Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading*

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
& CREATES, University of Aarhus
oebn@imf.au.dk

PETER REINHARD HANSEN
Department of Economics, Stanford University, Landau Economics Building,
579 Serra Mall, Stanford, CA 94305-6072, USA
peter.hansen@stanford.edu

ASGER LUNDE
Department of Marketing and Statistics, Aarhus School of Business,
University of Aarhus, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark
& CREATES, University of Aarhus
alunde@asb.dk

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

July 1, 2008

Abstract

We propose a multivariate realised kernel to estimate the ex-post covariation of log-prices. We show this new consistent estimator is guaranteed to be positive semi-definite and is robust to measurement noise of certain types and can also handle non-synchronous trading. It is the first estimator which has these three properties which are all essential for empirical work in this area. We derive the large sample asymptotics of this estimator and assess its accuracy using a Monte Carlo study. We implement the estimator on some US equity data, comparing our results to previous work which has used returns measured over 5 or 10 minutes intervals. We show the new estimator is substantially more precise.

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

*The second and fourth author are also affiliated with CREATES, a research center funded by the Danish National Research Foundation. The Ox language of Doornik (2001) was used to perform the calculations reported here. We thank Ron Gallant for comments on a previous version of this paper.

1 Introduction

The last seven years has seen dramatic improvements in the way econometricians think about time-varying financial volatility, first brought about by harnessing high frequency data and then by mitigating the influence of market microstructure effects. Extending this work to the multivariate case is challenging as this needs to additionally remove the effects of non-synchronous trading while simultaneously requiring that the covariance matrix estimator be guaranteed to be positive semi-definite. In this paper we provide the first estimator which achieves all these objectives. This will be called the multivariate realised kernel, which we will define in equation (1).

We study a d-dimensional log price process $X = \left(X^{(1)}, X^{(2)}, ..., X^{(d)}\right)'$. These prices are observed irregularly and non-synchronous over the interval $[0, T]$. For simplicity of exposition we take $T = 1$ throughout the paper. These observations could be trades or quote updates. The observation times for the i-th asset will be written as $t_1^{(i)}, t_2^{(i)}, \ldots$. This means the available database of prices is $X^{(i)}(t_j^{(i)})$, for $j = 1, 2, ..., N^{(i)}(1)$, and $i = 1, 2, ..., d$. Here $N^{(i)}(t)$ counts the number of distinct data points available for the i-th asset up to time t.

X is assumed to be driven by Y, the efficient price, abstracting from market microstructure effects. The efficient price is modelled as a *Brownian semimartingale* ($Y \in \mathcal{BSM}$) defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$,

$$Y(t) = \int_0^t a(u)\mathrm{d}u + \int_0^t \sigma(u)\mathrm{d}W(u),$$

where a is a vector of elements which are predictable locally bounded drifts, σ is a càdlàg volatility matrix process and W is a vector of independent Brownian motions. For reviews of the econometrics of this type of process see, for example, Ghysels, Harvey & Renault (1996). If $Y \in \mathcal{BSM}$ then its ex-post covariation, which we will focus on for reasons explained in a moment, is

$$[Y](1) = \int_0^1 \Sigma(u)\mathrm{d}u, \quad \text{where} \quad \Sigma = \sigma\sigma',$$

where

$$[Y](1) = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \left\{Y(t_j) - Y(t_{j-1})\right\} \left\{Y(t_j) - Y(t_{j-1})\right\}',$$

(e.g. Protter (2004, p. 66–77) and Jacod & Shiryaev (2003, p. 51)) for any sequence of deterministic synchronized partitions $0 = t_0 < t_1 < ... < t_n = 1$ with $\sup_j\{t_{j+1} - t_j\} \to 0$ for $n \to \infty$. This is the quadratic variation of Y.

The contribution of this paper is to construct a consistent, positive semi-definite estimator of $[Y](1)$ from our database of asset prices. The challenges of doing this are three fold: (i) there are market microstructure effects $U = X - Y$, (ii) the data is irregularly spaced and non-synchronous, (iii) the market microstructure effects are not statistically independent of the Y process.

Quadratic variation is crucial to the economics of financial risk. This is reviewed by, for example, Andersen, Bollerslev & Diebold (2008) and Barndorff-Nielsen & Shephard (2007), who provide very extensive references. The economic importance of this line of research has recently been reinforced by the insight of Bollerslev, Tauchen & Zhou (2008) who have showed that expected stock returns seem well explained by the variance risk premium (the difference between the implied and realised variance) and this risk premium is only detectable using the power of high frequency data. See also the paper by Drechsler & Yaron (2008).

Our analysis builds upon earlier work on the effect of noise on univariate estimators of $[Y](1)$ by, amongst others, Zhou (1996), Andersen, Bollerslev, Diebold & Labys (2000), Bandi & Russell (2005), Zhang, Mykland & Aït-Sahalia (2005), Andersen, Bollerslev & Meddahi (2006), Hansen & Lunde (2006), Kalnina & Linton (2008), Zhang (2006), Renault & Werker (2008), Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*a*) and Jacod, Li, Mykland, Podolskij & Vetter (2007). The case of no noise is dealt with in the same spirit as the papers by Andersen, Bollerslev, Diebold & Labys (2001) and Barndorff-Nielsen & Shephard (2002), Barndorff-Nielsen & Shephard (2004), Mykland & Zhang (2006), Mykland & Zhang (2008) and Jacod & Protter (1998).

A distinctive feature of multivariate financial data is the phenomenon of non-synchronous trading or nontrading. These two terms are distinct. The first refers to the fact that any two assets rarely trade at the same instant. The latter to situations where one assets is trading frequently over a period while some other assets do not trade. The treatment of non-synchronous trading effects dates back to Fisher (1966). For several years researchers focused mainly on the effects that stale quotes have on daily closing prices. Campbell, Lo & MacKinlay (1997, chapter 3) provides a survey of this literature. When increasing the sampling frequency beyond the inter-hour level several authors have demonstrated a severe bias towards zero in covariation statistics. This phenomenon is often referred to as the Epps effect. Epps (1979) found this bias for stock returns, and it has also been demonstrated to hold for foreign exchange returns, see Guillaume, Dacorogna, Dave, Müller, Olsen & Pictet (1997). This is confirmed in our empirical work where realised covariances computed using high frequency data, over specified fixed time periods such as 15 seconds, dramatically underestimate the degree of dependence between assets. Some recent econometric work on this topic includes Malliavin & Mancino (2002), Reno (2003), Martens (2003), Hayashi & Yoshida (2005), Voev & Lunde (2007), Griffin & Oomen (2006), Large (2007) and Zhang (2005). We will draw ideas from this work.

The form of multivariate realised kernel we propose is, in the univariate special case, subtly different from that studied in the univariate paper by Barndorff-Nielsen et al. (2008*a*). Their flat-top kernel, which has the advantage of being unbiased and fully efficient, is not guaranteed to be non-negative. It also could not directly deal with non-synchronous data. This is essential in the multivariate case, which motivates the specific form of the multivariate realised kernel proposed here. We discuss in some detail the specifics

of the differences between these estimates in Section 6.1. The change to our preferred estimator means the rate of convergence, bandwidth choice and asymptotic distribution of our new estimator differs from the flat-top version. In particular, our theory can be used to tune the bandwidth selection for estimating particular correlations, betas, inverse covariance matrices or just covariances.

The structure of the paper is as follows. In Section 2 we synchronize the timing of the multivariate data using what we call Refresh Time. This allows us to refine high frequency returns and in turn the multivariate realised kernel. Further we make precise the assumptions we make use of in our Theorems to study the behaviour of our statistics. In Section 3 we give a detailed discussion of the asymptotic distribution of realised kernels in the univariate case. The analysis is then extended to the multivariate case. Section 4 contains a summary of a simulation experiment designed to investigate the finite sample properties of our estimator. Section 5 contains some results from implementing our estimators on some US stock price data taken from the TAQ database. This is followed by a Section on extensions and further remarks, while the main part of the paper is finished by a Conclusion. This is followed by an Appendix which contains the proofs of various theorems given in the paper, and an Appendix with results related to Refresh Time sampling. More details of our empirical results and simulation experiments are given in a web Appendix which can be found at `http://www.hha.dk/~alunde/BNHLS/BNHLS.htm`.

2 Defining the multivariate realised kernel

2.1 Synchronizing data

2.1.1 Refresh time

Non-synchronous trading delivers fresh (trade or quote) prices at irregularly spaced times which differ across stocks. Dealing with non-synchronous trading has been an active area of research in financial econometrics in recent years, e.g. Hayashi & Yoshida (2005), Voev & Lunde (2007) and Large (2007). Stale prices are a key feature of estimating covariances in financial econometrics as recognised at least since Epps (1979), for they introduce cross-autocorrelation amongst asset price returns.

Write the number of observations in the i-th asset made up to time t as the counting process $N^{(i)}(t)$, and the times at which trades are made as $t_1^{(i)}, t_2^{(i)}, ...$ We now define a time scale which will be key to the construction of multivariate realised kernels.

Definition 1 *Refresh Time*[1] *for* $t \in [0,1]$. *We define the first refresh time as* $\tau_1 = \max\left(t_1^{(1)}, ..., t_1^{(d)}\right)$, *and then subsequent refresh times as*

$$\tau_{j+1} = \max\left(t^{(1)}_{N^{(1)}_{\tau_j}+1}, ..., t^{(d)}_{N^{(d)}_{\tau_j}+1}\right).$$

The resulting Refresh Time sample size is N, *while we write* $n^{(i)} = N^{(i)}(1)$.

