma}_1(Y_\delta) \\ \vdots \\ \tilde{\gamma}_H(Y_\delta) \end{pmatrix} \xrightarrow{Ls} MN\left(0, A_1 \times t \int_0^t \sigma_u^4 \mathrm{d}u\right), \quad A_1 = \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 4 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 4 \end{pmatrix}. \tag{6}$$

Here Ls denotes convergence in law stably. If, in addition, $U \in \mathcal{WN}$ and $Y \perp\!\!\!\perp U$ then $\widetilde{\gamma}(Y_\delta, U_\delta) \xrightarrow{Ls} MN\left(0, 2\omega^2[Y]B\right)$, where B is a $(H+1) \times (H+1)$ symmetric matrix with block structure

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

$B_{12} = B_{21}^\intercal$. *Here B_{22} is a $(H-1) \times (H-1)$ symmetric matrix.*

Finally, when $U \in \mathcal{WN}$ and writing $n_\delta = \lfloor t/\delta \rfloor$, for $n_\delta \geq H$

$$\mathrm{E}\left\{\widetilde{\gamma}(U_\delta)\right\} = 2\omega^2 n_\delta \left(1, -1, 0, 0, ..., 0\right)^\intercal, \quad \textit{and} \quad \mathrm{Cov}\left\{\widetilde{\gamma}(U_\delta)\right\} = 4\omega^4 \left(n_\delta C + \widetilde{D}\right).$$

Here the $(H+1) \times (H+1)$ symmetric matrices C and $\widetilde{D}$ have block structure

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad \widetilde{D} = \begin{pmatrix} \widetilde{D}_{11} & \widetilde{D}_{12} \\ \widetilde{D}_{21} & \widetilde{D}_{22} \end{pmatrix},$$

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

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

$$\widetilde{D}_{22} = \begin{pmatrix} -7 & \bullet & \bullet & \bullet & \bullet & \bullet \\ 6 & -10 & \bullet & \bullet & \bullet & \bullet \\ -2 & 8 & -13 & \bullet & \bullet & \bullet \\ 0 & -2.5 & 10 & -16 & \bullet & \bullet \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{H}{2} & 2H & -3H-1 \end{pmatrix}, \ \widetilde{D}_{21} = \begin{pmatrix} -1 & 4 \\ 0 & -\frac{3}{2} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix},$$

where $C_{12} = C_{21}^{\intercal}$ *and* $\widetilde{D}_{12} = \widetilde{D}_{21}^{\intercal}$. *The* 2×2 *matrices* C_{11} *and* $\widetilde{D}_{11}$ *are*

$$C_{11} = \begin{pmatrix} 1+\lambda^2 & -2-\lambda^2 \\ -2-\lambda^2 & 5+\lambda^2 \end{pmatrix}, \widetilde{D}_{11} = \begin{pmatrix} -\lambda^2/2 & \lambda^2/2+1 \\ \lambda^2/2+1 & -\lambda^2/2-7/2 \end{pmatrix}.$$

3 Subsampled realised autocovariances

The subsampled realised autocovariances are defined by

$$\gamma_h^s(X_\delta) = \sum_{j=1}^{n_\delta} x_j^s x_{j-h}^s, \quad x_j^s = X_{\delta(j+(s-1)/S)} - X_{\delta(j+(s-1)/S-1)},$$

for $s = 1, \ldots, S$, where x_j^s are intraday returns over intervals of length δ, see Figure 1 for an illustration. For each of the S subsamples the realised kernel is given by,

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

and we define the *subsampled realised kernel* as

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

Here

$$\gamma_h(X_\delta; S) = \frac{1}{S}\sum_{s=1}^{S} \gamma_h^s(X_\delta).$$

Notice that the subsampled realised kernel computes returns over intervals of length δ but uses prices measured every δ/S periods. Hence this statistic works the database of high frequency returns more intensively than each of the realised kernels, $\tilde{K}^s(X_\delta)$. While the sample size used to construct each of the realised kernels is n_δ, the *effective sample size* used by the subsampled realised kernel, $\tilde{K}(X_\delta; S)$, is $n = S \times n_\delta$.

3.1 General theory

The extension of the terms involving noise to the subsampling case is straightforward under $U \in \mathcal{WN}$, as the market microstructure terms, U_t, are uncorrelated (actually independent) cross subsamples. This implies

$$\widetilde{\gamma}(Y_\delta, U_\delta; S) \xrightarrow{Ls} MN\left(0, \frac{2\omega^2}{S}[Y]B\right), \tag{7}$$

$$\mathrm{E}\left\{\widetilde{\gamma}(U_\delta;S)\right\} = 2\omega^2 n_\delta \left(1,-1,0,0,...,0\right)^\intercal, \tag{8}$$

$$\mathrm{Cov}\left\{\widetilde{\gamma}(U_\delta;S)\right\} = \frac{4\omega^4}{S}\left(n_\delta C + \widetilde{D}\right). \tag{9}$$

The contributions from the $\tilde{D}$ matrix are known as end-effects because they are tied to the S first and the S last observations. For most estimators, this term does not show up in the asymptotic variance because it is of lower order than the other terms, such as those associated with the C matrix. The only exception is when the estimator is based on a smooth kernel and a fixed S. However, the end effects are also negligible in this case, because it follows from Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) that their contribution to the asymptotic variance is of order $O\left(\omega\right)$, which is known to be small in practice. For this reason, we will ignore these end effects, as it simplifies the exposition of our analyses.

We need to extend (6) to the subsampling case. This is given in the following Theorem.

Theorem 2 *Suppose that $Y \in \mathcal{BSM}$ and (3) holds, then as $\delta \downarrow 0$*

$$n_\delta^{1/2}\begin{pmatrix} \gamma_0(Y_\delta;S) - \int_0^t \sigma_u^2 \mathrm{d}u, \\ \widetilde{\gamma}_1(Y_\delta;S) \\ \vdots \\ \widetilde{\gamma}_H(Y_\delta;S) \end{pmatrix} \xrightarrow{Ls} MN\left(0, A_S \times t\int_0^t \sigma_u^4 \mathrm{d}u\right),$$

where

$$A_S = \frac{2}{3}\begin{pmatrix} 2+S^{-2} & \bullet & 0 & \cdots \\ 1-S^{-2} & 4+2S^{-2} & \bullet & \ddots \\ 0 & 1-S^{-2} & 4+2S^{-2} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \to \frac{2}{3}\begin{pmatrix} 2 & 1 & 0 & \cdots \\ 1 & 4 & 1 & \ddots \\ 0 & 1 & 4 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} = A_\infty,$$

and as $\delta \downarrow 0$ and $S \to \infty$

$$n_\delta^{1/2}\begin{pmatrix} \gamma_0(Y_\delta;S) - \int_0^t \sigma_u^2 \mathrm{d}u, \\ \widetilde{\gamma}_1(Y_\delta;S) \\ \vdots \\ \widetilde{\gamma}_H(Y_\delta;S) \end{pmatrix} \xrightarrow{Ls} MN\left(0, A_\infty \times t\int_0^t \sigma_u^4 \mathrm{d}u\right).$$

3.2 Comments

Key to the asymptotic distribution of the $\widetilde{\gamma}_h(Y_\delta)$ is the A_S matrix. Important special cases of this result are A_1 (defined in Theorem 1) and

$$A_2 = \begin{pmatrix} 3/2 & \bullet & \bullet & \cdots \\ 1/2 & 3 & \bullet & \ddots \\ 0 & 1/2 & 3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \quad A_3 = \frac{2}{3}\begin{pmatrix} 2(1+\frac{1}{9}) & \bullet & \bullet & \cdots \\ (1-\frac{1}{9}) & 4(1+\frac{1}{18}) & \bullet & \ddots \\ 0 & (1-\frac{1}{9}) & 4(1+\frac{1}{18}) & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The limiting result is a good approximation even for very small S. Subsampling does improve the accuracy of the realised autocovariances, however the improvements are very modest indeed and the potential gains are almost exhausted for very small values of S.

These matrices include a number of important special cases which have influenced the recent econometric analysis of realised volatility. The asymptotic distribution

$$n_\delta^{1/2}\left(\gamma_0(Y_\delta)-[Y]_t\right)\overset{Ls}{\to} MN\left(0,2t\int_0^t\sigma_u^4\mathrm{d}u\right) \tag{10}$$

appears in the work of Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002). The extension of (10) to the subsampled case

$$n_\delta^{1/2}\left(\gamma_0(Y_\delta;S)-[Y]_t\right)\overset{Ls}{\to} MN\left(0,\frac{4}{3}\left(1+2S^{-2}\right)t\int_0^t\sigma_u^4\mathrm{d}u\right), \tag{11}$$

is in Zhang, Mykland, and Aït-Sahalia (2005b). Note that $\frac{2}{3}\left(2+S^{-2}\right)$ falls from 2 to 4/3 as S rises from 1 to infinity, so

$$n_\delta^{1/2}\left(\gamma_0(Y_\delta;S)-[Y]_t\right)\overset{Ls}{\to} MN\left(0,\frac{4}{3}t\int_0^t\sigma_u^4\mathrm{d}u\right),\quad \text{as } S,n_\delta\to\infty. \tag{12}$$

So in the absence of noise, the subsampled realised variance, $\gamma_0(Y_\delta;S)$, produces a slightly more precise estimator than the realised variance, $\gamma_0(Y_\delta)$, by exploiting more of the data. Goncalves and Meddahi (2004) and Zhang, Mykland, and Aït-Sahalia (2005a) have studied Edgeworth expansions of these types of results, while the former also derived a bootstrapped version to improve the finite sample performance of the feasible version of the theory.

4 Subsampled realised kernel

In this section, we study subsampled realised kernels based on smooth and kinked kernel functions. Specifically, we require that $k(s)$ is continuous and twice differentiable on $[0,1]$ and that $k(0)=1$ and $k(1)=0$. Naturally, the derivatives at the end points are defined by

$$k'(0)=\lim_{x\downarrow 0}\frac{k(x)-k(0)}{x}\qquad\text{and}\qquad k'(1)=\lim_{x\uparrow 1}\frac{k(1)-k(x)}{1-x}.$$

In the framework without subsampling, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) showed that

$$k'(0)=0\quad\text{and}\quad k'(1)=0, \tag{13}$$

is a necessary condition for a realised kernel to have the best rate of convergence, and this property is also key for subsampled realised kernels. So we shall refer to kernels that satisfy (13) as *smooth*, and we use *kinked* to refer to the kernels that violate (13).

In some of our proofs it is convenient to extend the support of the kernel functions beyond the unit interval, using the conventions: $k(x) = 0$ for $x > 1$ and $k(-x) = k(x)$.

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) showed that kernel functions of the type just described, can be used to produce consistent estimators with mixed Gaussian asymptotic distributions. These results do not require any subsampling. It is therefore interesting to analyze whether there are any gain from subsampling realised kernels or not. Perhaps surprisingly we find that subsampling is harmful or, at best, impotent, for realised kernels that are based on smooth or kinked kernel functions.

Below we formulate limit results for subsampled realised kernels using the notation

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

and it is convenient to introduce the notation

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

to simplify the expressions for the asymptotic variance.

Theorem 3 *For large H and n the asymptotic distributions of*

$$\tilde{K}(Y_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u, \quad \tilde{K}(Y_\delta, U_\delta; S) + \tilde{K}(U_\delta, Y_\delta; S), \quad \textit{and} \quad \tilde{K}(U_\delta; S),$$

are mixed Gaussian with mean zero and asymptotic variances given by

$$4\frac{H}{n_\delta} k_{\bullet}^{0,0} t \int_0^t \sigma_u^4 \mathrm{d}u, \tag{14}$$

$$8\omega^2 \int_0^t \sigma_u^2 \mathrm{d}u k_{\bullet}^{1,1} H^{-1} \Big/ S \tag{15}$$

$$4\omega^4 n_\delta \left[\left\{ k'(0)^2 + k'(1)^2 \right\} H^{-2} + k_{\bullet}^{2,2} H^{-3} \right] / S. \tag{16}$$

respectively. Furthermore, $\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u$ is mixed Gaussian with a zero mean and variance

$$4k_{\bullet}^{0,0} t \int_0^t \sigma_u^4 \mathrm{d}u \left\{ \frac{H}{n_\delta} + \frac{2\xi\rho k_{\bullet}^{1,1} H^{-1} + \xi^2 n_\delta \left[\left\{ k'(0)^2 + k'(1)^2 \right\} H^{-2} + k_{\bullet}^{2,2} H^{-3} \right]}{S} \right\}. \tag{17}$$

A very interesting observation is that subsampling has no impact on the first term, (14). The implication is that the asymptotic distribution of the realised kernel, $\tilde{K}(Y_\delta)$, is identical to that of the subsampled realised kernel $\tilde{K}(Y_\delta; S)$. So despite the fact that subsampling lowers the variance of the individual realised autocovariances, $\tilde{\gamma}_h(Y_\delta)$, the variance of the realised kernel is unaffected. The reason is that subsampling introduces positive correlation between $\tilde{\gamma}_h(Y_\delta; S)$ and $\tilde{\gamma}_{h+1}(Y_\delta; S)$

that exactly offsets the reduction in the variance of the realised autocovariances. This follows from the fact that

$$[A_S]_{i,i} + [A_S]_{i,i-1} + [A_S]_{i-1,i} = \frac{2}{3}\left\{4 + 2S^{-2} + 2(1 - S^{-2})\right\} = 4, \qquad i > 1,$$

does not depend on S.

Subsampling does have an effect on the terms that are due to noise, (15) and (16), where the contribution to the asymptotic variance is reduced by a factor of S. So it is (15) and (16) that will characterize the gains from increasing the sample size by a factor of S.