[1] Refresh time was used in a cointegration study of price discovery by Harris, McInish, Shoesmith & Wood (1995). Martens (2003) used the same idea in the context of realised covariances, but his estimator is inconsistent.

The τ_1 is the time it has taken for all the assets to trade, i.e. all their posted price have been updated. τ_2 is the first time when all the prices are again refreshed. This process is displayed in Figure 1 for $d = 3$.

Our analysis will now be based on this time clock $\{\tau_j\}$. Our approach will be to:

- Assume the entire vector of up to date prices are seen at these refreshed times $X(\tau_j)$, which is not correct — for we only see a single new price and $d-1$ stale prices[2].
- Show these stale pricing errors have no impact on the asymptotic distribution of the realised kernels.

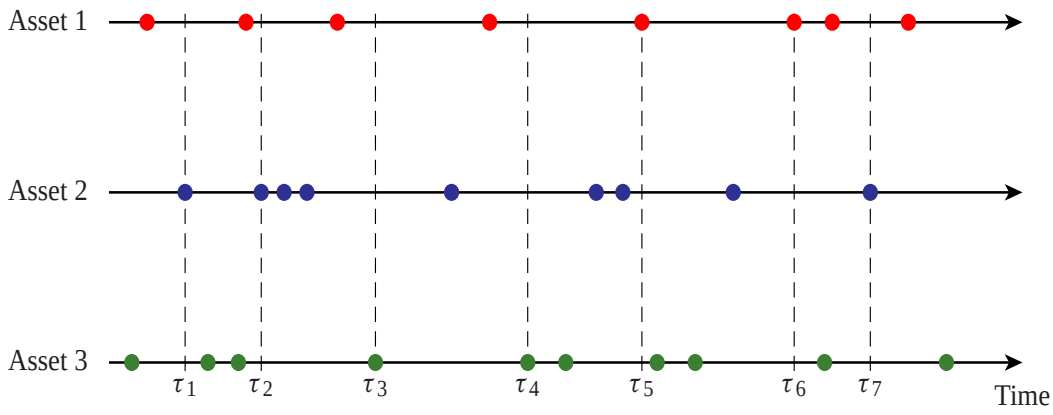

Figure 1: *This figure illustrates Refresh Time in a situation with three assets. The dots represent the times* $\{t_j^{(i)}\}$. *The vertical lines represent the sampling times generated from the three assets with refresh time sampling. Note, in this example,* $N = 7$, $n^{(1)} = 8$, $n^{(2)} = 9$ and $n^{(3)} = 10$.

This approach to dealing with non-synchronous data converts the problem into one where the Refreshed Times' sample size N is determined by the degree of non-synchronicity and $n^{(1)}, n^{(2)}, \ldots, n^{(d)}$.[3] The degree to which we keep data is measured by the size of the retained data over the original size of the database. For Refresh Time this is $p = dN/\sum_{i=1}^{d} n^{(i)}$. For the data in Figure 1, $p = 21/27 \simeq 0.78$.

2.1.2 Jittering end conditions

Realised kernels are built out of n high frequency returns computed from synchronized vector prices recorded at N times. It turns out that our asymptotic theory dictates we need to average m prices at the

[2]Their degree of staleness will be limited by their Refresh Time construction to a single lag in Refresh Time. The extension to a finite number of lags is given in Section 6.6.

[3]Suppose trade times arrive as independent standard Poisson process with common intensity λ, so that $\mathrm{E}\{N^{(i)}(t)\} = \lambda t$. Then $t_1^{(i)} \sim \exp(\lambda)$ and as $\tau_1 = \max\{t_1^{(1)}, t_1^{(2)}, ..., t_1^{(d)}\}$, so, e.g. Embrechts, Kluppelberg & Mikosch (1997, pp. 125 & 176) $\tau_1/\log d \xrightarrow{d} \lambda^{-1}$, or more refined $\Pr(\tau_1 - \log d \leq x) = \{1 - \exp(-\lambda x)/d\}^d \to \exp\{-\exp(-\lambda x)\}$. Hence the sample size from the refreshed analysis will fall with $\log d$, the dimension of the asset prices. The situation where the intensity varies across assets, i.e. $\mathrm{E}\{N^{(i)}(t)\} = \lambda_i t$, will not substantially change this result.

The loss of observations is relatively cheap here, because the rate of convergence for our realised kernel will be $n^{1/5}$. In a standard situation where an estimator converges at rate $n^{1/2}$, one can expect confidence intervals to widen by about 100% when the sample size is reduced by a factor of 4. When the rate of convergence is $n^{1/5}$, confidence intervals only widen by about 32%.

very beginning and end of the day to define these returns[4]. The theory behind this will be explained in Section 6.5, but for now we just define what we mean by jittering. Let $n, m \in \mathbb{N}$, with $n - 1 + 2m = N$, then set the vector observations $X_0, X_1, ..., X_n$ as $X_j = X(\tau_{j+m})$, $j = 1, 2, ..., n-1$,and

$$X_0 = \frac{1}{m}\sum_{j=1}^{m} X\left(\tau_j\right) \quad \text{and} \quad X_n = \frac{1}{m}\sum_{j=1}^{m} X(\tau_{N-m+j}).$$

So X_0 and X_n are constructed by jittering initial and final time points. By allowing m to be moderately large but very small in comparison with n, it means these observations record the efficient price without much error, as the error is averaged away. Experimentation suggests m should be around 2 for the kind of data we see in this paper: see Section 6.5 for a discussion of this issue.

These prices allow us to define the high frequency vector returns as $x_j = X_j - X_{j-1}$, $j = 1, 2, ..., n$.

2.2 Realised kernel

Having synchronized the high frequency vector returns $\left\{x_j\right\}$ we can define our class of positive semi-definite *multivariate realised kernels* (RK). It takes on the following form

$$K(X) = \sum_{h=-n}^{n} k(\tfrac{h}{H+1})\Gamma_h. \tag{1}$$

Here the non-stochastic $k(x)$ for $x \in \mathbb{R}$ is a weight function. The h-th realised autocovariance is

$$\Gamma_h = \begin{cases} \sum_{j=|h|+1}^{n} x_j x_{j-h}', & h \geq 0 \\ \sum_{j=|h|+1}^{n} x_{j-h} x_j', & h < 0. \end{cases}$$

We focus on the class of kernel functions, $\mathcal{K}$, that is characterized by:

(i) $k(0) = 1, k'(0) = 0$;

(ii) k is twice differentiable with continuous derivatives;

(iii) $k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2} < \infty$, where $k_\bullet^{0,0} = \int_0^1 k(x)^2 \mathrm{d}x$, $k_\bullet^{1,1} = \int_0^1 k'(x)^2 \mathrm{d}x$, $k_\bullet^{2,2} = \int_0^1 k''(x)^2 \mathrm{d}x$;

(iv) $\int_{-\infty}^{\infty} k(x)\exp(ix\lambda)\mathrm{d}x \geq 0$ for all $\lambda \in \mathbb{R}$.

The assumption $k(0) = 1$ means Γ_0 gets unit weight, while $k'(0) = 0$ means the kernel gives close to unit weight to Γ_h for small values of $|h|$. Condition (iv) guarantees $K(X)$ to be positive semi-definite, (e.g. Bochner's theorem and Andrews (1991)).

The multivariate realised kernel has the same form as a standard heteroskedasticity and autocorrelated (HAC) covariance matrix estimator familiar in econometrics (e.g. Gallant (1987), Newey & West (1987), and Andrews (1991)). But there are a number of important differences. For example, the sums that

[4]This kind of averaging appears in, for example, Jacod et al. (2007).

define the realised autocovariances are not divided by the sample size, while $k'(0) = 0$ is critical in our framework. Unlike the situation in the standard HAC literature, an estimator based on the Bartlett kernel will not be consistent for the ex-post variation of prices, measured by quadratic variation, in the present setting. Later we will recommend using the Parzen kernel (its form is given in Table 1) instead.

2.3 Assumptions about the noise and refresh time

Having defined the positive semi-definite realised kernel, we will now write out our assumptions about the market microstructure effects U and the $\{\tau_j\}$ which govern the properties of the vector returns $\{x_j\}$ and so $K(X)$.

The assumptions about the noise are stated in observations time — that is we only model the noise at exactly the times where there are trades or quote updates. This type of assumption is familiar from the work of, for example, Zhou (1998), Bandi & Russell (2005), Zhang et al. (2005), Barndorff-Nielsen et al. (2008*a*) and Hansen & Lunde (2006). We define

$$U_j = X(\tau_j) - Y(\tau_j), \quad j = 0, 1, ..., N,$$

which is noise associated with $X(\tau_j)$, the observation at time τ_j.

Assumption 1 *Suppose that, conditional on* $\{Y\}$, $\{U_j\}$ *is covariance stationary (*$U \in \mathcal{CS}$*) with* $\mathrm{E}(U_j) = 0$ *and* $\sum_h |h\,\Omega_h| < \infty$*, where* $\Omega_h = \mathrm{cov}(U_j, U_{j-h})$*. Let* $M = \max\{|i-j|, |h-l|\}$*. For* $h, l \geq 0$*, there exists* ϱ_M *such that,* $\left|\mathrm{Cov}(U_i U'_{i-h}, U_j U'_{j-l})\right| \leq \varrho_M$*, where* $\sum_{i=1}^{\infty} \varrho_i (1+\epsilon)^i < \infty$*, for some* $\epsilon > 0$*.*

A key quantity in our analysis is the, so-called, long-run variance:

$$\Omega = \sum_{h=-\infty}^{\infty} \Omega_h,$$

which is a non-stochastic $d \times d$ matrix.

On occasions we refer to a white noise assumption about the U process ($U \in \mathcal{WN}$) which means we assume it has $\mathrm{E}\left(U_j\right) = 0$, $\mathrm{Var}\left(U_j\right) = \Omega$ and $U_i \perp\!\!\!\perp U_j$, for all $i \neq j$. This white noise assumption is unsatisfactory from a number of viewpoints (e.g. Phillips & Yu (2008) and Kalnina & Linton (2008)) and will not be used to derive our limit theorems.

Throughout the paper we follow Barndorff-Nielsen et al. (2008*a*) in making this assumption about the times that we have Refresh Time data[5].

[5]This means that $Z(t) = Y(T(t))$ is a Brownian semimartingale with $[Z]_1 = [Y]_{T(1)}$ and spot volatility $\lambda(t) = \tau(t)\{\sigma \circ T(t)\}$. The point of this assumption is that $Z(j/N) = Y(T(j/N)) = Y(\tau_j)$, where $\tau_j = T(j/N)$. So irregularly spaced data on Y can be thought of as equally spaced on Z.

An implication of this is that $\sup_j\{\tau_{j+1} - \tau_j\} = O_p(n^{-1})$ for $n \to \infty$, which means that $\sup_{i,j}\{t_{j+1}^{(i)} - t_j^{(i)}\} = O_p(n^{-1})$ by construction of refresh time.

Assumption 2 *Define* $T(t) = \int_0^t \tau^2(u)\mathrm{d}u$, *where* $\tau(u)$ *is strictly positive, càdlàg univariate process. Then we assume Refresh Times occur at* $\tau_j = T(j/n)$. *We also assume that* τ *is adapted to* $\mathcal{F}$. *When both conditions hold we write* $\tau \in \mathcal{T}$.

3 Asymptotic results

3.1 Consistency

We first give a consistency result for the multivariate realised kernel $K(X)$, which can be written $K(X) = K(Y) + K(Y,U) + K(U,Y) + K(U)$, where $K(Y,U) = \sum_{h=-n}^{n} k(\frac{h}{H+1}) \sum_j y_j u'_{j-h}$, with $y_j = Y_j - Y_{j-1}$ and $u_j = U_j - U_{j-1}$.

Theorem 1 *Let* $k \in \mathcal{K}$ *and* $n \to \infty$. *If* $K(U) \xrightarrow{p} 0$ *and* $K(Y) \xrightarrow{p} [Y]$ *then*

$$K(X) \xrightarrow{p} [Y].$$

If $H \propto n^{\eta}$ *with* $\eta \in (0,1)$ *and* $\tau \in \mathcal{T}$ *then* $K(Y) \xrightarrow{p} [Y]$. *If* $H \propto n^{\eta}$ *with* $\eta \in (1/2,1)$, $U \in \mathcal{CS}$, *and* $m \to \infty$, *then* $K(U) \xrightarrow{p} 0$. *Further, if* $K(Y) - [Y] = O_p(n^{-\epsilon})$ *and* $K(U) = O_p(n^{-2\epsilon})$ *for some* $\epsilon > 0$, *then* $K(X) - [Y] = O_p(n^{-\epsilon})$.

Note in particular that, whatever the relationship between Y and U, if $K(U) \xrightarrow{p} 0$ and $K(Y) \xrightarrow{p} [Y]$ then $K(X) \xrightarrow{p} [Y]$. Hansen & Lunde (2006) have shown that endogenous noise is empirically important, particularly for mid-quote data. The above theorem is comparatively clean, it means endogeneity does not matter for consistency. What matters is that the realised kernel applied to the noise process would converge to zero as $n \to \infty$.[6]

3.2 Central limit theory

3.2.1 Univariate asymptotic analysis of realised kernels

Before introducing the results on the multivariate case, it is helpful to consider the univariate case

$$X(t_j) = \int_0^{t_j} a(u)\mathrm{d}u + \int_0^{t_j} \sigma(u)\mathrm{d}W(u) + U_j.$$

In order to present the results for the univariate case, we write ω^2 in place of Ω, so $\omega^2 = \sum_{h=-\infty}^{\infty} \mathrm{E}(U_i U_{i-h})$.

Proposition 1 *Let* $k \in \mathcal{K}$, $\tau \in \mathcal{T}$, $H = c_0 n^{3/5}$, $U \in \mathcal{CS}$, $Y \perp\!\!\!\perp U$ *and* $m^{-1} = o(n^{-1/5})$. *Then*

$$n^{1/5}\left\{\mathrm{K}(X) - \int_0^1 \sigma^2(u)\mathrm{d}u\right\} \xrightarrow{Ls} \mathrm{MN}\left\{c_0^{-2}\left|k''(0)\right|\omega^2, 4c_0 k_{\bullet}^{0,0}\mathrm{IQ}\right\},$$

where $\mathrm{IQ} = \int_0^1 \lambda^4(u)\mathrm{d}u$ *is the integrated quarticity, and* $\lambda(t) = \tau(t)\{\sigma \circ T(t)\}$.

[6] Of course, if U had a component V, which evolved in calendar time, e.g. V is an Ornstein-Uhlenbeck process, then $U \notin \mathcal{CS}$ and $K(U)$ would not vanish in probability.

The notation $\xrightarrow{Ls}$ MN means stable convergence to a mixed Gaussian distribution. The notion of stable convergence is important for the construction of confidence intervals and the use of the delta method. The reason is that IQ is random, and stable convergence guarantees joint convergence that is needed here. Stable convergence is discussed, for example, in Mykland & Zhang (2006) and Mykland & Zhang (2008), who also provides extensive references. The presence of λ in the limit theory is due to the irregularly spaced nature of the data, if it was equally spaced then $\tau(t) = 1$ and $T(t) = t$, so IQ $= \int_0^1 \sigma^4(u)\mathrm{d}u$ as usual.

Remark. The asymptotic distribution in Proposition 1 has a non-zero asymptotic mean which implies that the upward asymptotic bias of the realised kernel is roughly $n^{-1/5}c_0^{-2}\left|k''(0)\right|\omega^2$. Having an asymptotic bias term in the asymptotic distribution is familiar from kernel density estimation with the optimal bandwidth. Here the situation is slightly easier for in principle the bias term can be estimated from the data.

We now explain why Proposition 1 is the most interesting to us. The rest of this subsection can be skipped on first reading if the reader is not interested in these background results.

To start this consider first some moments of various quantities.

Proposition 2 *Let $k \in \mathcal{K}$ and $U \in \mathcal{CS}$. Then* $\mathrm{E}\{\mathrm{K}(U)\} = \frac{n}{H^2}|k''(0)|\omega^2 + O(m^{-1}) + o(n/H^2)$ *and if, additionally $\tau \in \mathcal{T}$, the asymptotic variance of $K(Y)$ and $K(U)$ are given by*

$$\frac{H}{n}4k_\bullet^{0,0}\mathrm{IQ} \qquad \textit{and} \qquad \frac{n}{H^3}4k_\bullet^{2,2}\omega^4. \tag{2}$$

Remark. The second term in $\mathrm{E}\{K(U)\}$ highlights the need for the averaging at the end-points. The $O(m^{-1})$ term roughly equals $2m^{-1}\omega^2$, so we need $m \to \infty$ for the bias to vanish. Empirically ω^2 is tiny so $2m^{-1}\omega^2$ will be small even with $m = 1$, but theoretically this is an important observation.

Remark. The result shows that estimators in this class of realised kernels are generally biased due to the kernels not being entirely flat-top, but the bias is modest so long as H increases at a faster rate than $\sqrt{n}$. For a weight function with $k''(0) = 0$ we could take $H \propto n^{1/2}$ which would result in a faster rate of convergence. However, no weight function with $k''(0) = 0$ can guarantee a positive semi-definite estimate, see Andrews (1991, p. 832, comment 5).

Remark. If $m^{-1} = o(n^{-1/5})$, then the mean square error optimal rate for H is $H \propto n^{3/5}$, equalising the rate of the squared bias and the variance. All but the first term in (2) vanish as $n \to \infty$ when $H \propto n^{3/5}$. Note that the asymptotic bias is tied to $k''(0)$ whereas the asymptotic variance is tied to $k_\bullet^{0,0}$.

Remark. This result looks rather weak compared to the corresponding result for the flat-top kernel $K^F(X)$ introduced by Barndorff-Nielsen et al. (2008*a*) with $k'(0) = 0$. They had the nicer result that[7]

$$n^{1/4}\left\{K^F(X) - \int_0^1 \sigma^2(u)\mathrm{d}u\right\} \xrightarrow{Ls} \mathrm{MN}\left\{0, 4ck_\bullet^{0,0}\mathrm{IQ} + \tfrac{8}{c}k_\bullet^{1,1}\omega^2\int_0^1\sigma^2(u)\mathrm{d}u + \tfrac{4}{c^3}k_\bullet^{2,2}\omega^4\right\},$$

[7] See also Zhang (2006) who independently obtained a $n^{1/4}$ consistent estimator using a multiscale approach.

when $H = cn^{1/2}$, under the assumption that $U \in \mathcal{WN}$. Hence the use of non-flat top kernels comes at an asymptotic cost, but ensures positive semi-definiteness. Section 6.1.2 also shows that $K(X)$ is more robust to endogeneity and serial dependence in U than $K^F(X)$.