The most obvious generalisation of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) is to think of the case where S is fixed and we allow H to increase with n_δ. When (13) holds, we can follow Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) and set $H = c(\xi n_\delta)^{1/2}$. Then we obtain the result that

$$n_\delta^{1/4}\left\{\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \overset{Ls}{\to} MN\left\{0, 4\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4}\left(ck_\bullet^{0,0} + \frac{2c^{-1}\rho k_\bullet^{1,1} + c^{-3}k_\bullet^{2,2}}{S}\right)\right\},$$

which Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) saw was the best rate possible for this problem. Whether or not (13) holds, when we set $H = c(\xi n_\delta)^{2/3}$ we have

$$n_\delta^{1/6}\left\{\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \overset{Ls}{\to} MN\left[0, 4\omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}\left\{ck_\bullet^{0,0} + \frac{k'(0)^2 + k'(1)^2}{c^2 S}\right\}\right].$$

Here S plays a relatively simple role, reducing the impact of noise — by in effect reducing the noise variance from ω^2 to $\omega^2/\sqrt{S}$. If (13) does hold then we get the very simple result that

$$n_\delta^{1/6}\left\{\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \overset{Ls}{\to} MN\left\{0, 4ck_\bullet^{0,0}\omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}\right\}.$$

The latter result is interesting, for it has no asymptotic gains at all from subsampling.

Until now, we have stated asymptotic results using n_δ, even though the subsampled statistics are based on a larger sample size – one that is about S times larger. Next we make the transition to the effective sample size.

4.1 Effective Sample Size

For the purpose of discussing the effects of subsampling it is useful to make the comparison in terms of the effective sample size, $n = n_\delta S$. This makes it explicit that a larger S reduces the sample size, n_δ, that is available for each to the realised kernels. Then we ask if it is better to increase n_δ or S for a given n — i.e. should we split time into lengthy returns and lots of subsampling, or use shorter returns and less subsampling.

In terms of the effective sample size, (17) becomes

$$4t\int_0^t \sigma_u^4 \mathrm{d}u \left[\frac{HS}{n} k_\bullet^{0,0} + \frac{2\xi\rho k_\bullet^{1,1}}{HS} + n\xi^2 \left\{ \frac{k'(0)^2 + k'(1)^2}{(HS)^2} + S\frac{k_\bullet^{2,2}}{(HS)^3} \right\} \right]. \tag{18}$$

Here HS appears in the variance expression in a way that is almost identical to H when there is no subsampling ($S = 1$). The only difference is the impact on the last term. This term vanishes when $k'(0) = k'(1) = 0$ does not hold, because the second last term is then $O\left(n/(SH)^2\right)$ whereas the last term is only $O\left(H^{-1}\right) O\left(n/(SH)^2\right)$. This feature of the asymptotic variance holds the key to the different results we derive for smooth and kinked kernels.

4.2 Kinked Kernels: When $k'(0) = k'(1) = 0$ does not hold

When (13) does not hold the asymptotic variance is given by

$$4t\int_0^t \sigma_u^4 \mathrm{d}u \left\{ \frac{HS}{n} k_\bullet^{0,0} + \frac{2\xi\rho k_\bullet^{1,1}}{HS} + n\xi^2 \frac{k'(0)^2 + k'(1)^2}{(HS)^2} \right\}.$$

While this expression depends on the product HS, it is invariant to the particular values of H and S, though we do need $H \to \infty$ to justify the terms, $k_\bullet^{0,0}$, $k_\bullet^{1,1}$, etc. We have the following result.

Theorem 4 *(i) If $SH = c(\xi n)^{2/3}$ we have*

$$n^{1/6}\left(\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u \right) \xrightarrow{Ls} MN\left(0, 4\omega^{4/3} \left(t\int_0^t \sigma_u^4 \mathrm{d}u \right)^{2/3} \left\{ ck_\bullet^{0,0} + \frac{k'(0)^2 + k'(1)^2}{c^2} \right\} \right), \tag{19}$$

as $n \to \infty$, so long as H increase with n. (ii) The asymptotic variance is minimised by

$$c = \left\{ 2\frac{k'(0)^2 + k'(1)^2}{k_\bullet^{0,0}} \right\}^{1/3}, \quad \textit{and} \quad 6ck_\bullet^{0,0}\omega^{4/3} \left(t\int_0^t \sigma_u^4 \mathrm{d}u \right)^{2/3}$$

is the lower bound for the asymptotic variance.

An interesting observation is that the asymptotic distribution (19) is not influenced by S, not even the rate of growth in S. All that matters is that H grows and that HS grows at the right rate. The implication is that there are no gains from subsampling when $k'(0)^2 + k'(1)^2 \neq 0$. So we might as well set $S = 1$ and use the realised kernel that does not require any subsampling.

The second part of Theorem 4 shows that

$$ck_\bullet^{0,0} = 6\left[2\left(k_\bullet^{0,0} \right)^2 \left\{ k'(0)^2 + k'(1)^2 \right\} \right]^{1/3}$$

controls the asymptotic efficiency of estimators in this class.

Example 1 *The Bartlett kernel, $k(x) = 1 - x$, has $k_\bullet^{0,0} = 1/3$ and $k'(0)^2 + k'(1)^2 = 2$, so that $6ck_\bullet^{0,0} = 2 \cdot 12^{1/3} \simeq 4.58$, whereas the quadratic kernel, $k(x) = 1 - 2x + x^2$, is more efficient, because it has $k_\bullet^{0,0} = 1/5$ and $k'(0)^2 + k'(1)^2 = 4$, so that $6ck_\bullet^{0,0} = 12 \cdot 5^{-2/3} \simeq 4.10$.*

4.3 Smooth Kernels: When $k'(0) = k'(1) = 0$ holds

In this Section we consider smooth kernel functions. Some examples of smooth kernel functions are given in Table 1, where $k_{\text{TH}_1}(x) = \sin^2\left\{\frac{\pi}{2}(1-x)\right\} = \left[1-\cos\left\{\pi(1-x)\right\}\right]/2 = \left\{1+\cos\left(\pi x\right)\right\}/2$ is the Tukey-Hanning kernel.

Table 1: Some smooth kernel functions.

Cubic kernel	$k_C(x) = 1 - 3x^2 + 2x^3$
Parzen kernel	$k_P(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1 \end{cases}$
TH_p	$k_{TH_p}(x) = \sin^2\{\pi/2\,(1-x)^p\}$

We know from Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) that the rate of convergence of realised kernels improves when $k'(0) = k'(1) = 0$. This smoothness condition will also improve the rate of convergence for subsampled realised kernels. For smooth kernel functions, the asymptotic variance is given by

$$4t\int_0^t \sigma_u^4 \mathrm{d}u \left\{ \frac{HS}{n} k_\bullet^{0,0} + \frac{2\xi\rho k_\bullet^{1,1}}{HS} + \xi^2 nS \frac{k_\bullet^{2,2}}{(HS)^3} \right\}. \tag{20}$$

Because the last term is multiplied with S it is evident that the asymptotic distribution will depend on whether S is constant or increases with n. This is made precise in the following Theorem.

Theorem 5 *(i.a) When S is fixed we set $HS = c(\xi n)^{1/2}$ and have*

$$n^{1/4}\left\{\tilde{K}(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \xrightarrow{Ls} MN\left[0, 4\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4}\left\{ck_\bullet^{0,0} + \frac{2\rho}{c}k_\bullet^{1,1} + \frac{S}{c^3}k_\bullet^{2,2}\right\}\right]. \tag{21}$$

(i.b) When $S = an^\alpha$ for some $0 < \alpha < 2/3$, we set $HS = c(\xi n)^{1/2} n^{\alpha/4}$ and have

$$n^{\frac{1-\alpha/2}{4}}\left(\tilde{K}(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u\right) \xrightarrow{Ls} MN\left[0, 4\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4}\left\{ck_\bullet^{0,0} + \frac{a}{c^3}k_\bullet^{2,2}\right\}\right].$$

(ii) Whether S is constant or not, the asymptotic variance is minimized by

$$HS = (\xi n)^{1/2}\sqrt{\frac{\rho k_\bullet^{1,1}}{k_\bullet^{0,0}}\left\{1 + \sqrt{1 + 3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}\right\}},$$

and the lower bound is

$$n^{-1/2}\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4} g(S), \tag{22}$$

where

$$g(S) = \frac{16}{3}\sqrt{\rho k_{\bullet}^{1,1} k_{\bullet}^{0,0}} \left\{ \frac{1}{\sqrt{1+\sqrt{1+3S\frac{k_{\bullet}^{0,0}k_{\bullet}^{2,2}}{(\rho k_{\bullet}^{1,1})^2}}}} + \sqrt{1+\sqrt{1+3S\frac{k_{\bullet}^{0,0}k_{\bullet}^{2,2}}{(\rho k_{\bullet}^{1,1})^2}}} \right\}. \tag{23}$$

Remark. In $(i.b)$ we impose $\alpha < 2/3$. The reason is that $H \propto n^{1/2+\alpha/4-\alpha} = n^{(1-\frac{3}{2}\alpha)/2}$ and we need $(1-\frac{3}{2}\alpha)/2 > 0$ to ensure that H grows with n.

The relative efficiency in this class of estimators is given from $g(S)$, and we have the following important result for subsampling of smooth kernels

Corollary 1 *The asymptotic variance of $\tilde{K}(X_\delta; S)$ is strictly increasing in S.*

The implication is that subsampling is always harmful for smooth kernels. Furthermore, $(i.b)$ shows that there is an efficiency loss from allowing S to grow with n. See Table 2 for the values of $g(S)$ for some selected kernel functions.

Another implication of Theorem 5 concerns the best way to sample high frequency returns. This result is formulated in the next corollary and will require some explanation.

Corollary 2 *The asymptotic variance, (22), as a function of ρ, is minimized for $\rho = 1$.*

The parameter $\rho = \int_0^t \sigma_u^2 \mathrm{d}u / \sqrt{t\int_0^t \sigma^4 \mathrm{d}u}$ may appear to be fixed, which would make the Corollary rather uninteresting. However, ρ is not fixed because the integrated quarticity, $\int_0^t \sigma^4 \mathrm{d}u$, depends on the sampling scheme. Rather than equidistant sampling in calendar time we can generate the sampling times by,

$$t_j = t \times \tau\left(\frac{j}{n}\right), \qquad j = 0, 1, \ldots, n.$$

Here τ is simply a time changing mapping (for the unit interval), i.e. $\tau(0) = 0$, $\tau(1) = 1$, and τ is monotonically increasing, so that $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$. A change of time does not affect $\int_0^t \sigma_u^2 \mathrm{d}u$ but does influence the integrated quarticity $\int_0^t \sigma_u^4 \mathrm{d}u$, see e.g. Mykland and Zhang (2006). A particularly interesting sampling scheme is that known as *business time sampling*, see e.g. Oomen (2005, 2006). Under this sampling scheme intraday returns are sampled in a way that makes them homogeneous, i.e. $\int_{t_{i-1}}^{t_i} \sigma_u^2 \mathrm{d}u = n^{-1}\int_0^t \sigma_u^2 \mathrm{d}u$. The integrated quarticity is minimized under this sampling scheme as was shown by Hansen and Lunde (2006, p. 135), where the minimum has $t\int_0^t \sigma^4 \mathrm{d}u = \left(\int_0^t \sigma_u^2 \mathrm{d}u\right)^2$, implying $\rho = 1$. It follows that the τ for business time sampling, τ_{BTS} say, must solve $\int_0^{t\times\tau(s)} \sigma_u^2 \mathrm{d}u = s \times \int_0^t \sigma_u^2 \mathrm{d}u$. So by the implicit function theorem we have $\tau'_{\text{BTS}}(s) \propto 1/\sigma^2(\tilde{s})$, where $\tilde{s} = t \times \tau_{\text{BTS}}(s)$. Thus, under this scheme the returns are sampled more frequently when the volatility is high and less frequent when the volatility is low. In general we have $\rho \leq 1$ and Corollary 2 shows that business time sampling ($\rho = 1$) is the ideal sample scheme

(for a given sample size, n). Sampling in business time is infeasible because τ_{BTS} depends on the unknown volatility path. In practice, tick time sampling, where sampling occurs every fixed number of transactions, seems to be a better proxy for business time sampling than is calendar time sampling. In this situation, Corollary 2 states that it is better to sample returns in tick time.

Given S and ρ it is easy to compute the optimal H, as $H = c_S(\xi n)^{1/2}$ for this class of kernels, where

$$c_S = S^{-1}\sqrt{\frac{\rho k_\bullet^{1,1}}{k_\bullet^{0,0}}\left\{1+\sqrt{1+3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}\right\}}. \tag{24}$$

Table 2: *Key quantities for some smooth-continuous kernels.*

	$k_\bullet^{0,0}$	$k_\bullet^{1,1}$	$k_\bullet^{2,2}$	$\sqrt{k_\bullet^{0,0}k_\bullet^{1,1}}$	$\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(k_\bullet^{1,1})^2}$	c_1	$g(S)$			
							$S=1$	$S=2$	$S=3$	$S=10$
Cubic	0.371	1.20	12.0	0.67	3.09	3.68	9.03	9.81	10.39	12.72
Parzen	0.269	1.50	24.0	0.64	2.87	4.77	8.53	9.25	9.78	11.94
TH_1	0.375	1.23	12.2	0.68	3.00	3.70	9.18	9.96	10.55	12.89
TH_2	0.218	1.71	41.8	0.61	3.11	5.75	8.27	8.99	9.51	11.65
TH_5	0.097	3.50	489.0	0.58	3.85	8.07	8.07	8.82	10.19	11.57
TH_{10}	0.050	6.57	3610.6	0.57	4.19	24.79	8.04	8.80	10.19	11.59
TH_{16}	0.032	10.26	14374.0	0.57	4.33	39.16	8.02	8.80	10.20	11.60

Key quantities for some smooth kernels. Key is $g(S)$ that measures the relative efficiency in this class of estimators. Here computed for the case with constant volatility ($\rho = 1$) such that these numbers are comparable with the maximum likelihood estimator that has $g = 8.00$. No subsampling ($S = 1$) produces the best estimator and kernels with a relative large $k_\bullet^{0,0}k_\bullet^{2,2}/(k_\bullet^{1,1})^2$ tend to be more sensitive to subsampling.

In Table 2 we present key quantities for some smooth kernels. Perhaps the most interesting quantitiy is $g(S)$ of (23), as it enable us to compare the relative efficiency across estimators. In Table 2 we have computed $g(S)$ for the case where $\rho = 1$. So $g(S)$ can be compared to 8.00 which is the corresponding constant for the maximum likelihood estimator in the parametric version of the problem. We see that most kernels are only slightly less efficient than the maximum likelihood estimator, TH_{16} almost reaching this lower bound. Comparing $g(S)$ for different degrees of subsampling, reminds us that $S = 1$ (no subsampling) yields the most efficient estimator. The larger the value of $k_\bullet^{0,0}k_\bullet^{2,2}/(k_\bullet^{1,1})^2$ the more sensitive is the kernel to subsampling.

In Figure 2 we have plotted some smooth kernel functions, $k(x/c_1)$ using their respective optimal value for c_1, see Table 2. We see that the TH_1 kernel is almost identical to the cubic kernel. The TH_{16} kernel is somewhat flatter, putting less weight on realised autocovariance of lower order and higher weight on realised autocovariance of higher order. The Parzen kernel is typically between TH_1 and TH_{16}.

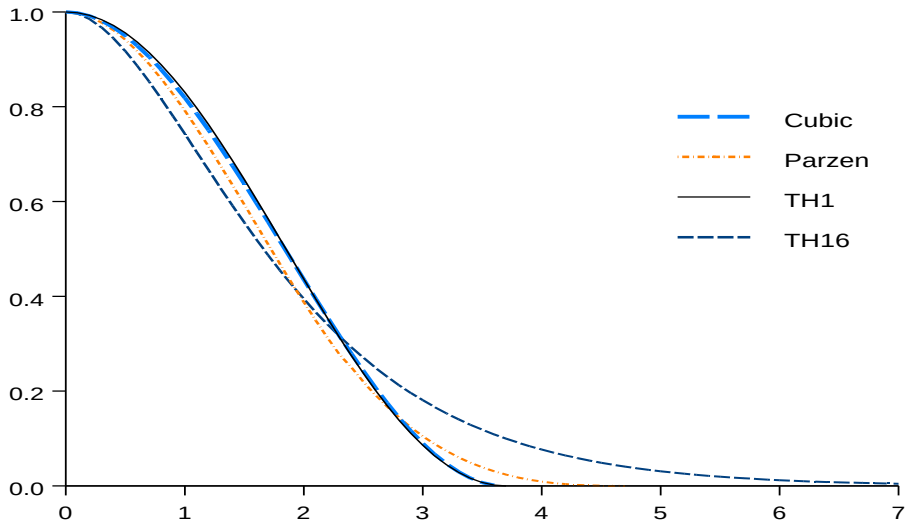

Figure 2: Plots of some selected smooth kernels, $k(x/c_1)$, using their repective optimal value of c when $S = 1$.

While the smooth kernels improve the rate of convergence over the kinked kernels, the improvements may be modest in finite samples. The reason is the following. When the noise is small the optimal H is small, and H may actually be quite similar for kinked and smooth kernels. For instance with $\xi = 0.01$ and $n = 780$, which corresponds to sampling twice per minute on a typical trading day, the Bartlett kernel has $c_{\text{BARTLETT}}(\xi n)^{2/3} = 9.00$ whereas the cubic kernel has $c_{\text{CUBIC}}(\xi n)^{1/2} = 10.78$. So in this case the two types of estimators are rather similar and despite the fact that H_{BARTLETT} grows at the faster rate $n^{2/3}$, the cubic kernels includes more lags in this situation. Consistent with this observation, Bandi and Russell (2006) find that the finite sample properties of kinked and smooth kernels are quite similar, although they do report some gains from the smooth kernels.

5 Intuition: Subsampled realised kernels are realised kernels

A closer inspection of subsampled realised kernels reveals that these can approximately be represented by a realised kernel. Lemma A.1 in the appendix shows that

$$\gamma_h(X_\delta; S) \simeq \sum_{s=-S+1}^{S-1} k_B\left(\tfrac{s}{S}\right) \gamma_{Sh+s}(X_{\delta/S}), \quad \text{where} \quad k_B(x) = 1 - |x|\,,$$

where the approximation is due to minor end-effects. See the proof of Lemma A.1 for details. The implication is that

$$\tilde{K}(X_\delta; S) \simeq \sum_{s=-S+1}^{S-1} k_B\left(\tfrac{s}{S}\right) \gamma_s(X_{\delta/S}) + \sum_{h=1}^{H} k\left(\tfrac{h-1}{H}\right) \sum_{s=-S}^{S} k_B\left(\tfrac{s}{S}\right) \left\{\gamma_{Sh+s}(X_{\delta/S}) + \gamma_{-Sh-s}(X_{\delta/S})\right\}$$

$$= \sum_{h=0}^{HS} k_S\left(\tfrac{h-1}{HS}\right)\tilde{\gamma}_{Sh+s}(X_{\delta/S}).$$

So a subsampled realised kernel is a realised kernel simply operating on a higher frequency (setting aside minor end-effects). The implied kernel weights, $k_S(\frac{h}{HS})$, $h = 1, \ldots, SH$, are simply convex combinations of neighboring weights of the original kernel,

$$k_S\left(\tfrac{hs}{HS}\right) = \tfrac{S-s}{S}k\left(\tfrac{h}{S}\right) + \tfrac{s}{S}k\left(\tfrac{h+1}{S}\right), \qquad h = 0, \ldots, H, \quad s = 1, \ldots, S. \tag{25}$$

This provides intuition for the theoretical results we have established for subsampled realised kernels.

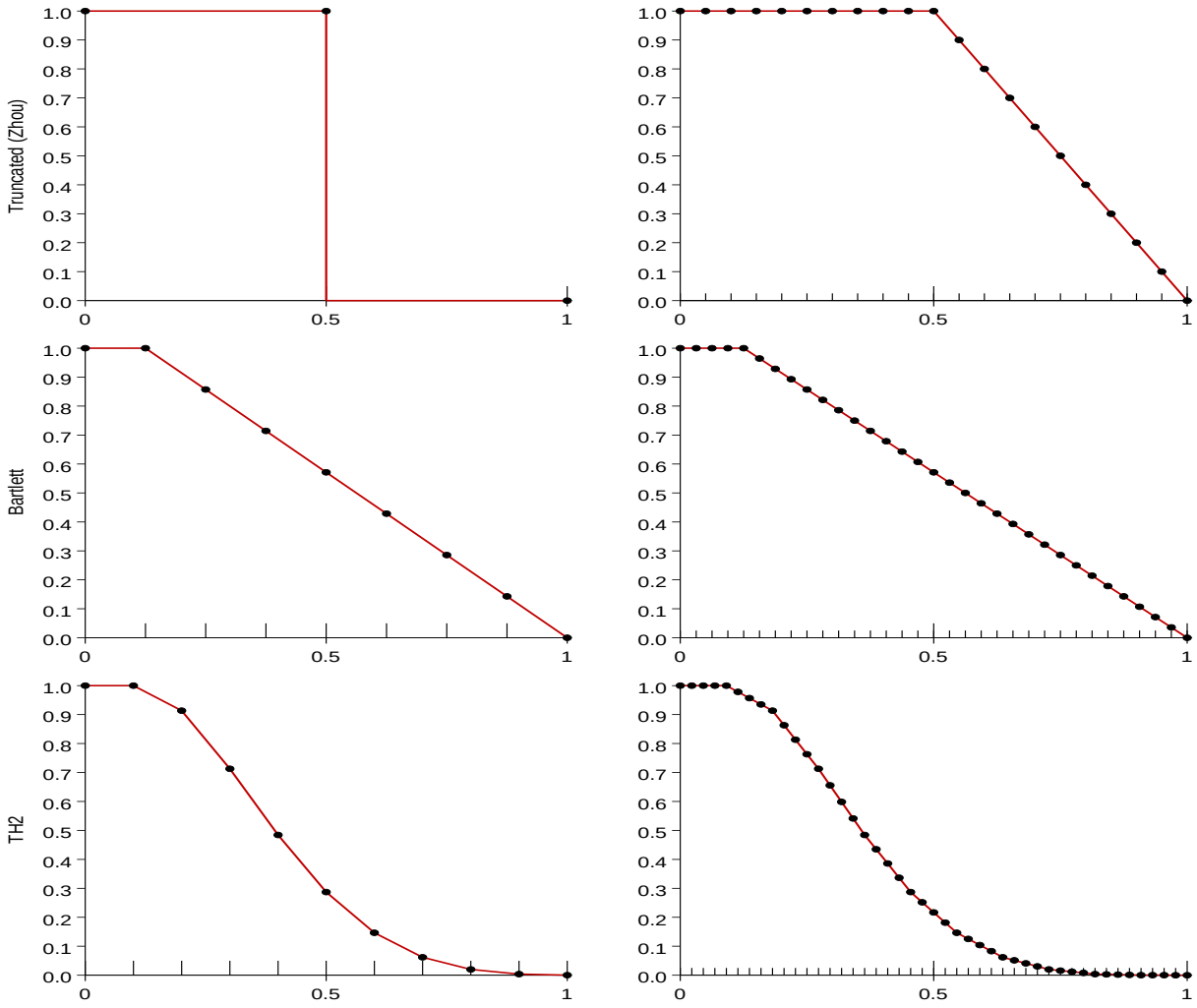

Figure 3: The effects of subsampling some kernels. The left panels display the original kernel functions and the right panels display their implied kernel functions that are induced by subsampling. For the truncated (discontinuous) kernel the two are very different. So subsampling makes an important difference in this case. For the (kinked) Bartlett kernel the two are virtually identical, which suggests that subsampling has no effect on this kernel. Finally, for the smooth kernel in the lower panels subsampling has only a small effect by making the kernel function piecewise linear.

In Figure 3 we trace out the implied kernel weights for three subsampled realised kernels.

The left panels display the original kernel functions and right panels display the implied kernel functions. The first kernel is the truncated kernel ($H = 1$), where we see that subsampling leads to a substantially different implied kernel function. For the kinked Bartlett kernel we see that subsampling leads to the same kernel function. For the smooth TH_2 kernel function, we see that the original and implied kernel functions are fairly similar, however subsampling does impose some piecewise linearity which is the reason that subsampling of smooth kernels increases the asymptotic variance.

The connection between subsampled realised kernels and realised kernels is perhaps not too surprising, because Bartlett (1950) motivated his kernel with the subsampling idea, see also Anderson (1971, p. 512) and Priestley (1981, pp. 439–440), where the latter have a discussion of end effects. Similar relations are found between estimators of the long-run variance, for instance Politis, Romano, and Wolf (1999) noted that the subsample-estimator of Carlstein (1986) is identical to the moving block bootstrap estimator and the Jackknife estimator in this context. An interesting observation from our analysis is that subsampled kinked kernels are essentially unaffected, whereas subsampling changes the shape of smooth kernels in an unfortunate way.

6 The case with discontinuous kernel functions

In this section we consider the kernel functions we have labelled as discontinuous kernels. Such kernels lead to estimators with poor asymptotic properties. We shall see that subsampling can substantially improve such estimators, making them consistent with mixed Gaussian distributions. So for such kernels, subsampling is a saviour.

Lemma 1 *Let* $\tilde{K}_w(X_\delta) = \sum_{h=0}^{H} w_h \tilde{\gamma}_h(X_\delta)$, *where* $H = o(n)$ *(possibly constant). Then*

$$w_0 = 1 + o(1) \qquad \textit{and} \qquad w_0 - w_1 = o(n^{-1}),$$

are necessary conditions for $E\left(\tilde{K}_w(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right) \to 0$; *and*

$$\sum_{h=0}^{H} (w_{h+1} - 2w_h + w_{h-1})^2 = o(n^{-1}), \tag{26}$$

is a necessary condition for $\mathrm{Var}\left(\tilde{K}_w(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right) \to 0$, *where we set* $w_{H+1} = 0$ *and* $w_{-1} = w_0$.

The lemma shows that realised kernels using a fixed H cannot converge to $\int_0^t \sigma_u^2 \mathrm{d}u$ in mean squares, because such estimators will not satisfy (26).

Now consider the case where we construct w_h from a kernel function and let $H \to \infty$, as we did in the previous section. In this situation it is clear that any discontinuous kernel will violate (26),

because

$$n\sum_{h=0}^{H}\left(w_{h+1}-2w_h+w_{h-1}\right)^2\simeq n\times J^2,$$

where

$$J^2=\sum_{x_j\in\mathcal{D}_k}\left\{\lim_{x\uparrow x_j}k(x)-\lim_{x\downarrow x_j}k(x)\right\}^2.$$

Here $\mathcal{D}_k$ is the set of discontinuity points for $k(x)$.

Next, we consider the truncated kernel which does not satisfies (26). We will see that subsampling this kernel produces an estimator that is consistent and mixed Gaussian. This is true whether H is finite or is allowed to grow with the sample size.

6.1 Zhou (1996) estimator

First we will look at estimators which are thought of as having H fixed and allowing the degree of subsampling to increase. This is outside the spirit of the realised kernels of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) which need H to get large with n_δ for consistency, however it is close to the important early work of Zhou (1996) and is strongly intellectually connected to the influential work on two scale estimators by Zhang, Mykland, and Aït-Sahalia (2005b).

The Zhou (1996) estimator can be written as

$$\gamma_0(X_\delta;S)+\widetilde{\gamma}_1(X_\delta;S)$$

which is the subsampled realised kernel based on the truncated kernel function using $H=1$. Zhou (1996) noticed that the variance of his estimator was of order $O(\frac{S}{n_\delta})+O(\frac{1}{S})+O(\frac{n_\delta}{S^2})$, but did not realize that by allowing S to increase with n_δ his estimator is consistent. In fact, in a subsequent paper Zhou stated that his subsampled realised kernels was inconsistent, see Zhou (1998, p. 114). The following Theorem gives its asymptotic distribution.

Theorem 6 *Suppose $S=c^3n_\delta^2$, then as $n_\delta\to\infty$*

$$n_\delta^{1/2}\left\{\gamma_0(X_\delta;S)+\widetilde{\gamma}_1(X_\delta;S)-\int_0^t\sigma_u^2\mathrm{d}u\right\}\overset{Ls}{\to}MN\left(0,\frac{16}{3}\int_0^t\sigma_u^4\mathrm{d}u+8\omega^4/c^3\right).$$

This asymptotics is not particularly attractive for its seeming $n_\delta^{1/2}$ rate of convergence hides the fact that it assumes massive databases in order to allow S to increase rapidly with n_δ. A more interesting way of thinking about this estimator is in terms of the effective sample size

$$n=S\times n_\delta.$$

Again we ask if it is better to increase n_δ or S for a given n. This leads to the following result.