3.2.2 Choosing the bandwidth H and weight function

Next we turn to the optimal (mean square error) choice for the bandwidth parameter H.

Proposition 3 *Let $k \in \mathcal{K}$, $\tau \in \mathcal{T}$, $U \in \mathcal{CS}$ and set $H = c^* \xi^{4/5} n^{3/5}$, where $c^* = \left\{k''(0)^2/k_\bullet^{0,0}\right\}^{1/5}$ and $\xi^2 = \omega^2/\sqrt{IQ}$, then*

$$n^{1/5}\left\{\mathrm{K}(X) - \int_0^1 \sigma^2(u)\mathrm{d}u\right\} \xrightarrow{Ls} \mathrm{MN}(\kappa, 4\kappa^2), \quad \textit{where} \quad \kappa = \left|k''(0)(k_\bullet^{0,0})^2\right|^{1/5}\{\omega I Q\}^{2/5}.$$

The relative efficiency of different realised kernels in this class are determined solely by the constant $\left|k''(0)(k_\bullet^{0,0})^2\right|^{1/5}$ and so can be universally determined for all Brownian semimartingales and noise processes. This constant is computed for a variety of kernel weight functions in Table 1. This shows that the Quadratic Spectral (QS), Parzen and Fejér weight functions are attractive in this context. The optimal weight function minimizes, $\left|k''(0)(k_\bullet^{0,0})^2\right|^{1/5}$, which is also the situation for HAC estimators, see Andrews (1991). Thus, using Andrews' analysis of HAC estimators, it follows from our results that the QS kernel is the optimal weight function within the class of weight functions that are guaranteed to produce a non-negative realised kernel estimate. A drawback of the QS and Fejér weight functions is that they, in principle, require n (all) realised autocovariances to be computed, whereas the number of realised autocovariances needed for the Parzen kernel is only H — hence we advocate the use of Parzen weight functions. We will discuss estimating ξ^2 in Section 3.3.1.

Kernel function, $k(x)$		$\lvert k''(0)\rvert$	$k_\bullet^{0,0}$	$\left\lvert k''(0)(k_\bullet^{0,0})^2\right\rvert^{1/5}$
Parzen	$k_P(x) = \begin{cases} 1-6x^2+6x^3 & 0 \le x \le 1/2 \\ 2(1-x)^3 & 1/2 \le x \le 1 \\ 0 & x > 1 \end{cases}$	12	0.269	0.97
Quadratic Spectral	$k_{QS}(x) = \frac{3}{x^2}\left(\frac{\sin x}{x} - \cos x\right), \quad x \ge 0$	1/5	$3\pi/5$	0.93
Fejér	$k_F(x) = \left(\frac{\sin x}{x}\right)^2, \quad x \ge 0$	2/3	$\pi/3$	0.94
Tukey-Hanning$_\infty$	$k_{TH_\infty}(x) = \sin^2\left\{\frac{\pi}{2}\exp(-x)\right\}, \quad x \ge 0$	$\pi^2/2$	0.52	1.06
BNHLS (2008)	$k(x) = (1+x)\,e^{-x} \quad x \ge 0$	1	5/4	1.09

Table 1: *Properties of some realised kernels. $\left|k''(0)(k_\bullet^{0,0})^2\right|^{1/5}$ measures the relative asymptotic efficiency of $k \in \mathcal{K}$.*

3.2.3 Some multivariate notation

To start we define some terms. Let

$$\Psi = \int_0^1 \tau^2(u)\left\{\Sigma \circ T(u) \otimes \Sigma \circ T(u)\right\}\mathrm{d}u,$$

which is the $d^2 \times d^2$ random matrix analog of integrated quarticity.

Our result will use the *matrix normal distribution*. For $M \in \mathbb{R}^{q\times q}$, $M \sim N(A, B)$ simply means that $\mathrm{vec}(M)$ is Gaussian distributed with mean $\mathrm{vec}(A)$ and the covariance between $a'Mb$ and $c'Md$ is given by $\mathrm{cov}(a'Mb, c'Md) = v'_{ab}Bv_{cd}$, with $v_{ab} = \mathrm{vec}(\frac{ab'+ba'}{2})$ and $v_{cd} = \mathrm{vec}(\frac{cd'+dc'}{2})$.

3.2.4 Multivariate central limit theorem

Theorem 2 *Suppose $H = c_0 n^{3/5}$, $\tau \in \mathcal{T}$, $U \in \mathcal{CS}$, $m^{-1} = o(n^{-1/5})$, $Y \perp\!\!\!\perp U$ and $k \in \mathcal{K}$ then*

$$n^{1/5}\left\{K(X) - \int_0^1 \Sigma(u)\mathrm{d}u\right\} \xrightarrow{Ls} \mathrm{MN}\{c_0^{-2}|k''(0)|\Omega, 4c_0 k_\bullet^{0,0}\Psi\}.$$

This is the multivariate extension of Proposition 1, yielding a limit theorem for the consistent multivariate estimator in the presence of noise. The bias is determined by the long-run variance Ω, the variance solely by integrated quarticity.

Corollary 1 *An implication of Theorem 2 is that for $a, b \in \mathbb{R}^d$ we have*

$$n^{1/5}a'\left\{K(X) - \int_0^1 \Sigma(u)\mathrm{d}u\right\} b \xrightarrow{Ls} \mathrm{MN}\left\{c_0^{-2}|k''(0)|a'\Omega b, 4c_0 k_\bullet^{0,0} v'_{ab}\Psi v_{ab}\right\}.$$

So once a consistent estimator for Ψ is obtained, Corollary 1 makes it straightforward to compute a confidence interval for any element of the integrated variance matrix.

Example 1 *In the bivariate case we can write the results as*

$$n^{1/5}\begin{pmatrix} K(X^{(i)}) - \int_0^1 \Sigma_{ii}\mathrm{d}u \\ K(X^{(i)}, X^{(j)}) - \int_0^1 \Sigma_{ij}\mathrm{d}u \\ K(X^{(j)}) - \int_0^1 \Sigma_{jj}\mathrm{d}u \end{pmatrix} \xrightarrow{Ls} \mathrm{MN}(A, B), \tag{3}$$

where

$$A = c_0^{-2}|k''(0)|\begin{pmatrix} \Omega_{ii} \\ \Omega_{ij} \\ \Omega_{jj} \end{pmatrix} \quad \textit{and} \quad B = 2c_0 k_\bullet^{0,0}\int_0^1 \tau^2 \begin{pmatrix} 2\Sigma_{ii}^2 & 2\Sigma_{ii}\Sigma_{ij} & 2\Sigma_{ij}^2 \\ \bullet & \Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2 & 2\Sigma_{ii}\Sigma_{ji} \\ \bullet & \bullet & 2\Sigma_{jj}^2 \end{pmatrix} \circ T\mathrm{d}u,$$

which has features in common with the noiseless case discussed in Barndorff-Nielsen & Shephard (2004, eq. 18). By the delta method we can deduce the asymptotic distribution of the kernel based regression and correlation. For example

$$n^{1/5}\left(\frac{K(X^{(i)}, X^{(j)})}{K(X^{(j)})} - \beta_{ij}\right) \xrightarrow{Ls} \mathrm{MN}(A, B),$$

where

$$A = \frac{c_0^{-2}|k''(0)|}{\int_0^1 \Sigma_{jj}\mathrm{d}u}\left(\Omega_{ij} - \Omega_{jj}\beta_{ij}\right), \quad \beta_{ij} = \frac{\int_0^1 \Sigma_{ij}\mathrm{d}u}{\int_0^1 \Sigma_{jj}\mathrm{d}u},$$

and

$$B = \frac{2c_0 k_\bullet^{0,0}}{\left(\int_0^1 \Sigma_{jj}\mathrm{d}u\right)^2}\left(1, -\beta_{ij}\right)\left[\int_0^1 \tau^2 \begin{pmatrix} \Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2 & 2\Sigma_{ii}\Sigma_{ji} \\ 2\Sigma_{jj}\Sigma_{ij} & 2\Sigma_{jj}^2 \end{pmatrix} \circ T\mathrm{d}u\right]\begin{pmatrix} 1 \\ -\beta_{ij} \end{pmatrix}.$$

To produce the result (3) we notice the asymptotic variance consists of terms $4c_0 k_\bullet^{0,0} v'_{e_i e_h}\Psi v_{e_j e_k}$ *where* e_i *denotes the* i*-th unit vector in* $\mathbb{R}^d$*. Consider, for simplicity, the case with equidistant sampling times, so that* $\Psi = \int_0^1 \Sigma(u)\otimes\Sigma(u)\mathrm{d}u$*. Then*

$$\begin{aligned} 4v'_{e_i e_h}\Psi v_{e_j e_k} &= \mathrm{vec}(e_i e_h' + e_h e_i')'\left\{\int(\Sigma\otimes\Sigma)\mathrm{d}u\right\}\mathrm{vec}(e_j e_k' + e_k e_j') \\ &= 2\int \mathrm{tr}\{\Sigma e_i e_h'\Sigma e_k e_j' + \Sigma e_i e_h'\Sigma e_j e_k'\}\mathrm{d}u = 2\int(\Sigma_{ik}\Sigma_{jh}+\Sigma_{ih}\Sigma_{jk})\mathrm{d}u, \end{aligned}$$

and the result follows by using various combinations of (i, h, k, j)*.*

3.3 Some practical issues

3.3.1 Choice of H in practice

A main feature of multivariate kernels is that there is a single bandwidth parameter H which controls the number of leads and lags used for all the series. It must grow with n at rate $n^{3/5}$, the key question here is how to estimate a good constant of proportionality — which controls the efficiency of the procedure.