Lemma 2 *If $S = c(\xi n)^{2/3}$ then the Zhou estimator has*

$$n^{1/6}\left(\gamma_0(X_\delta;S)+\widetilde{\gamma}_1(X_\delta;S)-\int_0^t\sigma_u^2\mathrm{d}u\right)\overset{Ls}{\to}MN\left(0,\omega^{4/3}\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3}\left(\tfrac{16}{3}c+\tfrac{8}{c^2}\right)\right).$$

The minimum asymptotic variance is

$$\underbrace{8\sqrt[3]{3}}_{\simeq 11.54}\omega^{4/3}\left(\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3},\quad \textit{with}\quad c=\sqrt[3]{3}.$$

The relationship $S = c(\xi n)^{2/3}$, which implies that $S \propto n_\delta^2$, gives the impression that this estimator will require massive subsampling to work, however if the noise is small this is not necessarily the case.

An interesting feature of the Zhou estimator is that its asymptotic variance is of the form obtained by the kinked non-subsampled realised kernels, i.e. ones which do not satisfy the $k'(0) = k'(1) = 0$ condition.

Example 2 *Suppose n corresponds to using prices every 1 second for a whole trading day on the NYSE, so $n = 23,400$. If $\omega^2 = 0.001$ and $t\int_0^t\sigma_u^4\mathrm{d}u = 1$, which is roughly right in empirical work from 2004 for thickly traded stocks, then for the Zhou estimator the optimal degree of subsampling is $S \simeq 25$ so that $n_\delta \simeq 920$. Thus the procedure is suggesting subsampled 25 second returns. Hence the degree of subsampling is rather modest. In 2000, $\omega^2 = 0.01$ and $\int_0^t\sigma_u^4\mathrm{d}u = 1$ would be more reasonable, in which case $S = 118$ and $n_\delta = 198$, which corresponds to returns measured every roughly 2 minutes.*

6.2 2-lag flat-top Bartlett estimator

A natural extension of Zhou (1996) is to allow H to be larger than one but fixed. The theory of realised kernels suggested this may well produce more efficient estimators, which we now show is true.

Lemma 3 *Let $w_0 = w_1 = 1$ and $w_2 = 1/2$. With $S = c(\xi n)^{2/3}$ we have*

$$n^{1/6}\left\{\gamma_0(X_\delta;S)+\widetilde{\gamma}_1(X_\delta;S)+\frac{1}{2}\widetilde{\gamma}_2(X_\delta;S)-\int_0^t\sigma_u^2\mathrm{d}u\right\}\overset{Ls}{\to}MN\left(0,\omega^{4/3}\left(t\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3}\left(\frac{20}{3}c+\frac{2}{c^2}\right)\right),$$

and the minimum variance is

$$\underbrace{10\sqrt[3]{3/5}}_{\simeq 8.43}\omega^{4/3}\left(\int_0^t\sigma_u^4\mathrm{d}u\right)^{2/3},\quad \textit{with}\quad c=\sqrt[3]{3/5}.$$

The constant in the asymptotic variance is here reduced from about 11.54 to 8.43. So this estimator is quite a bit more efficient than the Zhou estimator.

In the previous Theorem we added $w_2 = 1/2$ times $\tilde{\gamma}_2(X_\delta)$ to Zhou's estimator, which led to a reduction of the asymptotic variance. Now we proceed by adding additional realised autocovariances to Zhou's estimator, using the Bartlett weights, $w_h = k(\frac{h-1}{H})$, $h = 2, \ldots, H$. An interesting question is what happens as we increase H further? For moderately large H we have that

$$n^{1/6}\left\{\tilde{K}(X_\delta) - \int_0^t \sigma_u^2 \mathrm{d}u\right\}$$

has an asymptotic variance of approximately

$$\frac{4}{3}\left\{2 + (H+1)\right\} c \int_0^t \sigma_u^4 \mathrm{d}u + \frac{8\omega^4}{c^2 H^2}.$$

This suggests

$$c^3 = \frac{12\omega^4}{H^3 \int_0^t \sigma_u^4 \mathrm{d}u} + o(1),$$

so the asymptotic variance (using $\frac{4}{3}12^{1/3} + 8/12^{2/3} = 2\sqrt[3]{12}$) is

$$\underbrace{2\sqrt[3]{12}}_{\simeq 4.58}\omega^{4/3}\left(\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3} + o(1).$$

So we achieve an additional reduction of the asymptotic variance. Not surprisingly, this is the asymptotic variance of the Bartlett realised kernel applied to a sample of size n when $H \propto n^{2/3}$, see Example 1. Here, by allowing H to grow we approach the situation with kinked kernels so we observe the eventual impotence of subsampling – a property we have shown holds for all kinked kernels. Hence as H gets large the optimal degree of subsampling rapidly falls and the best thing to do is simply to run a Bartlett realised kernel on the data without subsampling, i.e. take $n_\delta = n$.

Figure 4 shows the implied kernel functions that are generated by subsampling Zhou's estimator ($H = 1$) and the two estimators that have been enhanced by adding Bartlett weights. The relative asymptotic efficiency for these estimators are simply given by $k_\bullet^{0,0}$ of the implied kernel. From Figure 4 it is evident that $k_\bullet^{0,0}(H=1) > k_\bullet^{0,0}(H=2) > k_\bullet^{0,0}(H=\infty)$ which explains that the asymptotic variance of this estimator is decreasing in H.

6.3 Relationship to two scale estimator

The main idea in the two scale estimator of Zhang, Mykland, and Aït-Sahalia (2005b) was to use $\gamma_0(X_\delta; S)$ and bias correct it using the estimator

$$\hat{\omega}^2 = \frac{\gamma_0(X_{\delta/S})}{2n}$$

which exploits very high frequency returns over time intervals of length δ/S. Their results are reproved here, exploiting our previous results to make the proofs very short.

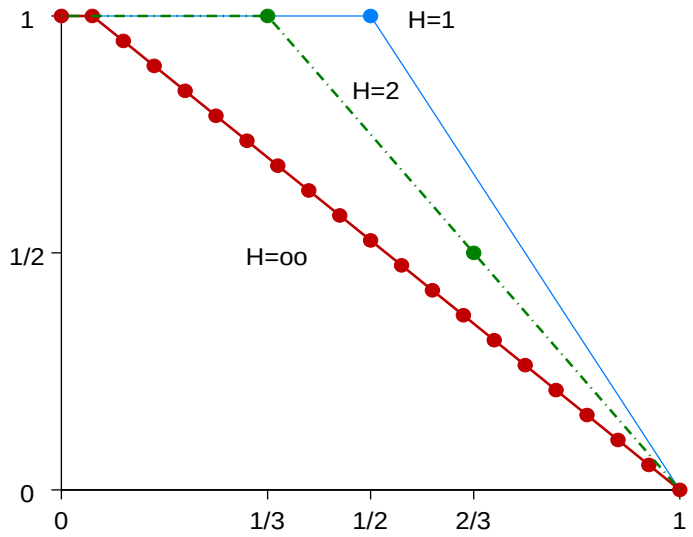

Figure 4: The implied kernels that arise from subsampling some simple kernels. $H = 1$ corresponds to the subsampled version of Zhou's estimator; $H = 2$ is that for Zhou's estimator after adding $1/2\tilde{\gamma}_2(X_\delta)$; and $H = \infty$ (here approximated by $H = 18$) illustrates the implied kernel for Zhou's estimator that is enhanced by an increasing number of Bartlett-weighted realised autocovariances.

We set

$$S = c^3\xi^2 n_\delta^2, \qquad \text{or equivalently} \qquad S = c(\xi n)^{2/3},$$

which imposes the optimal rate for S in this context.

Theorem 7 *With $S = c(\xi n)^{2/3}$ we have*

$$n^{1/6}\left\{\gamma_0(X_\delta;S) - n_\delta 2\omega^2 - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \overset{Ls}{\to} MN\left\{0, \omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}\left(\frac{4}{3}c + 4\frac{1+\lambda^2}{c^2}\right)\right\}. \quad (27)$$

This asymptotic result is infeasible in the sense that $\gamma_0(X_\delta;S) - n_\delta 2\omega^2$ is not an estimator of $\int_0^t \sigma_u^2 \mathrm{d}u$, because it involves the unknown parameter, ω^2. Shortly we will analyse the feasible estimator, where $\hat{\omega}^2 = \gamma_0(X_{\delta/S})/2n$ is substituted for ω^2. The following result address the asymptotic consequences of this substitution.

Theorem 8 *With $S = c(\xi n)^{2/3}$ we have*

$$n^{1/6}\left\{\begin{array}{c} \frac{1}{S}\sum_{j=1}^{Sn_\delta}\left(U_{j\delta/S} - U_{(j-S)\delta/S}\right)^2 - n_\delta 2\omega^2 \\ \frac{1}{S}\sum_{j=1}^{Sn_\delta}\left(U_{j\delta/S} - U_{(j-1)\delta/S}\right)^2 - n_\delta 2\omega^2 \end{array}\right\} \overset{L_s}{\to} N\left\{0, \frac{4\omega^4}{c^2\xi^{4/3}}\left(\begin{array}{cc} 1+\lambda^2 & \lambda^2 \\ \lambda^2 & 1+\lambda^2 \end{array}\right)\right\}.$$

The structure of the asymptotic covariance matrix can now be exploited to eliminate the nuisance parameter, λ^2. The implication is

$$n^{1/6}\left\{\frac{1}{S}\sum_{j=1}^{Sn_\delta}\left(U_{j\delta/S} - U_{(j-S)\delta/S}\right)^2 - \frac{1}{S}\sum_{j=1}^{Sn_\delta}\left(U_{j\delta/S} - U_{(j-1)\delta/S}\right)^2\right\} \overset{L_s}{\to} N\left(0, \frac{8\omega^4}{c^2\xi^{4/3}}\right),$$

allowing $\gamma_0(U_\delta;S) - n_\delta 2\omega^2$ to be replaced by $\gamma_0(U_\delta;S) - n_\delta 2\hat{\omega}^2$, yielding a feasible estimator which remarkably also reduces the variance compared to the infeasible estimator. This is stated next.

Theorem 9 *With $S = c(\xi n)^{2/3}$ we have*

$$n^{1/6}\left\{\gamma_0(X_\delta;S) - 2n_\delta\hat{\omega}^2 - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \overset{Ls}{\to} MN\left\{0, \omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}\left(\frac{4}{3}c + \frac{8}{c^2}\right)\right\}.$$

The minimum asymptotic variance is

$$\underbrace{2\sqrt[3]{12}}_{\simeq 4.58}\omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}, \quad \textit{with} \quad c = \sqrt[3]{12}.$$

Thus the two scale estimator is significantly more efficient than the Zhou estimator and is as efficient as the Bartlett realised kernel with H lags where $H = c(\xi n)^{2/3}$. Interesting this is solely due to replacing ω^2 by $\hat{\omega}^2$ — for if we used (27) it would deliver a less efficient estimator than the Zhou estimator in the case of Gaussian noise where $\lambda^2 = 2$. In effect $\hat{\omega}^2$ plays the role of a control variable, reducing the variance of the estimator.

Example 3 *(continued from Example 2). If $\omega^2 = 0.001$ and $t\int_0^t \sigma_u^4 \mathrm{d}u = 1$, then $S \simeq 40$ and $n_\delta \simeq 580$. Hence the degree of subsampling is larger than that used by Zhou.*

The two scale estimator by Zhang, Mykland, and Aït-Sahalia (2005b) combines a subsampled realised variance with $\gamma_0(X_\delta)$ for some delta. So it follows from our results in Section 5 that this estimator (apart from end effects) is a realised kernel – in this case the implied kernel is a Bartlett kernel. The two scales estimator converges at rate $n^{1/6}$, whereas the related multiscale estimator by Zhang (2006) converges at the efficient rate $n^{1/4}$. The latter combines multiple subsampled realised variances, each using a different S. So the multiscale estimator can also be expressed as a realised kernel. In this case the implied kernel is a linear combination of Bartlett kernels using different lag lengths. Interestingly, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) have shown that the multiscale estimator is asymptotically equivalent to the realised kernel based on the cubic kernel function, see Table 1. Not surprisingly, it can be shown that the implied kernel for the multiscale estimator is the cubic kernel.

7 Some Empirical Recommendations

So far we have worked under the assumption that the noise is of the independent type defined in (5). This assumption seems reasonable for equity returns when prices are sampled at moderate high frequencies. For instance, for the liquid stocks in the Dow Jones Industrial Average this assumption seems reasonable when applied to 1 minute returns, see Hansen and Lunde (2006). In this context, our theoretical results have shown that the best approach to estimation is to use a smooth realised kernel without any subsampling. This, conveniently, permits one to use the feasible methods for inference developed in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). A shortcoming of

this approach is that this estimator does not make use of all available observations. For example, transactions on the most liquid stocks now take place every few seconds, but for $U \in \mathcal{WN}$ to be reasonable we can only sample every, say, 15th observation.

In this Section we discuss how to construct subsampled realised estimators that make use of all available data. We also discuss how valid inference can be made about such estimators under realistic assumptions about the noise in tick-by-tick data.

The main idea is to use a subsampled realised kernel, where S is chosen to be sufficiently large so that (5) is reasonable for a sample that only consists of every Sth observation. The asymptotic variance can be estimated from the coarsely sampled data, using the methods by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006), and this leads to valid inference that is robust to both time-dependent and endogenous noise in the tick-by-tick data.

Specifically we recommend the following procedure.

1. Choose S sufficiently large for (5) to be a plausible assumption for a sample that only consists of every Sth observation. One can check for violation of (5) by applying the diagnostics used in Hansen and Lunde (2006).

2. Construct S distinct subsamples, by jittering the initial starting point and sampling every Sth observation. So each subsample has approximately $n_\delta = n/S$ observations.

3. For each of the S subsamples, obtain estimates of ω^2 and $IQ = t\int_0^t \sigma_u^4 \mathrm{d}u$, and an initial estimate of $IV = \int_0^t \sigma_u^2 \mathrm{d}u$. See Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) for ways to do this. Average each of these estimators to construct the subsampled estimators, $\hat{\omega}^2 = S^{-1}\sum_{s=1}^S \hat{\omega}_s^2$ and $\widehat{IV}_{\text{initial}} = S^{-1}\sum_{s=1}^S \widehat{IV}_{\text{initial},s}$ and $\widehat{IQ} = S^{-1}\sum_{s=1}^S \widehat{IQ}_s$.

4. Obtain an estimate, $\hat{H}$, for the optimal H, by plugging the subsampled estimates into the expression for the optimal H. Use this $\hat{H}$ to compute the S realised kernels, $\tilde{K}^s(X_{\text{skip-}S})$, using a smooth kernel and the weights $w_0 = w_1 = 1$ and $w_h = k\left(\frac{h-1}{\hat{H}}\right)$, for $h = 2, \ldots, \hat{H}$. Form their average to obtain the actual estimator, $\widehat{IV}_{\text{final}} = \tilde{K}(X_{\text{skip-}S}; S)$.

5. Finally, compute the conservative estimate for avar $\left\{\tilde{K}(X_{\text{skip-}S}; S)\right\}$ using the finite sample expressions

$$\widehat{\text{Var}}\left\{\tilde{K}(X_{\text{skip-}S}; S)\right\} = \widehat{IQ}\left(w'Aw\right) \times \frac{1}{n_\delta} + 8\hat{\omega}^2\widehat{IV}_{\text{final}}\left(w'Bw\right) + 4\hat{\omega}^4\left(w'Cw\right) \times n_\delta, \qquad (28)$$

where $w = (w_0, w_1, \ldots, w_{\hat{H}})$. Here one can use that

$$w'Aw \;=\; 2 + 4\sum_{h=1}^{H}(w_h)^2 \simeq 4Hk_\bullet^{0,0},$$

$$
\begin{aligned}
w'Bw &= 1 + 2\sum_{h=2}^{H} w_h(w_h - w_{h-1}) \simeq H^{-1}k_\bullet^{1,1}, \\
w'Cw &= 4 + \sum_{h=2}^{H} w_h(6w_h - 8w_{h-1} + 2w_{h-2}) \simeq H^{-3}k_\bullet^{2,2}.
\end{aligned}
$$

The variance estimate in (28) is the sum of the finite sample versions of (14-16) with $S = 1$. So this expression completely ignores subsampling, and the expression is really an estimator of $\text{Var}(\tilde{K}^s(X_{\text{skip-}S}))$. The reason is that subsampling does not reduce the noise-variance by a factor of S, unless the noise is uncorrelated across subsamples. This is unrealistic when the subsamples exploit all the tick-by-tick data. However, we still have

$$
\text{avar}\left\{\tilde{K}(X_{\text{skip-}S}; S)\right\} \leq \text{avar}(\tilde{K}^s(X_{\text{skip-}S})),
$$

even if $U_t \perp\!\!\!\perp U_s$ is violated for some $s \neq t$. So (28) is simply a robust estimator that is expected to yield a conservative estimate of the variance. It is interesting to have some notion of how conservative this estimator is.

Recall our result in Theorem 3 that $\text{avar}\left\{\tilde{K}(Y_{\text{skip-}S}; S)\right\} = \text{avar}(\tilde{K}^s(Y_{\text{skip-}S}))$, see (14). So subsampling cannot reduce the contribution to the asymptotic variance from this term, while the contributions from the two other terms (15) and (16), potentially can be driven all the way to zero.

Consider the realised TH_2 kernel. When $\rho = 1$ its asymptotic variance is proportional to

$$
c_1 + 2\frac{k_\bullet^{1,1}}{k_\bullet^{0,0}}c_1^{-1} + \frac{k_\bullet^{2,2}}{k_\bullet^{0,0}}c_1^{-3} = 5.75 + \frac{1.71}{0.218}\frac{2}{5.75} + \frac{41.8}{0.218}(5.75)^{-3} \simeq 9.50.
$$

Subsampling this estimator with $S = 10$, say, will reduce this term to a number no less than

$$
5.75 + \frac{1}{10}\frac{1.71}{0.218}\frac{2}{5.75} + \frac{1}{10}\frac{41.8}{0.218}(5.75)^{-3} \simeq 6.12.
$$

Hence the variance reduction will be less than 36%, and even with $S \to \infty$ the reduction will be less than 40%. In practice, the reduction is likely to be much smaller, because the noise in the different subsamples is dependent. So even though (28) is a conservative estimator – it is not perversely conservative.

8 Simulation study

In this section we analyse the finite sample properties of $\tilde{K}(X_\delta; S)$, using both a smooth TH_2 kernel and a kinked Bartlett kernel. We are particularly interested in the MSE of $\tilde{K}(X_\delta; S)$, as a function of δ and S, and to see whether the simulation based results differs from our theoretical results in any significant way.

8.1 Simulated model

We consider the following SV model,

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

where ρ is a leverage parameter. This model is frequently used for simulation is this context, see e.g. Huang and Tauchen (2005), Goncalves and Meddahi (2004), and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).

In our simulated model, we set $\mu = 0.03$, $\beta_1 = 0.125$, $\alpha = -0.025$ and $\rho = -0.3$. Further, we set $\beta_0 = \beta_1^2/(2\alpha)$ in order to standardize $\mathrm{E}\left(\sigma_t^2\right) = 1$. With this configuration the variance of $\int_0^t \sigma_u^2 du$ is comparable to the empirical results found in Hansen and Lunde (2005). For the variance of market microstructure noise we set $\omega^2 = 0.1$.

8.2 Design

The process is generated using an Euler scheme based on $N = 23,400$ intervals, where each interval is thought to correspond to one second so that the entire interval corresponds to 6.5 hours, which is the length of a typical trading day.[1] The volatility process is restarted at its mean value $\sigma_0 = 1$ every day by setting $\tau_0 = 5/2$. This keeps the noise-to-signal ratio, $\xi = \omega^2/\sqrt{\int_0^1 \sigma_u^4 du}$, comparable across simulations. In our Monte Carlo designs we let the effective sample size, n, be either $1,560$, $4,680$, or $23,400$, which correspond to sampling every 15, 5, or 1 seconds, respectively. So a sample with $4,680$ observations, say, is obtained by including every fifth observation of the $N = 23,401$ simulated data points over the $[0, t]$ interval.

8.3 Implementation of realised kernels and subsampled realised kernels

From the simulated data, $X_0, \ldots, X_n$, we define the "skip-S returns" $\Delta_S X_j = X_j - X_{j-S}$. The subsampled realised autocovariances are computed by,

$$\hat{\gamma}_h^s = \sum_{j=1}^{n_\delta} \Delta_S X_{jS+s-1} \Delta_S X_{(j-h)S+s-1}, \qquad s = 1, \ldots, S, \qquad h = -H, \ldots, 0, \ldots, H,$$

where $n_\delta = n/S$. The subsampled realised kernel is defined by

$$\widehat{\tilde{K}(X;S)} = \frac{1}{S} \sum_{s=1}^{S} \widehat{\tilde{K}^s(X)}, \qquad \text{where} \qquad \widehat{\tilde{K}_H^s(X)} = \hat{\gamma}_0^s + \sum_{h=1}^{H} k(\tfrac{h-1}{H}) \left(\hat{\gamma}_h^s + \hat{\gamma}_{-h}^s\right).$$

So for $S = 1$ we simply have

$$\widehat{\tilde{K}_H(X)} = \hat{\gamma}_0 + \sum_{h=1}^{H} k(\tfrac{h-1}{H}) \left(\hat{\gamma}_h + \hat{\gamma}_{-h}\right), \quad \text{where} \quad \hat{\gamma}_h = \sum_{j=1}^{n} x_j x_{j-h}.$$

[1] In practice we generate intraday returns for $33,400$ intervals, and treat the first and and last $5,000$ returns as out-of-period returns ($x_{-1}, x_{-2}, \ldots$ and $x_{N+1}, x_{N+2}, \ldots$).

We use our expression for the optimal choice for H. So when $S = 1$ we use $H^*_{\text{TH}_2,1} = 5.75(\xi n)^{1/2}$ for the smooth TH_2 kernel and $H^*_{\text{Bartlett},1} = \sqrt[3]{12(\xi n)^2}$ for the kinked Bartlett kernel. The "noise-to-signal" parameter, $\xi = \omega^2/\sqrt{\int_0^1 \sigma_u^4 \mathrm{d}u}$ need not be estimated in our simulations, because ω^2 is known and the integrated quarticity, $\int_0^1 \sigma_u^4 \mathrm{d}u \simeq N \sum_{j=1}^N \sigma_{j/N}^4$, can be computed from the simulated data. The parameter $\rho = \int_0^1 \sigma_u^2 \mathrm{d}u / \sqrt{\int_0^1 \sigma_u^4 \mathrm{d}u}$ can be computed from the simulated volatility path. When $S \geq 2$ the optimal H for the Bartlett kernel is simply given by $H^*_{\text{Bartlett},S} = S^{-1}\sqrt[3]{12(\xi n)^2}$, and the TH_2 kernel has $H^*_{\text{TH}_2,S} = c_S^{1/2}(\xi n)$, where

$$c_S = S^{-1}\sqrt{7.84\rho\left(1 + \sqrt{1 + 9.33S}\right)},$$

as defined in (24).

8.4 Simulation Results

Figure 5 shows the Monte Carlo results with the number of subsamples, S, increasing along the horizontal axis and the MSE on the vertical axis. The lines represent different sample sizes.

Consider first the results based on the Bartlett kernel. Our theoretical results in Theorem 4 dictate that these lines should be horizontal. This result is confirmed, especially for the large sample size $n = 23,400$. Still, a small increase in the MSE as S increases is observed for the smaller sample sizes. The reason is that the lag length of the implied kernel, H_{implied}, can only attain values that are divisible by S. While the Bartlett kernel without subsampling has $H^*_{\text{Bartlett},1} = \left\lceil \sqrt[3]{12(\xi n)^2} \right\rceil$, the implied Bartlett kernel has $H_{\text{implied}} = S \times \left\lceil S^{-1}\sqrt[3]{12(\xi n)^2} \right\rceil$. So as S increases the implied kernels' H_{implied} is more likely to deviate from $H^*_{\text{Bartlett},1}$, which causes an increase in the mean squared error. The smaller is the sample size, n, the smaller is the optimal value for H. So it is not surprising that the impact on MSE is seen earlier when n is small. In this design, the optimal lag length, $H^*_{\text{Bartlett},1}$, is about 67, 140, and 403, for $n = 1,560$, $n = 4,680$, and $n = 23,400$, respectively. Though the is some variation in the optimal H across simulations because it through ξ, depends on the simulated volatility path. The lower panels present the results for the smooth TH_2 kernel. Here, our theoretical results in Theorem 5 state that the MSE is increasing in S, and this phenomenon is evident for all sample sizes. For each of the sample sizes, $n = 1,560$, $n = 4,680$, and $n = 23,400$, the optimal $H^*_{\text{TH}_2,1}$, is typically 72, 125, and 279, respectively. The results when $\omega^2 = 0.01$ and $\omega^2 = 0.001$ (not reported) are similar. Here the optimal H is smaller and this causes subsampling to have a larger impact on the MSE. Naturally, the implied kernels must have $H_{\text{implied}} \geq S$, so that $H_{\text{implied}} = S$ whenever $S \geq H^*$. This constraint is relevant for our simulations with small levels of noise because subsampling takes H_{implied} further away from its optimal value, as S increases beyond the optimal H.

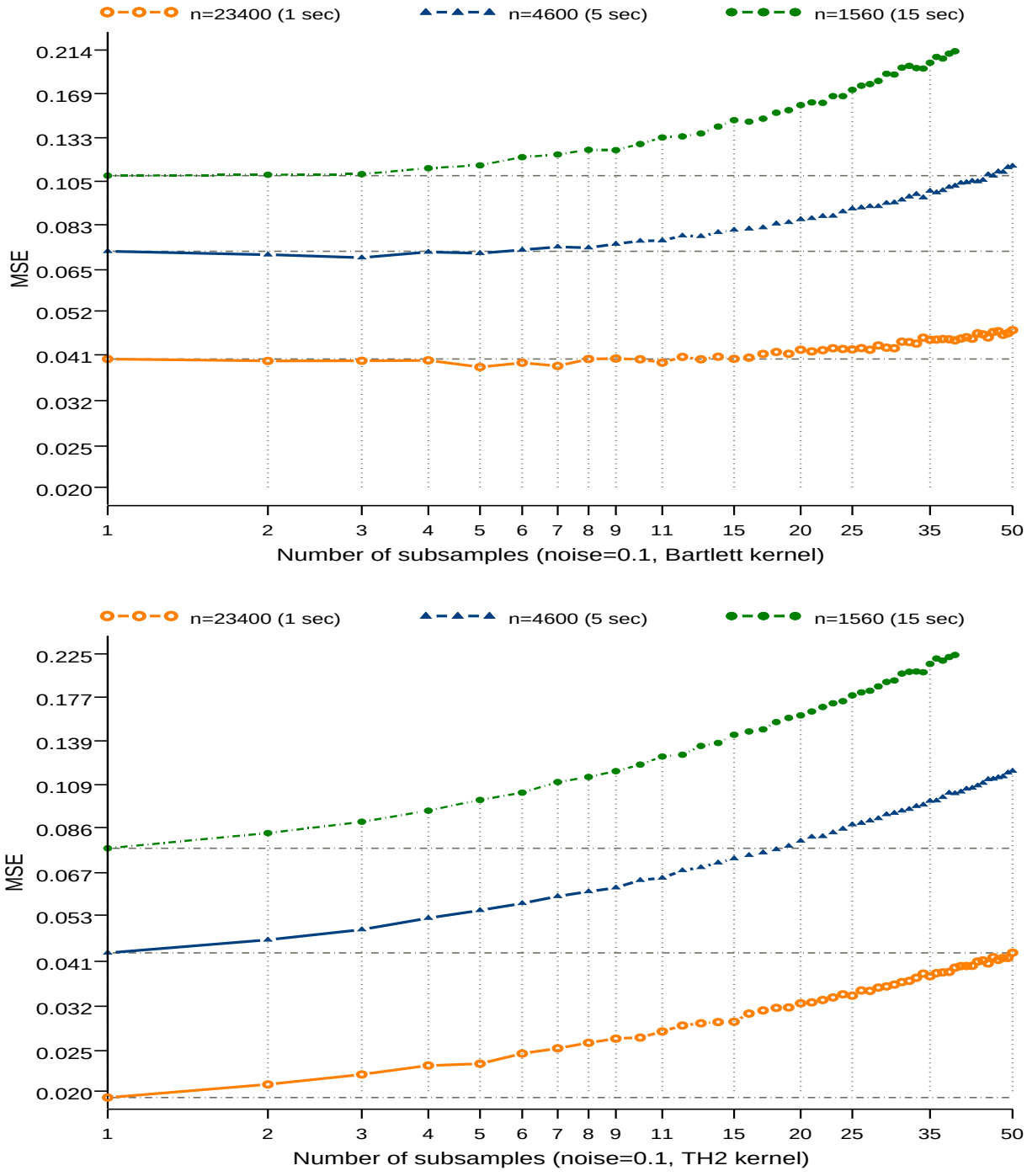

Figure 5: Mean squares errors (MSEs) for subsampled realised kernels using three different sample sizes. The upper panel presents the results for the Bartlett kernel and the lower panels presents the results for the TH_2 kernel. For the (kinked) Bartlett kernel we see that the MSE is fairly insensitive to S, whereas the (smooth) TH_2 kernel has MSEs that are slightly increasing in S. These findings are fully consistent with the theoretical results in Theorems 4 and 5.