If we applied the univariate optimal mean square error bandwidth selection to each asset price individually we would get d bandwidths $H^{(i)} = c^*\xi_i^{4/5}n^{3/5}$, where $c^* = \left\{k''(0)^2/k_\bullet^{0,0}\right\}^{1/5}$ and $\xi_i^2 = \Omega_{ii}/\sqrt{IQ_{ii}}$, where $\Sigma_{ii}(u)$ is the spot variance for the i-th asset. In practice we usually approximate $\sqrt{IQ_{ii}}$ by $\int_0^1\Sigma_{ii}(u)\mathrm{d}u$ and use $\xi_i^2 = \Omega_{ii}/\int_0^1\Sigma_{ii}(u)\mathrm{d}u$, which can be estimated relatively easily by using a low frequency estimate of $\int_0^1\Sigma_{ii}(u)\mathrm{d}u$ and one of many sensible estimators of Ω_{ii} which use high frequency data. Then we could construct some ad hoc rules for choosing the global H, such as

$$H_{\min} = \min(H^{(1)}, ..., H^{(d)}), \quad H_{\max} = \max(H^{(1)}, ..., H^{(d)}), \quad \text{or} \quad \bar{H} = d^{-1}\sum_{i=1}^d H^{(i)},$$

or many others. In our empirical work we have used $\bar{H}$, while our web Appendix provides an analysis of the impact of this choice.

An interesting alternative is to optimise the problem for a portfolio, e.g. letting ι be a d-dimensional vector of ones then $d^{-2}\iota' K(X)\iota = K\left(d^{-1}\iota' X\right)$, which is like a "market portfolio" if X contains many assets. This is easy to carry out, for having converted everything into Refresh Time one computes the market $(\iota' X/\iota'\iota)$ return and then carry out a univariate analysis on it, choosing an optimal H for the market. This single H is then applied to the multivariate problem.

From the results in Example 1 it is straightforward to derive the optimal choice for H, when the objective is to estimate a covariance, a correlation, the inverse covariance matrix (which is important for

portfolio choice) or β_{ij}. For example, for β_{12} the trade-off is between $c_0^{-4}|k''(0)|^2\left(\Omega_{12}-\Omega_{22}\beta_{12}\right)^2$, and

$$2c_0k_\bullet^{0,0}\int_0^1\tau^2\left(\Sigma_{11}\Sigma_{22}+\Sigma_{12}^2-4\beta_{12}\Sigma_{11}\Sigma_{22}+2\beta_{12}^2\Sigma_{22}\right)\circ T\mathrm{d}u.$$

3.3.2 Realised kernel based beta and correlation

A key reason for needing our realised kernel to be positive semi-definite is that elements of it can be combined to consistently estimate the quadratic variation version of the beta and correlation between assets i and j

$$\beta^{(i,j)}=\frac{[Y^{(i)},Y^{(j)}]}{[Y^{(j)}]}\quad\text{and}\quad\rho^{(i,j)}=\frac{[Y^{(i)},Y^{(j)}]}{\sqrt{[Y^{(i)}][Y^{(j)}]}},$$

where we have written $[Y]=\left\{[Y^{(i)},Y^{(j)}]\right\}_{i,j=1,2,...}$. The quantities $\beta^{(i,j)}$ and $\rho^{(i,j)}$ have been highlighted in previous research by, for example, Andersen, Bollerslev, Diebold & Labys (2003), Barndorff-Nielsen & Shephard (2004) and Dovonon, Goncalves & Meddahi (2007), but their work was hampered by only being able to use 5-15 minute returns due to the effect of noise and irregularly spaced data.

The realised kernel estimators of these quantities are straightforward and the asymptotic distribution simply follows by the application of the delta method. In particular

$$\hat{\beta}^{(i,j)}=\frac{K\left(X^{(i)},X^{(j)}\right)}{K\left(X^{(j)}\right)}\quad\text{and}\quad\hat{\rho}^{(i,j)}=\frac{K\left(X^{(i)},X^{(j)}\right)}{\sqrt{K\left(X^{(i)}\right)K\left(X^{(j)}\right)}}\in[-1,1],$$

where we have written the elements of the realised kernel matrix $K(X)$ as $K(X)=\left\{K\left(X^{(i)},X^{(j)}\right)\right\}_{i,j=1,2,...}$.

4 Simulation Study

So far the analysis has been asymptotic, based on $n\to\infty$. Here we reinforce this by carrying out a simulation analysis to assess the accuracy of the asymptotic predictions in finite samples. We simulate over the interval $t\in[0,1]$.

The following multivariate factor stochastic volatility model is used

$$\mathrm{d}Y^{(i)}=\mu^{(i)}\mathrm{d}t+\mathrm{d}V^{(i)}+\mathrm{d}F^{(i)},\quad\mathrm{d}V^{(i)}=\rho^{(i)}\sigma^{(i)}\mathrm{d}B^{(i)},\quad\mathrm{d}F^{(i)}=\sqrt{1-\left(\rho^{(i)}\right)^2}\sigma^{(i)}\mathrm{d}W.$$

where the elements of B are independent standard Brownian motions and $W\perp\!\!\!\perp B$. Here $F^{(i)}$ is the common factor, whose strength is determined by $\sqrt{1-\left(\rho^{(i)}\right)^2}$.

This model means that each $Y^{(i)}$ is a diffusive SV model with constant drift $\mu^{(i)}$ and random spot volatility $\sigma^{(i)}$. In turn the spot volatility obeys the independent processes $\sigma^{(i)}=\exp\left(\beta_0^{(i)}+\beta_1^{(i)}\varrho^{(i)}\right)$ with $\mathrm{d}\varrho^{(i)}=\alpha^{(i)}\varrho^{(i)}\mathrm{d}t+\mathrm{d}B^{(i)}$. Thus there is perfect statistical leverage (correlation between their innovations) between $V^{(i)}$ and $\sigma^{(i)}$, while the leverage between $Y^{(i)}$ and $\varrho^{(i)}$ is $\rho^{(i)}$. The correlation between $Y^{(1)}(t)$ and $Y^{(2)}(t)$ is $\sqrt{1-\left(\rho^{(1)}\right)^2}\sqrt{1-\left(\rho^{(2)}\right)^2}$.

The price process is simulated via an Euler scheme[8], and the fact that the OU-process have an exact discretization (see e.g. Glasserman (2004, pp. 110)). Our simulations are based on the following configuration (the same for both processes) $(\mu^{(i)}, \beta_0^{(i)}, \beta_1^{(i)}, \alpha^{(i)}, \rho^{(i)}) = (0.03, -5/16, 1/8, -1/40, -0.3)$, so that $\beta_0^{(i)} = (\beta_1^{(i)})^2/(2\alpha^{(i)})$. Throughout we have imposed that $\mathrm{E}\left(\int_0^1 \sigma^{(i)2}(u)\mathrm{d}u\right) = 1$. The stationary distribution of $\varrho^{(i)}$ is utilised in our simulations to restart the process each day at $\varrho^{(i)}(0) \sim N(0, (-2\alpha^{(i)})^{-1})$. For our design we have that the variance of σ^2 is $\exp(-2(\beta_1^{(i)})^2/\alpha^{(i)}) - 1 \simeq 2.5$. This is comparable to the empirical results found in e.g. Hansen & Lunde (2005) which motivate our choice for $\alpha^{(i)}$.

We add noise simulated as

$$U_j^{(i)}|\sigma, Y \overset{i.i.d.}{\sim} N\left(0, \omega^2\right) \quad \text{with} \quad \omega^2 = \xi^2 \sqrt{N^{-1} \sum\nolimits_{j=1}^{N} \sigma^{(i)4}(j/N)},$$

where the noise-to-signal ratio, ξ^2 takes the values 0, 0.001 and 0.01. This means that the variance of the noise increases with the volatility of the efficient price (e.g. Bandi & Russell (2006)). The observed process is then given by $X(j/N) = Y(j/N) + U_j$, $j = 0, \ldots, N$.

To model the non-synchronously spaced data we use two independent Poisson process sampling schemes to generate the times of the actual observations $\left\{t_j^{(i)}\right\}$ to which we apply our realised kernel. We control the two Poisson processes by $\lambda = (\lambda_1, \lambda_2)$, such that for example $\lambda = (5, 10)$ means that on average $X^{(1)}$ and $X^{(2)}$ is observed every 5 and 10 second, respectively. This means that the simulated number of observations will differ between repetitions, but on average the processes will have $23400/\lambda_1$ and $23400/\lambda_2$ observations, respectively.