9 Empirical study of General Electric trades

In this section we revisit the empirical application in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). Our objective is to compare subsampled realised kernels with standard realised kernels and other estimators. The estimation problem is here to estimate the daily increments of $[Y]$ for the logarithm price of General Electric (GE) shares, using high frequency transaction data carried out on the New York Stock Exchange in 2000 and in 2004. The reason that we analyse data from both periods is that the variance of the noise was around 10 times higher in 2000 than in 2004. A more detailed analysis on 29 other major stocks is provided in a Web Appendix to this paper available from `www.hha.dk/~alunde/bnhls/bnhls.htm`. This appendix also describes the exact implementation of our estimators. Precise details on the cleaning we carried out on the raw data before it was analysed are described in the web appendix to Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).

Table 3 shows Summary statistics for seven estimators. The first estimator is the realised TH_2 kernel using approximate 1 minute returns. The approximate 1 minute returns are obtained by skipping a fixed number of transactions, such that the average time between observations is one minute. In 2000 we had to skip every 9.7 observations on average to construct the approximate 1 minute returns, and in 2004 we had to skip every 13.7 observations on average. The second estimator is the subsampled realised TH_2 kernel. So in 2000 we have $S \simeq 9.7$ and in 2004 we have $S \simeq 13.7$. The third estimator is the realised TH_2 kernel that is based on tick-by-tick data (i.e. all available trades) and an H that is S times larger than that used by the first estimator. This estimator should be quite similar to the subsampled realised kernel, according to our results in Section 5.

The following three estimators are subsampled realised variances. These are based on returns that are sampled in calendar time, where each intraday return spans 20 minutes, 5 minutes, or 1 minute, as indicated in the subscript of these estimators. To exhaust data sampled every second, the number of subsamples are $S = 1200$, $S = 300$, and $S = 60$, respectively. For instance, the estimator $[X_{5 \text{ minutes}}; 300]$ is the average of 300 realised variances, where each of the realised variances are based on 5 minute intraday returns, simply changing the initial place that prices are recorded by one second. The last estimator, $TSRV(K, J)$, by Aït-Sahalia, Mykland, and Zhang (2005), is the two-scale estimator that is designed to be robust to deviations from i.i.d. noise. Here we use their *area adjusted* estimator, which involves a bias correction.

Table 3: Summary statistics for subsampled $[Y]$ estimators, GE.

	Mean	Std. (HAC)	$\overline{H}$	$\rho([\widehat{Y}], \widetilde{K})$	acf(1)	acf(2)	acf(5)	acf(10)
			Sample period: 2000					
Modified Tukey-Hanning kernel ($\widetilde{H} = cn^{1/2}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{\text{ap. 1 min}})$	4.747	3.216 (6.133)	6.558	1.000	0.43	0.25	0.03	0.15
Subsampled Modified Tukey-Hanning kernel ($H = cn^{1/2}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{\text{ap. 1 min}}; S)$	4.709	3.220 (6.170)	6.558	0.997	0.43	0.25	0.03	0.16
Modified Tukey-Hanning kernel ($H = S \cdot \widetilde{H}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{1\text{ tick}})$	4.702	2.946 (5.793)	62.44	0.986	0.46	0.27	0.05	0.13
Simple RV subsampled								
$[X_{20\text{ minutes}}; 1200]$	4.417	3.650 (6.046)		0.894	0.26	0.17	-0.01	0.17
$[X_{5\text{ minutes}}; 300]$	4.908	3.018 (5.611)		0.984	0.44	0.23	0.01	0.14
$[X_{1\text{ minutes}}; 60]$	5.545	2.376 (5.167)		0.787	0.55	0.36	0.10	0.18
AMZ (2006)								
$TSRV(K, J)$	3.511	2.846 (5.265)		0.941	0.36	0.21	0.01	0.23
$TSRV(K, J)$ - *aa*	4.514	3.657 (6.766)		0.941	0.36	0.21	0.01	0.23
			Sample period: 2004					
Modified Tukey-Hanning kernel ($\widetilde{H} = cn^{1/2}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{\text{ap. 1 min}})$	0.962	0.568 (1.195)	5.723	1.000	0.34	0.32	0.28	0.08
Subsampled Modified Tukey-Hanning kernel ($H = cn^{1/2}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{\text{ap. 1 min}}; S)$	0.954	0.561 (1.202)	5.723	0.995	0.37	0.32	0.28	0.09
Modified Tukey-Hanning kernel ($H = S \cdot \widetilde{H}$)								
$\widetilde{K}_w^{\text{TH2}}(X_{1\text{ tick}})$	0.947	0.522 (1.130)	78.27	0.990	0.37	0.31	0.30	0.08
Simple RV subsampled								
$[X_{20\text{ minutes}}; 1200]$	0.885	0.516 (1.036)		0.933	0.27	0.27	0.27	0.08
$[X_{5\text{ minutes}}; 300]$	0.943	0.503 (1.088)		0.984	0.37	0.32	0.30	0.08
$[X_{1\text{ minutes}}; 60]$	0.942	0.376 (0.921)		0.899	0.46	0.43	0.38	0.12
AMZ (2006)								
$TSRV(K, J)$	0.736	0.436 (0.929)		0.944	0.33	0.35	0.28	0.11
$TSRV(K, J)$ - *aa*	0.946	0.560 (1.194)		0.944	0.33	0.35	0.28	0.11

Summary statistics for three types of kernel based estimators: First the realised Modified Tukey-Hanning kernel using approximate 1 minute returns. Then the corresponding subsampled kernel. Next, the kernel computed using the inefficient rate and based on all available trades, after which a version with $H = \widetilde{H} \cdot S$ follows. Next, subsampled versions of simple RV statistics based on 20, 5 and 1 minute returns are given. For instance, the subsampled $[X_{5\ minutes}; 300]$ calculates RV over 5 minutes, averaged over 300 times, just changing the initial place prices are recorded. The AMZ (2006) are two-scale estimators designed to be robust to deviations from i.i.d. noise. The second estimator which scales $TSRV(K, J)$ overcome its finite sample bias.

From Table 3 we see that the estimators are very tightly correlated. The two realised kernels and the subsampled realised kernel are almost perfectly correlated, and all reported statistics are quite similar for these estimators. The two scale estimator is also quite similar to the realised kernels. Interestingly, amongst the subsampled realised variances, it is that based on 5 minute returns that is most similar to the realised kernels. This suggest that 20 minute returns leads to too much sampling error, whereas 1 minute returns are being influenced by the bias due to market microstructure noise.

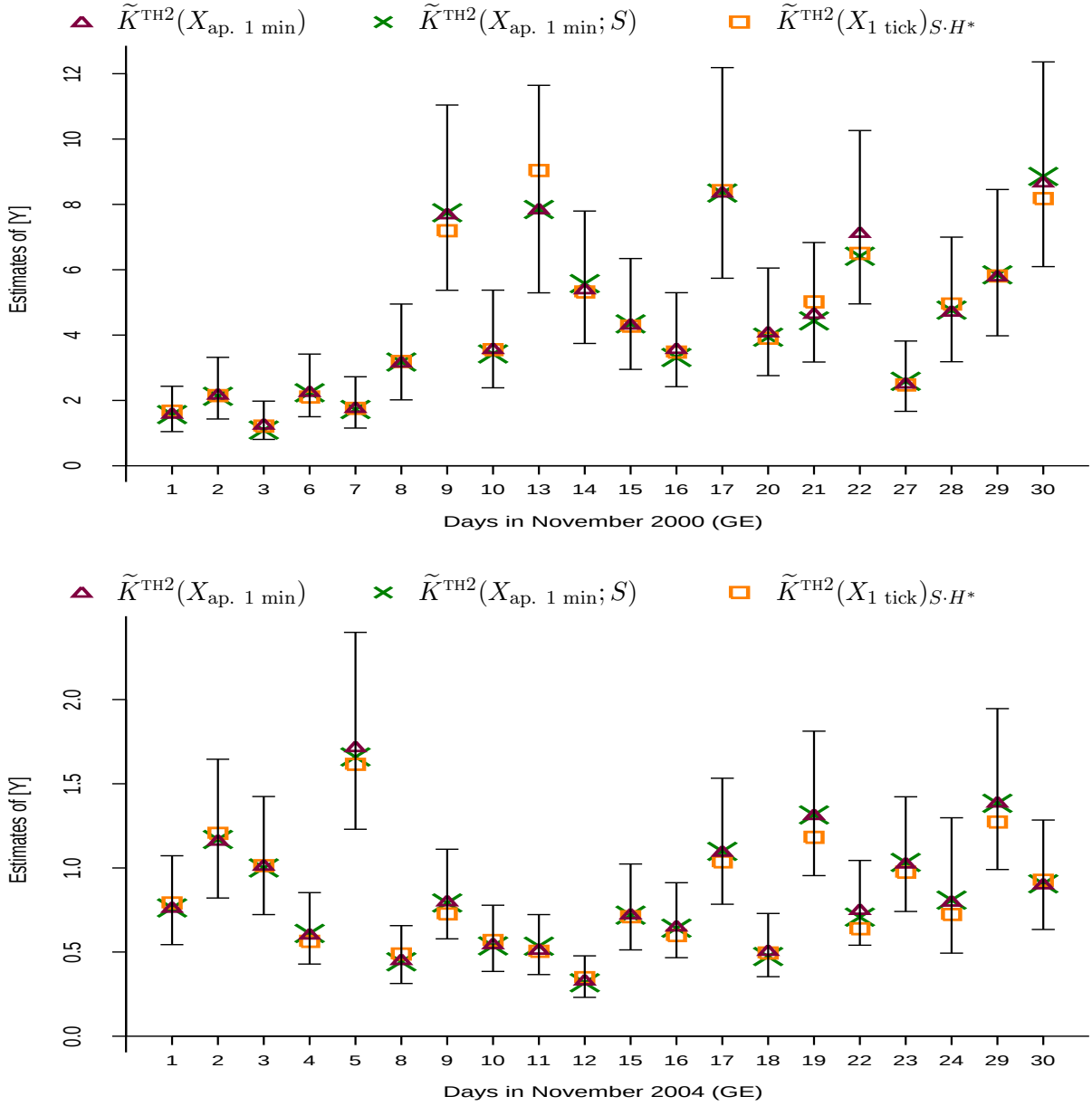

Figure 6: Three estimators for the daily increments to $[Y]$ for General Electrics in November 2000 and 2004. Triangles are the estimates of the realised kernel using roughly 1 minute returns. Diamonds are the estimates produced by the subsampled realised kernel. Circles are the estimates of the realised kernel that uses tick-by-tick returns and an H that is S times larger than that used by the first realised kernel. The intervals are the 95% confidence intervals for $\tilde{K}^{\text{TH2}}(X_{\text{ap. 1min}})$.

Time series for some of these estimators are drawn in Figure 6, where we plot daily point estimates for November 2000 and November 2004. We also include the confidence intervals for $\tilde{K}^{\text{TH2}}(X_{\text{ap. 1 min}})$ using the method of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). The three estimators are virtually almost identical. While the subsampled realised kernel may be slightly more precise than the moderately sampled realised kernel, $\tilde{K}^{\text{TH2}}(X_{\text{ap. 1 min}})$, Figure 6 does not suggest there is a big difference between these two. The realised kernel that is based on

tick-by-tick data is slightly different from the other estimators, but always inside the confidence interval for $\tilde{K}^{\text{TH2}}(X_{\text{ap. 1 min}})$.

10 Conclusions

In this paper we have studied the properties of subsampled realised kernels. Subsampling is a very natural addition to realised kernels, for it can be viewed as averaging over realised kernels with slightly different starts of the day. We have provided a first asymptotic study of the properties of subsampling for these statistics, allowing the degree of subsampling or the number of lags to go to infinity or being fixed. Included in our analysis is the asymptotic distribution of the estimator proposed by Zhou (1996).

Subsampling leads to surprisingly little gains in our analysis. In fact, we found that subsampling is harmful for the best class of realised kernel estimators. The main advantage of subsampling is that it can overcome the inefficiency that results from a poor choice of kernel weights in the first place. For example, when the truncated kernel is used to design estimators, the resulting estimator has poor asymptotic properties. whereas the subsampled estimator is consistent and converges at rate $n^{1/6}$.

In the realistic situation where the noise is endogenous and time dependent, subsampled realised kernels do provide a simple way to make use of all the available data. We have discussed how to make valid inference about such estimators.

We also provide a slightly different and rather simple way of thinking of the two scale estimator proposed by Zhang, Mykland, and Aït-Sahalia (2005b).

References

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). Ultra high frequency volatility estimation with dependent microstructure noise. Working Paper w11380, NBER.

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

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

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: John Wiley and Sons.

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

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

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

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2006). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Unpublished paper: Nuffield College, Oxford.

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.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrica 37*, 1–16.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics 14*, 1171–1179.

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

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

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

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

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

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

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

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

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

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

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

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

Oomen, R. A. A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics 24*, 219–237.

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

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. New York: Springer.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Volume 1. London: Academic Press.

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

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

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

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005a). Edgeworth expansions for realized volatility and related estimators. Unpublished paper, University of Illinois at Chicago.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005b). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*, 1394–1411.

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

Zhou, B. (1998). Parametric and nonparametric volatility measurement. In C. L. Dunis and B. Zhou (Eds.), *Nonlinear Modelling of High Frequency Financial Time Series*, Chapter 6, pp. 109–123. John Wiley Sons Ltd.

Appendix: Proofs

Lemma A.1 *We have*

$$\gamma_h(X_\delta; S) = \sum_{s=-S}^{S} (1 - \frac{|s|}{S})\gamma_{Sh+s}(X_{\frac{\delta}{S}}) + \frac{R_S^x}{S}.$$

The remainder R_S^x/S is a relatively small term, due to end effects. The term is defined explicitly in the proof, and the expression shows that R_S^x can be made zero by tweaking the first $S-1$ and last $S-1$ intraday returns.

Proof. Define the intraday returns $x_j = X_{\frac{\delta}{S}j} - X_{\frac{\delta}{S}j-\frac{\delta}{S}}$, and write

$$X_{\delta(j+\frac{s-1}{S})} - X_{\delta(j-1+\frac{s-1}{S})} = X_{\frac{\delta}{S}(jS+s-1)} - X_{\frac{\delta}{S}(jS-S+s-1)} = x_{jS+s-1} + \cdots + x_{jS-S+s}.$$

So $x_j,\ j = \ldots, 1, \ldots, Sn_\delta, \ldots$ are intraday returns over short intervals, each having length δ/S.

Consider first the usual realised autocovariance:

$$\begin{aligned}
\gamma_h^1(X_\delta) &= \sum_{j=1}^{n_\delta} \left(X_{\delta j} - X_{\delta(j-1)}\right)\left(X_{\delta(j-h)} - X_{\delta(j-h-1)}\right) \\
&= \sum_{j=1}^{n_\delta} \left(x_{jS-S+1} + \cdots + x_{jS}\right)\left(x_{jS-S+1-hS} + \cdots + x_{jS-hS}\right) \\
&= \sum_{j=1}^{n} x_j x_{j-Sh} \\
&\quad + \sum_{\substack{j=1 \\ j \bmod S \neq 0}}^{n} x_j x_{j-Sh+1} + \sum_{\substack{j=1 \\ j \bmod S \notin \{0,S-1\}}}^{n} x_j x_{j-Sh+2} + \cdots + \sum_{\substack{j=1 \\ j \bmod S = 1}}^{n} x_j x_{j-Sh+S-1} \\
&\quad + \sum_{\substack{j=1 \\ j \bmod S \neq 1}}^{n} x_j x_{j-Sh-1} + \sum_{\substack{j=1 \\ j \bmod S \notin \{1,2\}}}^{n} x_j x_{j-Sh-2} + \cdots + \sum_{\substack{j=1 \\ j \bmod S = 0}}^{n} x_j x_{j-Sh-S+1}.
\end{aligned}$$

Similarly for $s > 1$ we have

$$\gamma_h^s(X_\delta) = \sum_{j=1}^{n_\delta} \left(X_{\delta j+\frac{s-1}{S}} - X_{\delta(j-1)+\frac{s-1}{S}}\right)\left(X_{\delta(j-h)+\frac{s-1}{S}} - X_{\delta(j-h-1)+\frac{s-1}{S}}\right)$$

$$
\begin{aligned}
&= \sum_{j=1}^{n_\delta} \left(x_{jS-S+1+s-1} + \cdots + x_{jS+s-1}\right)\left(x_{jS-S+1+s-1-hS} + \cdots + x_{jS+s-1-hS}\right) \\
&= \sum_{j=s}^{n+s-1} x_j x_{j-Sh} \\
&\quad + \sum_{\substack{j=s \\ j \bmod S \neq s-1}}^{n+s-1} x_j x_{j-Sh+1} + \sum_{\substack{j=s \\ j \bmod S \notin \{s-1,s-2\}}}^{n+s-1} x_j x_{j-Sh+2} + \cdots + \sum_{\substack{j=s \\ j \bmod S = s}}^{n+s-1} x_j x_{j-Sh+S-1} \\
&\quad + \sum_{\substack{j=s \\ j \bmod S \neq s}}^{n+s-1} x_j x_{j-Sh-1} + \sum_{\substack{j=s \\ j \bmod S \notin \{s,1\}}}^{n+s-1} x_j x_{j-Sh-2} + \cdots + \sum_{\substack{j=s \\ j \bmod S = s-1}}^{n+s-1} x_j x_{j-Sh-S+1}.
\end{aligned}
$$

By adding up the terms we have

$$
\gamma_h(X_\delta; S) = S^{-1} \sum_{s=1}^{S} \gamma_h^s(X_\delta) = \sum_{s=-S+1}^{S-1} (1 - \frac{s}{S}) \gamma_{Sh+s}(X_{\frac{\delta}{S}}) + \frac{R_S^x}{S},
$$

where

$$
\begin{aligned}
R_S^x = -\sum_{S=2}^{S} &\left(\sum_{j=1}^{s-1} x_j x_{j-Sh} + \sum_{j=1}^{s-2} x_j x_{j-Sh+1} + \cdots + \sum_{j=1}^{1} x_j x_{j-Sh+S-2} \right. \\
&\left. + \sum_{j=1}^{s-1} x_j x_{j-Sh-1} + \sum_{j=2}^{s-1} x_j x_{j-Sh-2} + \cdots + \sum_{j=s-1}^{s-1} x_j x_{j-Sh-S+1} \right) \\
+ \sum_{S=2}^{S} &\left(\sum_{j=n+1}^{n+s-1} x_j x_{j-Sh} + \sum_{j=n+1}^{n+s-2} x_j x_{j-Sh+1} + \cdots + \sum_{j=n+1}^{n+1} x_j x_{j-Sh+S-2} \right. \\
&\left. + \sum_{j=n+1}^{n+s-1} x_j x_{j-Sh-1} + \sum_{j=n+2}^{n+s-1} x_j x_{j-Sh-2} + \cdots + \sum_{j=n+s-1}^{n+s-1} x_j x_{j-Sh-h+1} \right).
\end{aligned}
$$

The term, R_S^x, is due to end effects and involves much fewer cross products, $x_i x_j$, than does $\sum_{s=1}^{S} \gamma_h^s(X_\delta)$. So that R_S^x/S is typically negligible. In fact, R_S^x can be made zero by assuming $x_1 = \cdots = x_{S-1} = x_{n+1} = \cdots = x_{n+S-1} = 0$.
□

Proof of Theorem 2. By Lemma A.1 we have

$$
\tilde{\gamma}_h(Y_\delta; S) \simeq \sum_{s=-S}^{S} (1 - \frac{|s|}{S}) \tilde{\gamma}_{Sh+s}(Y_{\frac{\delta}{S}}),
$$

and the asymptotic properties of $\gamma_h(Y_{\frac{\delta}{S}})$, $h = -SH, \ldots, SH$, using the small time gaps, δ/S, follows straightforwardly from Theorem 1, (6).

Write

$$
V_{0,S} = \frac{1}{S}\left(1 + 2\sum_{s=1}^{S}\left(1 - \tfrac{s}{S}\right)^2\right) = \frac{2}{3}\left(1 + \tfrac{S^{-2}}{2}\right) \to \frac{2}{3}
$$

$$V_{1,S} = \frac{1}{S}\left(0+\sum_{s=1}^{S}\tfrac{s}{S}\left(1-\tfrac{s}{S}\right)\right) = \frac{1}{6}\left(1-S^{-2}\right) \to \frac{1}{6}.$$

then for $h \geq 1$ we have

$$\begin{aligned}\text{Var}\left\{\tilde{\gamma}_h(Y_\delta;S)\right\} &= \text{Var}\left\{\sum_{s=-S}^{S}\tfrac{S-|s|}{S}\tilde{\gamma}_{Sh+s}(Y_{\delta_S})\right\} = \sum_{s=-S}^{S}\left(1-\tfrac{|s|}{S}\right)^2 \times \frac{1}{n}4t\int_0^t \sigma_u^4 \mathrm{d}u \\ &= 4V_{0,S} \times \frac{1}{n_\delta}t\int_0^t \sigma_u^4 \mathrm{d}u,\end{aligned}$$

and similarly for $h=0$ we find $Var\left\{\tilde{\gamma}_0(Y_\delta;S)\right\} = 2V_{0,S} \times \frac{1}{n_\delta}t\int_0^t \sigma_u^4 du$.

For $h \geq 0$ we find

$$\begin{aligned}\text{Cov}\left\{\tilde{\gamma}_h(Y_\delta;S),\tilde{\gamma}_{h+1}(Y_\delta;S)\right\} &= \text{Cov}\left\{\sum_{s=-S}^{S}\tfrac{S-|s|}{S}\tilde{\gamma}_{Sh+s}(Y_{\delta_S}),\sum_{s=-S}^{S}\tfrac{S-|s|}{S}\tilde{\gamma}_{Sh+S+s}(Y_{\delta_S})\right\} \\ &= \text{Var}\left\{\sum_{s=1}^{S}\tfrac{S-s}{S}\tfrac{s}{S}\tilde{\gamma}_{Sh+s}(Y_{\delta_S})\right\} = \sum_{s=1}^{S}\tfrac{S-s}{S}\tfrac{s}{S} \times \frac{1}{n}4t\int_0^t \sigma_u^4 \mathrm{d}u \\ &= 4V_{1,S} \times \frac{1}{n_\delta}t\int_0^t \sigma_u^4 \mathrm{d}u.\end{aligned}$$

Covariances between $\tilde{\gamma}_h(Y_\delta;S)$ and $\tilde{\gamma}_i(Y_\delta;S)$ are zero for $|h-i| \geq 2$, as they do not "share" any of the realised autocovariances $\tilde{\gamma}_{Sh+s}(Y_{\delta_S})$.

□

Proof of Theorem 3. The limit results for the subsampled realised kernels involving U_δ follow directly from Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) and (7)-(9). So we only need to show the result for the subsampled realised kernel on Y_δ.

We have

$$S\left(V_{0,S}+2V_{1,S}\right) = 1+2\sum_{s=1}^{S}\left(1-\tfrac{s}{S}\right)^2+2\sum_{s=1}^{S}\tfrac{s}{S}\left(1-\tfrac{s}{S}\right)1+\frac{S(S-1)}{S} = S,$$

so that $V_{0,S}+2V_{1,S}=1$. Given the structure of

$$A_S = \begin{pmatrix} 2V_{0,S} & \bullet & 0 & \cdots \\ 4V_{1,S} & 4V_{0,S} & \bullet & \ddots \\ 0 & 4V_{1,S} & 4V_{1,S} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

we have

$$\begin{aligned}& H^{-1}\sum_{i,j=0}^{H}k(\tfrac{i}{H})k(\tfrac{j}{H})[A_S]_{i,j} \\ =\ & 4V_{0,S}H^{-1}\sum_{h=0}^{H}k(\tfrac{h}{H})^2+8V_{1,S}H^{-1}\sum_{h=1}^{H}k(\tfrac{h}{H})k(\tfrac{h-1}{H})+O(H^{-1}) \\ =\ & 4(V_{0,S}+2V_{1,S})H^{-1}\sum_{h=0}^{H}k(\tfrac{h}{H})^2-8V_{1,S}H^{-2}\sum_{h=1}^{H}k(\tfrac{h}{H})\tfrac{k(\frac{h}{H})-k(\frac{h-1}{H})}{1/H}+O(H^{-1})\end{aligned}$$

$$= \quad 4\int_0^1 k(u)^2 \mathrm{d}x + O(H^{-1}),$$

which proves the result.
□

Proof of Theorem 4. (*i*) The mixed Gaussian result follows from Theorem 3. (*ii*) The best value for c is found by solving the first order condition

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

and substituting this c into (19) yields $\omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}$ times

$$4c\left\{k_\bullet^{0,0} + \frac{k'(0)^2 + k'(1)^2}{c^3}\right\} = 4c\left(k_\bullet^{0,0} + \frac{1}{2}k_\bullet^{0,0}\right) = 4ck_\bullet^{0,0}\left(1 + 1/2\right) = 6ck_\bullet^{0,0}.$$

Finally

$$ck_\bullet^{0,0} = \left\{2\frac{k'(0)^2 + k'(1)^2}{k_\bullet^{0,0}}\right\}^{1/3} k_\bullet^{0,0} = \left\{2\left(k_\bullet^{0,0}\right)^2\left(k'(0)^2 + k'(1)^2\right)\right\}^{1/3},$$

completes the proof.
□

Proof of Theorem 5. (*i.a*) The mixed Gaussian result is straight forward using Theorem 3.

(*i.b*) Substituting $HS = \xi^{1/2}cn^{1/2+\alpha/4}$ and $S = an^\alpha$ into (20) yields $4\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4}$ times

$$\frac{cn^{1/2+\alpha/4}}{n}k_\bullet^{0,0} + \frac{2\rho k_\bullet^{1,1}}{cn^{1/2+\alpha/4}} + nn^\alpha\frac{k_\bullet^{2,2}}{(cn^{1/2+\alpha/4})^3} = ck_\bullet^{0,0}n^{-1/2+\alpha/4} + c^{-3}k_\bullet^{2,2}n^{-1/2+\alpha/4},$$

because the second term, $c^{-1}2\rho k_\bullet^{1,1}n^{-1/2-\alpha/4}$, is of lower order that the first and the third term when $\alpha > 0$.