We vary λ though the following configurations $(3, 6)$, $(5, 10)$, $(10, 20)$, $(15, 30)$, $(30, 60)$, $(60, 120)$ motivated by the kind of data we see in databases of equity prices.

For each simulated day we compute the observed the price process, $X(j/N)$. In order to calculate $K(X)$ we need to select H. To do this we evaluate $\hat{\omega}_\delta^{(i)2} = [X_\delta^{(i)}](1)/(2n)$ and $[X_{1/900}^{(i)}](1)$, the realised variance estimator based on 15 minute returns. These give us the following feasible values $\hat{H}_i = cn^{3/5}\left(\hat{\omega}_\delta^{(i)2}/[X_{1/900}^{(i)}](1)\right)^{2/5}$. The results for H_{mean} are presented in Table 2.

Panel A of the table reports the univariate results of estimating integrated variance. We give the bias and root mean square error (MSE) for the realised kernel and compare it to the standard realised variance. In the no noise case of $\xi^2 = 0$ the RV statistic is quite a bit more precise, especially when n is large. The positive bias of the realised kernel can be seen when ξ^2 is quite large, but it is small compared to the estimators variance. In that situation the realised kernel is far more precise than the realised variance. None of these results are surprising or novel.

In Panel B we break new ground as it focuses on estimating the integrated covariance. We compare the realised kernel estimator with a realised covariance. The high frequency realised covariance is a very

[8]We normalize one second to be $1/23,400$, so that the interval $[0, 1]$ contains 6.5 hours. In generating the observed price, we discretize $[0, 1]$ into a number $N = 23,400$ of intervals.

precise estimator of the wrong quantity as its bias is very close to its very large mean square error. In this case its bias does not really change very much as n increases.

Table 2: Simulation results

Panel A: Integrated Variance

	Series A				Series B			
	RV^{1m}	RV^{15m}	$K(X)$		RV^{1m}	RV^{15m}	$K(X)$	
$\xi^2 = 0.0$	R.mse	R.mse	bias	R.mse	R.mse	R.mse	bias	R.mse
$\lambda = (3, 6)$	0.113	0.505	0.006	0.147	0.122	0.436	0.003	0.134
$\lambda = (10, 20)$	0.111	0.547	0.011	0.262	0.114	0.450	0.011	0.224
$\lambda = (60, 120)$	0.229	0.504	0.003	0.557	0.227	0.517	0.001	0.490
$\xi^2 = 0.001$								
$\lambda = (3, 6)$	1.509	0.654	0.040	0.253	1.417	0.488	0.033	0.215
$\lambda = (10, 20)$	1.432	0.660	0.041	0.359	1.318	0.492	0.035	0.295
$\lambda = (60, 120)$	1.013	0.559	0.014	0.557	0.636	0.554	0.013	0.551
$\xi^2 = 0.01$								
$\lambda = (3, 6)$	14.39	1.531	0.096	0.410	13.67	1.168	0.084	0.351
$\lambda = (10, 20)$	14.01	1.452	0.106	0.568	13.15	1.305	0.081	0.424
$\lambda = (60, 120)$	8.893	1.222	0.077	0.611	5.386	1.322	0.080	0.776

Panel B: Integrated Covariance/Correlation

		Cov^{1m}		Cov^{15m}		$K(X)$ Covar		$K(X)$ Corr		$K(X)$ beta	
$\xi^2 = 0.0$	#rets	bias	R.mse	bias	R.mse	bias	R.mse	bias	R.mse	bias	R.mse
$\lambda = (3, 6)$	3,121	-0.051	0.076	-0.004	0.183	-0.007	0.062	-0.012	0.016	-0.016	0.061
$\lambda = (5, 10)$	1,921	-0.085	0.108	-0.006	0.183	-0.009	0.076	-0.015	0.020	-0.019	0.064
$\lambda = (10, 20)$	982	-0.160	0.186	-0.011	0.186	-0.009	0.097	-0.018	0.026	-0.023	0.084
$\lambda = (30, 60)$	332	-0.342	0.395	-0.038	0.188	-0.021	0.142	-0.028	0.042	-0.035	0.125
$\lambda = (60, 120)$	166	-0.445	0.510	-0.071	0.203	-0.034	0.189	-0.036	0.054	-0.035	0.178
$\xi^2 = 0.001$											
$\lambda = (3, 6)$	3,121	-0.046	0.091	-0.005	0.191	-0.000	0.090	-0.027	0.032	-0.034	0.085
$\lambda = (5, 10)$	1,921	-0.082	0.123	-0.006	0.186	-0.002	0.099	-0.029	0.036	-0.033	0.083
$\lambda = (10, 20)$	982	-0.156	0.189	-0.010	0.195	-0.004	0.118	-0.032	0.040	-0.042	0.111
$\lambda = (30, 60)$	332	-0.344	0.400	-0.039	0.187	-0.019	0.150	-0.039	0.052	-0.049	0.153
$\lambda = (60, 120)$	166	-0.445	0.513	-0.074	0.206	-0.034	0.195	-0.044	0.060	-0.049	0.204
$\xi^2 = 0.01$											
$\lambda = (3, 6)$	3,121	-0.027	0.398	-0.009	0.263	0.000	0.123	-0.063	0.071	-0.072	0.132
$\lambda = (5, 10)$	1,921	-0.073	0.431	-0.005	0.257	-0.002	0.133	-0.067	0.076	-0.082	0.149
$\lambda = (10, 20)$	982	-0.139	0.407	-0.001	0.263	-0.005	0.153	-0.074	0.084	-0.099	0.198
$\lambda = (30, 60)$	332	-0.354	0.486	-0.044	0.236	-0.017	0.180	-0.089	0.104	-0.119	0.242
$\lambda = (60, 120)$	166	-0.451	0.561	-0.083	0.265	-0.032	0.222	-0.092	0.111	-0.120	0.310

Simulation results for the realised kernel using a factor SV model with non-syncronous observations and measurement noise. Panel A looks at estimating integrated variance using realised variance and the Parzen type realised kernel $K(X)$. Panel B looks at estimating integrated covariance and correlation using realised covariance and realised kernel. Bias and root mean square error are reported.

The realised kernel delivers a very precise estimator of the integrated covariance. It is downward biased due to the non-synchronous data, but the bias is very modest when n is large and its sampling vari-

ance dominates the root MSE. Taken together this implies the realised kernel estimators of the correlation and regression (beta) are less good. Peter says: Both are strongly negatively biased — which is due to it being a non-linear function of the noisy estimates of the integrated variance. The bias is the dominant component of the root MSE of the

5 Empirical illustration

We analyze high-frequency stock prices for ten assets, namely Alcoa Inc. (AA), American International Group Inc. (AIG), American Express Co. (AXP), Boeing Co. (BA), Bank of America Corp. (BAC), Citygroup Inc. (C), Caterpillar Inc. (CAT), Chevron Corp.(CVX), General Electric Co. (GE), and Standard & Poor's Depository Receipt (SPY). The SPY is an exchange-traded fund that holds all of the S&P 500 Index stocks and has enormous liquidity. The sample period runs from January 3, 2005 to June 29, 2007, delivering 626 distinct days. The data is the collection of trades and quotes recorded on the NYSE, taken from the TAQ database through the Wharton Research Data Services (WRDS) system. We present empirical results for both transaction and mid-quote prices.

Throughout our analysis we will estimate quantities each day, in the tradition of the realised volatility literature following, for example, Andersen et al. (2001) and Barndorff-Nielsen & Shephard (2002). This means the target becomes functions of $[Y]_s = [Y](s) - [Y](s-1)$, $s \in \mathbb{N}$. The functions we will deal with are covariances, correlations and betas.

5.1 Procedure for cleaning the high-frequency data

Careful data cleaning is one of the most important aspects of volatility estimation from high-frequency data. Numerous problems and solutions are discussed in Falkenberry (2001), Hansen & Lunde (2006), Brownless & Gallo (2006) and Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*b*). In this paper we follow the step-by-step cleaning procedure used in Barndorff-Nielsen et al. (2008*b*) who discuss in detail the various choices available and their impact on univariate realised kernels. For convenience we briefly review these steps.

All data: P1) Delete entries with a timestamp outside the 9:30 a.m. to 4 p.m. window when the exchange is open. P2) Delete entries with a bid, ask or transaction price equal to zero. P3) Retain entries originating from a single exchange (NYSE except for SPY for which all retained observations are from Pacific). Delete other entries.

Trade data only: T1) Delete entries with corrected trades. (Trades with a *Correction Indicator*, CORR $\neq$ 0). T2) Delete entries with abnormal *Sale Condition*. (Trades where COND has a letter code, except for "E" and "F"). T3) If multiple transactions have the same time stamp: use the median price. T4) Delete entries with prices that are above the *ask* plus the bid-ask spread. Similar for entries with prices below the *bid* minus the bid-ask spread.

Quote data only: Q1) When multiple quotes have the same timestamp, we replace all these with a single entry with the median bid and median ask price. Q2) Delete rows for which the spread is negative. Q3) Delete rows for which the spread is more that 10 times the median spread on that day. Q4) Delete rows for which the mid-quote deviated by more than 5 mean absolute deviations from a centered mean (excluding the observation under consideration) of 50 observations. We note steps P2, T1, T2, T4, Q2, Q3 and Q4 collectively reduce the sample size by less than 1%.