(*ii*) Minimizing (20) with respect to $x = HS$ has the first order condition,

$$n^{-1}k_\bullet^{0,0} - 2\xi\rho k_\bullet^{1,1}(HS)^{-2} - 3\xi^2 nSk_\bullet^{2,2}(HS)^{-4} = 0.$$

The unique positive solution is given by $HS = c_S(\xi n)^{1/2}$, where

$$c_S = \sqrt{\frac{\rho k_\bullet^{1,1}}{k_\bullet^{0,0}}\left(1 + \sqrt{1 + 3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}\right)} = \sqrt{\frac{\rho k_\bullet^{1,1}}{k_\bullet^{0,0}} + \sqrt{\frac{(\rho k_\bullet^{1,1})^2 + 3Sk_\bullet^{0,0}k_\bullet^{2,2}}{(k_\bullet^{0,0})^2}}}.$$

Now define

$$x = \frac{k_\bullet^{0,0}}{\rho k_\bullet^{1,1}} \qquad y = \frac{\rho k_\bullet^{1,1}}{Sk_\bullet^{2,2}}, \qquad \text{and} \qquad z = \sqrt{1 + 3x/y}$$

Then

$$c_S = \sqrt{\frac{1 + \sqrt{1 + 3x/y}}{x}} = \sqrt{\frac{1+z}{x}},$$

and $x/y = (z^2 - 1)/3 = (1 + z)(z - 1)/3$. So the minimum asymptotic variance is given by

$$4\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{3/4} k_\bullet^{0,0}\left(c_S + \frac{2}{c_S x} + \frac{1}{c_S^3 xy}\right),$$

which is proportional to

$$c_S + \frac{2}{c_S x} + \frac{1}{c_S^3 xy} \quad = \quad \sqrt{\tfrac{1+z}{x}} + 2\sqrt{\tfrac{1}{x(1+z)}} + \tfrac{\sqrt{x}}{(1+z)\sqrt{1+z}c^3 y} = \tfrac{1}{\sqrt{x}}\left\{\sqrt{1+z} + \tfrac{2}{\sqrt{1+z}} + \tfrac{x/y}{(1+z)\sqrt{1+z}}\right\}$$

$$
\begin{aligned}
&= \tfrac{1}{\sqrt{x}}\left\{\sqrt{1+z}+\tfrac{2}{\sqrt{1+z}}+\tfrac{(1+z)(z-1)/3}{(1+z)\sqrt{1+z}}\right\} = \tfrac{1}{\sqrt{x}}\left\{\tfrac{(1+z)+2+(z-1)/3}{\sqrt{1+z}}\right\} = \tfrac{1}{\sqrt{x}}\tfrac{4}{3}\tfrac{2+z}{\sqrt{1+z}} \\
&= \tfrac{1}{\sqrt{x}}\tfrac{4}{3}\left(\tfrac{1}{\sqrt{1+z}}+\sqrt{1+z}\right).
\end{aligned}
$$

Now substitute $z = \sqrt{1+3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}$ and $\frac{1}{\sqrt{x}} = \sqrt{\frac{\rho k_\bullet^{1,1}}{k_\bullet^{0,0}}}$ and (22) follows.
□

Lemma A.2 *Let $g(S)$ be as defined in Theorem 5. Then $g'(S) > 0$ for all $S > 0$.*

Proof. Consider the function

$$
f(x) = \frac{1}{\sqrt{1+\sqrt{1+ax}}} + \sqrt{1+\sqrt{1+ax}}, \qquad \text{for } a > 0.
$$

The first derivative $f'(x) = \frac{a}{4}\left(1+\sqrt{ax+1}\right)^{-3/2}$, is positive for all $x > 0$.
□
Proof of Corollary 1. From Lemma A.2 it follows that $g'(S) > 0$ for all $S > 0$, if we set $x = S$ and $a = 3k_\bullet^{0,0}k_\bullet^{2,2}/(\rho k_\bullet^{1,1})^2$. So any increment in S will increase the asymptotic variance.
□
Proof of Corollary 2. By substitution for the first ρ in $g(S)$ we find that (22) is proportional to

$$
\omega\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{1/2}\left(\int_0^t \sigma_u^2 \mathrm{d}u\right)^{1/2}\left\{\frac{1}{\sqrt{1+\sqrt{1+3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}}} + \sqrt{1+\sqrt{1+3S\frac{k_\bullet^{0,0}k_\bullet^{2,2}}{(\rho k_\bullet^{1,1})^2}}}\right\}.
$$

From Hansen and Lunde (2006, p. 135) it follows that business time sampling minimizes $t\int_0^t \sigma_u^4 du$ and by Lemma A.2 we have that also the second term is minimized for the largest possible value of ρ, (set $x = 1/\rho^2$). Since $\rho \leq 1$ the solution is $\rho = 1$.
□
Proof of Lemma 1. For simplicity, let $U_0 = U_n = 0$. We have

$$
\begin{aligned}
\tilde{\gamma}_h(U) &= \sum_{j=1}^{n}(U_j - U_{j-1})\left(U_{j-h} - U_{j-h-1} + U_{j+h} - U_{j+h-1}\right) \\
&= \sum_{j=0}^{n} U_j\left(U_{j-h} - U_{j-h-1} + U_{j+h} - U_{j+h-1}\right) - \sum_{j=0}^{n} U_j\left(U_{j-h+1} - U_{j-h} + U_{j+h+1} - U_{j+h}\right) \\
&= \sum_{j=0}^{n} U_jU_{j-h} - U_jU_{j-h-1} + U_jU_{j+h} - U_jU_{j+h-1} - U_jU_{j-h+1} + U_jU_{j-h} \\
&\qquad - U_jU_{j+h+1} + U_jU_{j+h} \\
&= \sum_{j=1}^{n} -U_j(U_{j-h-1} + U_{j+h+1}) + 2U_j\left(U_{j-h} + U_{j+h}\right) - U_j(U_{j+h-1} + U_{j-h+1}) \\
&= -V_{h+1,n} + 2V_{h,n} - V_{h-1,n}
\end{aligned}
$$

where $V_h = \sum_{j=1}^{n} U_j\left(U_{j-h} + U_{j+h}\right)$. So the realised kernel on the pure noise process, U, is

$$
\tilde{K}_w(U_\delta) \quad = \quad w_0(V_{0,n} - V_{1,n}) + \sum_{h=1}^{H} w_h(-V_{h-1,n} + 2V_{h,n} - V_{h+1,n})
$$

$$= \sum_{h=0}^{H} \left(-w_{h+1} + 2w_h - w_{h-1}\right) V_{h,n},$$

using the conventions $w_{-1} = w_0$ and $w_{H+1} = w_H$.

We have $\mathrm{Var}(V_h) = (4n - 2h)\omega^4$, and because V_h is entirely made up of $U_j U_{j-h}$ terms it follows that $\mathrm{Cov}(V_h, V_k) = 0$, for $h \neq k$. So

$$\begin{aligned}\mathrm{Var}\left\{\tilde{K}_w(U)\right\} &= \sum_{h=0}^{H} \left(w_{h+1} - 2w_h + w_{h-1}\right)^2 \left(4n\omega^4 - 2\omega^4 h\right) \\ &\geq 4\omega^4 (n - H/2) \sum_{h=0}^{H} \left(w_{h+1} - 2w_h + w_{h-1}\right)^2.\end{aligned}$$

Since $H = o(n)$ the result follows. The expressions are more involved without the simplifying assumption $U_0 = U_n = 0$. Here the conclusion is the same because the variance is also $\simeq 4\omega^4 n \sum_{h=0}^{H} \left(w_{h+1} - 2w_h + w_{h-1}\right)^2$ in this case. □

Proof of Theorem 6. The asymptotic distribution of

$$\gamma_0(X_\delta; S) + \widetilde{\gamma}_1(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u$$

is mixed Gaussian with variance of approximately, for moderate n_δ and S,

$$n_\delta^{-1} \frac{16}{3} t \int_0^t \sigma_u^4 \mathrm{d}u + \frac{8\omega^4 n_\delta}{S}. \tag{A.1}$$

The first term appears from (12), the second from Theorem 2 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006). □

Proof of Lemma 2. With $S = c(\xi n)^{2/3}$ we have $n_\delta^{-1} = S/n = c\xi^{2/3} n^{-1/3}$ and $\frac{n_\delta}{S} = n/S^2 = c^{-2}\xi^{-4/3} n^{1/3}$, so that (A.1) in the proof of Theorem 6 becomes $n^{1/3}$ times

$$\frac{16}{3} c\xi^{2/3} t \int_0^t \sigma_u^4 \mathrm{d}u + 8\omega^4 c^{-2} \xi^{-4/3} = \omega^{4/3} \left(t \int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3} \left(\frac{16}{3} c + 8c^{-2}\right).$$

So $n^{1/6} \left\{\gamma_0(X_\delta; S) + \widetilde{\gamma}_1(X_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u\right\}$ converges to a mixed Gaussian distribution with this variance. We can now minimise this asymptotic variance by selecting

$$c^3 = 3.$$

At this value the asymptotic variance is

$$\omega^{4/3} \left(t \int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3} \left\{\frac{16}{3} (3)^{1/3} + 8 (3)^{-2/3}\right\} \simeq 11.53 \omega^{4/3} \left(t \int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}.$$

□

Proof of Lemma 3. From Theorems 2 and 3 we obtain the following upper-left 3×3 submatrices of A_∞ and C,

$$[A_{\infty,3\times 3}] = \frac{2}{3} \begin{pmatrix} 2 & \bullet & \bullet \\ 1 & 4 & \bullet \\ 0 & 1 & 4 \end{pmatrix}, \qquad [C_{3\times 3}] = \begin{pmatrix} \lambda^2 + 1 & \bullet & \bullet \\ -\lambda^2 - 2 & \lambda^2 + 5 & \bullet \\ 1 & -4 & 6 \end{pmatrix}.$$

With $w = (1, 1, \frac{1}{2})^\intercal$ we have $w^\intercal[A_{\infty,3\times 3}]w = \frac{20}{3}$ and $w^\intercal[C_{3\times 3}]w = \frac{1}{2}$. The result now follows, as the asymptotic variance is

$$\begin{aligned} n^{1/3}\left(\tfrac{S}{n}t\int_0^1 \sigma_u^4 \mathrm{d}u \times \frac{20}{3} + 4\omega^4 \frac{n}{S^2} \times \frac{1}{2}\right) &= c\xi^{2/3} t \int_0^1 \sigma_u^4 \mathrm{d}u \frac{20}{3} + 2\omega^4 c^{-2}\xi^{-4/3} \\ &= \omega^{4/3}\left(t\int_0^1 \sigma_u^4 \mathrm{d}u\right)^{2/3}\left(\frac{20}{3}c + 2c^{-2}\right). \end{aligned}$$

□

Proof of Theorem 7. We have that $\gamma_0(X_\delta; S) - 2n_\delta\omega^2 - \int_0^t \sigma_u^2 du$ equals

$$\underbrace{\gamma_0(Y_\delta; S) - \int_0^t \sigma_u^2 \mathrm{d}u}_{n_\delta^{-1}\frac{4}{3}t\int_0^t \sigma_u^4 \mathrm{d}u} + \underbrace{2\gamma_0(U_\delta, Y_\delta; S)}_{S^{-1}8\omega^2\int_0^t \sigma_u^2 \mathrm{d}u} + \underbrace{\gamma_0(U_\delta; S) - 2n_\delta\omega^2}_{4\omega^4\frac{n_\delta}{S}(1+\lambda^2)},$$

which has mean zero and a variance that is the sum of the three terms given below the brackets. The three terms are given from (12), (7), and (9) respectively. For large $S = c(\xi n)^{2/3}$ (implying large $n_\delta = n/S = c^{-1}\xi^{-2/3}n^{1/3}$) we have

$$n^{1/6}\left\{\gamma_0(X_\delta; S) - 2n_\delta\omega^2 - \int_0^t \sigma_u^2 \mathrm{d}u\right\} \xrightarrow{Ls} MN\left\{0, 4\omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}\left(\tfrac{c}{3} + \tfrac{1+\lambda^2}{c^2}\right)\right\}.$$

□

Proof of Theorem 8. By the approximations

$$\frac{1}{S}\sum_{j=1}^n \left(U_{j\delta/S} - U_{(j-S)\delta/S}\right)^2 \simeq \frac{2}{S}\left(\sum_{j=1}^n U_{j\delta/S}^2 + \sum_{j=1}^n U_{j\delta/S}U_{(j-S)\delta/S}\right)$$

and

$$\frac{1}{S}\sum_{j=1}^n \left(U_{j\delta/S} - U_{(j-1)\delta/S}\right)^2 \simeq \frac{2}{S}\left(\sum_{j=1}^n U_{j\delta/S}^2 + \sum_{j=1}^n U_{j\delta/S}U_{(j-1)\delta/S}\right),$$

and using

$$\frac{2}{S} = \frac{2n^{1/2}}{c\xi^{2/3}n^{2/3}} \times n^{-1/2} = n^{-1/6}\sqrt{\frac{4}{c^2\xi^{4/3}}} \times n^{-1/2}$$

we see that

$$n^{-1/6}\begin{pmatrix} n^{-1/2}\sum_{j=1}^n \left(U_{j\delta/S} - U_{(j-S)\delta/S}\right)^2 - 2n_\delta\omega^2 \\ n^{-1/2}\sum_{j=1}^n \left(U_{j\delta/S} - U_{(j-1)\delta/S}\right)^2 - 2n_\delta\omega^2 \end{pmatrix} \xrightarrow{L} N\left\{0, \frac{4\omega^4}{c^2\xi^{4/3}}\begin{pmatrix} 1+\lambda^2 & \lambda^2 \\ \lambda^2 & 1+\lambda^2 \end{pmatrix}\right\}.$$

□

Proof of Theorem 9. Follows from Theorem 7 and Theorem 8, and

$$n^{-1/2}\gamma_0(X_{\delta/S}) = n^{-1/2}\gamma_0(U_{\delta/S}) + o_p(1) \quad \text{and} \quad \omega^4/\xi^{4/3} = \omega^{4/3}\left(t\int_0^t \sigma_u^4 \mathrm{d}u\right)^{2/3}.$$

□