5.2 Sampling schemes

We applied three different sampling schemes depending on the particular estimator. The simplest one is the HY estimator that uses all the available observations for a particular asset combination. Following Andersen et al. (2003) the realised covariation estimator is based on calender time sampling. Specifically, we consider 15 second, 5 minute, and 30 minute intraday return, aligned using the previous tick approach. This results in 1560, 78 and 13 daily observations, respectively.

For the realised kernel the Refresh Time sampling scheme discussed in section 2.1.1 is used. Our analysis first considers estimates for each of the 45 unique pairs of assets — delivering 45 distinct 2×2 covariance matrix estimates each day.

Table 3: Summary statistics for the refresh sampling scheme, 2×2 case

2×2 *case*

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA		0.601	0.597	0.594	0.601	0.594	0.587	0.570	0.596	0.568
AIG	0.673		0.600	0.602	0.624	0.628	0.590	0.603	0.625	0.603
AXP	0.665	0.670		0.600	0.602	0.585	0.590	0.552	0.585	0.548
BA	0.662	0.667	0.663		0.599	0.592	0.590	0.568	0.592	0.569
BAC	0.681	0.691	0.678	0.673		0.634	0.592	0.605	0.628	0.604
C	0.687	0.700	0.681	0.678	0.717		0.582	0.624	0.642	0.627
CAT	0.647	0.648	0.650	0.646	0.655	0.657		0.560	0.584	0.562
CVX	0.680	0.690	0.671	0.670	0.707	0.719	0.649		0.620	0.620
GE	0.686	0.699	0.677	0.675	0.719	0.733	0.653	0.726		0.619
SPY	0.678	0.696	0.658	0.665	0.721	0.747	0.633	0.743	0.762	

Average over daily number of high frequency observations available before the Refresh Time transformation

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
Trades	4,124	4,789	3,528	4,057	4,757	5,687	4,039	6,292	5,460	6,554
Quotes	11,222	11,738	10,482	10,717	12,562	13,393	9,937	13,573	14,189	18,587

Summary statistics for the refresh sampling scheme. In the two upper panels we present averages over the daily data reduction induced by the refresh sampling scheme, measured by $p = dN/\sum_{i=1}^{d} n^{(i)}$. *The upper panel display this in the 2×2 case. The upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. In the lower panel we average over the daily number of high frequency observations.*

The amount of data we discard by constructing Refresh Time is recorded in Table 3. It records the average of the daily p statistics defined in Section 2.1.1 for each pair. It emerges that we rarely lose more

that half the observations for most frequently traded assets. For the least active assets we typically lose between 30% to 40% of the observations.

We will also apply the realised kernel to the full 10×1 vector of returns. Here the data loss is more pronounced. Still, even in the worst case more that 20 percent of the observations remain in the sample. For transaction data the average number of Refresh Time observations in $1,222$, whereas the corresponding number is $3,942$ for the quote data. So in most cases we have an observation on average more often than every 8 seconds for quote data and 20 seconds for trade data.

5.3 Analysis of the covariance estimators: Cov_s^K, Cov_s^{HY} and $Cov_s^{\Delta \mathrm{m}}$

Throughout this subsection the target which we wish to estimate is $[Y^{(i)}, Y^{(j)}]_s$, $i, j = 1, 2, ..., d$, $s \in \mathbb{N}$. In what follows the pair i, j will only be referred to implicitly. All kernels are computed with Parzen weights.

We compute the realised kernel for (all possible) pairs of assets and for the full 10-dimensional vector of assets, and the resulting estimates of $[Y^{(i)}, Y^{(j)}]_s$ are denoted by $\mathrm{Cov}_s^{K_{2\times2}}$ and $\mathrm{Cov}_s^{K_{10\times10}}$, respectively. The two estimators differ in a number of ways, such as the bandwidth selection and the sampling times (due to the construction of Refresh Time).

To provide useful benchmarks for these estimators we also compute: Cov_s^{HY}, the Hayashi & Yoshida (2005) covariance estimator. Cov_s^{Δ}, the realised covariance based on intraday returns that span a interval of length Δ, e.g. 5 or 30 minutes (the previous-tick method is used). Cov_s^{OtoC}, the outer products of the open to close returns, which when averaged over many days provide an estimator of the average covariance between asset returns.

The empirical analysis of our estimators of the covariance is started by recalling the main statistical impact of market microstructure and the Epps effect. Table 4 contains the time series average covariance computed using the 15-second realised covariance $\mathrm{Cov}_s^{15\mathrm{s}}$, the Hayashi & Yoshida (2005) estimator Cov_s^{HY} and the open to close estimator Cov_s^{OtoC}. Quite a few of these types of tables will be presented and they all have the same structure. The numbers above the leading diagonal are results from trade data, the numbers below are from mid-quotes. Both $\mathrm{Cov}_s^{15\mathrm{s}}$ and the Cov_s^{HY} are typically much lower than Cov_s^{OtoC}. The numbers which are bolded are statistically significantly different from the Cov_s^{OtoC} numbers at the one percent level. This assessment is carried out in the following way.

For a given estimator, e.g. $\mathrm{Cov}_s^{K_{2\times2}}$, we consider the difference $d_s = \mathrm{Cov}_s^{K_{2\times2}} - \mathrm{Cov}_s^{OtoC}$, and compute the sample bias as $\bar{d}$ and robust (HAC) variance as $\overline{e}^2 = \gamma_0 + 2\sum_{h=1}^{q}\left(1 - \frac{h}{q+1}\right)\gamma_h$, where $\gamma_h = \frac{1}{T-h}\sum_{s=1}^{n-h}\eta_s\eta_{s-h}$. Here $\eta_s = d_t - \bar{d}$ and $q = \mathrm{int}\left\{4(T/100)^{2/9}\right\}$. The number is boldfaced if $\left|\sqrt{T}\bar{d}/\overline{e}\right| > 2.326$. The results in Table 4 indicate the $\mathrm{Cov}_s^{1/4\mathrm{m}}$ is severely downward bias, while Cov_s^{HY} is even more distorted. In both cases nearly every covariance estimator for every pair of assets for both trades and quotes seem statistically significantly biased.

Table 4: Average high frequency realised covariance

Average of realised covariances

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	2.370	0.212	**0.194**	**0.230**	0.182	0.213	**0.236**	**0.247**	0.167	**0.201**
AIG	0.212	1.126	**0.175**	0.192	**0.171**	**0.198**	**0.180**	0.201	**0.150**	**0.172**
AXP	**0.193**	**0.178**	0.901	**0.173**	**0.158**	**0.182**	**0.178**	**0.185**	**0.138**	**0.156**
BA	**0.230**	0.201	**0.185**	1.287	**0.162**	**0.195**	**0.215**	0.211	**0.155**	**0.176**
BAC	0.182	**0.171**	**0.160**	**0.168**	0.807	**0.192**	**0.162**	0.184	**0.139**	**0.151**
C	0.209	**0.197**	**0.186**	**0.197**	**0.189**	0.924	**0.194**	0.213	**0.161**	**0.183**
CAT	**0.249**	**0.194**	**0.190**	**0.227**	**0.171**	**0.203**	1.450	**0.214**	**0.151**	**0.174**
CVX	**0.251**	0.209	**0.194**	0.218	0.186	0.215	**0.228**	1.648	0.165	**0.205**
GE	0.161	**0.147**	**0.133**	**0.150**	**0.136**	**0.157**	**0.152**	0.163	0.887	**0.138**
SPY	**0.203**	**0.178**	**0.163**	**0.190**	**0.150**	**0.179**	**0.192**	**0.212**	**0.130**	0.300

Average of Hayashi-Yoshida covariances (all times)

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	2.842	0.185	**0.182**	**0.208**	**0.160**	**0.177**	**0.215**	**0.207**	0.154	**0.175**
AIG	**0.116**	1.318	**0.163**	0.179	**0.153**	**0.172**	**0.170**	0.171	**0.141**	**0.141**
AXP	**0.112**	**0.110**	1.017	**0.168**	**0.144**	**0.163**	**0.170**	**0.164**	**0.132**	**0.142**
BA	**0.127**	**0.122**	**0.112**	1.390	**0.150**	**0.170**	**0.208**	0.185	**0.146**	**0.152**
BAC	**0.096**	**0.103**	**0.091**	**0.093**	1.096	**0.161**	**0.154**	**0.158**	**0.129**	**0.126**
C	**0.111**	**0.115**	**0.106**	**0.110**	**0.103**	1.211	**0.170**	0.173	**0.142**	**0.143**
CAT	**0.140**	**0.122**	**0.120**	**0.137**	**0.099**	**0.120**	1.471	**0.197**	**0.149**	**0.154**
CVX	**0.131**	**0.119**	**0.115**	**0.123**	**0.105**	**0.118**	**0.131**	1.718	0.147	**0.156**
GE	**0.088**	**0.089**	**0.076**	**0.087**	**0.078**	**0.087**	**0.090**	**0.093**	1.655	**0.120**
SPY	**0.087**	**0.083**	**0.079**	**0.089**	**0.065**	**0.078**	**0.093**	**0.094**	**0.054**	0.292

Open-to-close covariance

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	1.637	0.259	0.350	0.456	0.264	0.307	0.664	0.618	0.227	0.405
AIG	0.259	0.871	0.356	0.268	0.287	0.322	0.351	0.268	0.256	0.283
AXP	0.347	0.353	0.867	0.323	0.377	0.422	0.435	0.344	0.304	0.360
BA	0.453	0.265	0.315	1.371	0.277	0.297	0.559	0.326	0.302	0.355
BAC	0.265	0.288	0.378	0.278	0.524	0.394	0.301	0.256	0.260	0.287
C	0.311	0.321	0.421	0.293	0.391	0.660	0.330	0.270	0.305	0.318
CAT	0.656	0.350	0.428	0.550	0.302	0.327	1.585	0.539	0.342	0.437
CVX	0.612	0.265	0.340	0.321	0.257	0.265	0.533	1.447	0.188	0.401
GE	0.232	0.257	0.307	0.301	0.264	0.304	0.340	0.185	0.532	0.262
SPY	0.409	0.283	0.375	0.363	0.295	0.315	0.427	0.398	0.261	0.349

The upper panel presents average estimates for Cov_s^{15s} *and the middle and lower panels display these for* Cov_s^{HY} *and* Cov_s^{OtoC}*, respectively. In all panels the upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. The diagonal elements are the average of IV estimates based on transactions. Outside the diagonals numbers are boldfaced if the bias is significant at the 1 percent level.*

5.4 Results for $Cov_s^{K_{2\times2}}$, $Cov_s^{K_{10\times10}}$ and Cov_s^{5m}

We now move on to more successful estimators. The upper panel of Table 5 presents the time series average estimates for $\text{Cov}_s^{K_{2\times2}}$, the middle panel for $\text{Cov}_s^{K_{10\times10}}$, and the lower panel give results for Cov_s^{5m}. The diagonal elements are the estimates based on transactions. Off-diagonal numbers are boldfaced if

they are significantly biased (compared to Cov_s^{OtoC}) at the 1 percent level.

Table 5: Averages for alternative integrated covariance estimators

Average of Parzen covariances (2×2)

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	2.278	0.307	0.351	0.388	0.326	0.357	0.576	0.560	0.308	0.402
AIG	0.310	0.999	0.286	0.247	0.299	0.310	0.308	0.212	0.258	0.281
AXP	0.352	0.284	0.833	0.275	0.323	0.341	0.341	0.239	0.264	**0.289**
BA	0.390	0.254	0.277	1.207	0.267	0.285	0.417	0.256	0.264	0.305
BAC	0.328	0.297	0.320	0.272	0.681	0.380	0.324	0.245	0.263	0.292
C	0.355	0.306	**0.331**	0.288	0.373	0.778	0.347	0.267	0.291	0.314
CAT	0.566	0.313	0.339	0.419	0.326	0.348	1.684	0.401	0.309	0.387
CVX	0.535	0.221	0.246	0.264	0.253	0.272	0.399	1.660	0.225	0.361
GE	0.308	0.256	0.261	0.261	0.264	0.286	0.306	0.229	0.639	0.274
SPY	0.401	0.282	**0.289**	0.310	0.291	0.311	0.389	0.361	0.270	0.325

Average of Parzen covariances (10×10)

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	2.168	0.289	0.346	0.405	0.327	0.357	0.649	0.619	0.275	0.396
AIG	0.292	0.943	0.294	0.234	0.288	0.310	0.283	0.188	0.251	0.259
AXP	0.343	0.295	0.838	0.296	0.352	0.355	0.370	**0.243**	0.268	**0.292**
BA	0.381	0.238	0.287	1.215	0.271	0.281	0.462	0.241	0.248	0.295
BAC	0.324	0.294	0.350	0.267	0.645	0.394	0.328	0.235	0.249	0.283
C	0.351	0.317	0.355	0.282	0.398	0.705	0.349	0.238	0.282	0.300
CAT	0.628	0.282	0.353	0.446	0.321	0.342	1.622	0.420	0.306	0.388
CVX	0.599	0.194	**0.235**	0.234	0.240	0.247	0.398	1.563	0.173	0.334
GE	0.280	0.257	0.269	0.250	0.254	0.285	0.302	0.182	0.585	0.247
SPY	0.391	0.264	**0.289**	0.291	0.285	0.304	0.379	0.338	0.252	0.296

Average of 5 min realised covariance (pre-tick times)

	AA	AIG	AXP	BA	BAC	C	CAT	CVX	GE	SPY
AA	2.315	0.312	0.347	0.378	0.318	0.356	0.539	0.526	0.303	0.397
AIG	0.310	0.996	0.274	0.254	0.272	0.292	0.300	0.219	0.239	0.269
AXP	0.342	0.275	0.833	0.272	0.309	**0.323**	0.327	0.240	0.251	**0.281**
BA	0.380	0.253	0.275	1.239	0.264	0.284	0.401	0.260	0.252	0.303
BAC	0.322	0.273	0.306	0.265	0.686	0.361	0.305	0.246	0.246	0.276
C	0.358	0.294	**0.323**	0.283	0.361	0.790	0.342	0.268	0.275	0.303
CAT	0.538	0.300	**0.322**	0.405	0.307	0.342	1.657	0.377	0.297	0.373
CVX	0.527	0.219	0.244	0.263	0.246	0.267	0.378	1.658	0.222	0.349
GE	0.303	0.243	0.250	0.249	0.247	0.275	0.298	0.223	0.644	0.256
SPY	0.393	0.269	**0.280**	0.303	0.274	0.303	0.376	0.350	0.254	0.324

The upper panel presents average estimates for $\text{Cov}_s^{K_{2\times2}}$*, the middle panel for* $\text{Cov}_s^{K_{10\times10}}$*, and the lower panel gives results for* $\text{Cov}_s^{5\text{m}}$*. In both panels the upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. The diagonal elements are the average of IV estimates based on transactions. Outside the diagonals numbers are boldfaced if the bias is significant at the 1 percent level.*

These results are quite encouraging for all three estimators. The average levels of the three estimators are roughly the same. $\text{Cov}_s^{K_{2\times2}}$ has three failures. $\text{Cov}_s^{K_{10\times10}}$ has four failures while $\text{Cov}_s^{5\text{m}}$ is rejected five times. All three estimators reject for the SPY/AXP combination, both for trades and quotes.

A much tougher comparison is to replace the noisy $d_s = \text{Cov}_s^K - \text{Cov}_s^{OtoC}$ with $d_s = \text{Cov}_s^{K_{2\times2}} - \text{Cov}_s^{K_{10\times10}}$. Our tests will then ask if there is a significant difference in the average. The results reported in our web Appendix suggest very little difference in the level of these two estimators. When we compute the same test based on $d_s = \text{Cov}_s^{K_{2\times2}} - \text{Cov}_s^{5\text{m}}$ we get consistent rejection of no difference between these estimators — now the levels of the $\text{Cov}_s^{K_{2\times2}}$ is judged to be above the corresponding result for $\text{Cov}_s^{5\text{m}}$ — particularly for GE and SPY stocks. The same thing happens when $\text{Cov}_s^{K_{10\times10}}$ is compared to $\text{Cov}_s^{5\text{m}}$.

The result in that analysis is reinforced by the information in the summary Table 6, which shows results averaged over all asset pairs for both trades and quotes. The results are not very different for most estimators as we move from trades to quotes, the counter example is Cov_s^{HY} which seems sensitive to this.

The Table shows $\text{Cov}_s^{K_{2\times2}}$ and $\text{Cov}_s^{K_{10\times10}}$ have roughly the same average value, which is slightly below Cov_s^{OtoC}. $\text{Cov}_s^{K_{2\times2}}$ has a nine times smaller variance than Cov_s^{OtoC}, which shows it is a lot more precise. Of course integrated variance is its self random so nine underestimates the efficiency gain of using $\text{Cov}_s^{K_{2\times2}}$. If volatility is close to being persistent then $\text{Cov}_s^{K_{2\times2}}$ is at least $\frac{1.067^2}{0.334^2(1-acf_1)} \simeq 17$ times more informative than the cross product of daily returns. The same observation holds for mid-quotes.

$\text{Cov}_s^{15\text{s}}$ and Cov_s^{HY} are very precise estimates of the wrong quantity. $\text{Cov}_s^{5\text{m}}$ is quite close to $\text{Cov}_s^{K_{2\times2}}$, the two measures have a correlation of 0.92.

The corresponding results for correlations are less good. All the estimates are biased, which is no surprise due to it being a non-linear transform of roughly unbiased and somewhat noisy observations. $\text{Corr}_s^{K_{2\times2}}$ looks like the most effective estimate.

In our web appendix we give time series plots and autocorrelogram for the various estimates of realised covariance for the AA-SPY assets combination using trade data. They show $\text{Cov}_s^{K_{2\times2}}$ performing much better than the 30 minute realised covariance but there not being a great deal of difference between the statistics when the realised covariance is based on 5 minute returns. The web appendix also presents scatter plots of estimates based on transaction prices (vertical axis) against the same estimate based on mid-quotes (horizontal axis) for the same days. These show a remarkable agreement between estimates based on $\text{Cov}_s^{K_{2\times2}}$, $\text{Cov}_s^{5\text{m}}$ and $\text{Cov}_s^{30\text{m}}$, while once again Cov_s^{HY} struggles. Overall $\text{Cov}_s^{K_{2\times2}}$ and $\text{Cov}_s^{5\text{m}}$ behave in a similar manner, with $\text{Cov}_s^{K_{2\times2}}$ slightly stronger. $\text{Cov}_s^{K_{10\times10}}$ estimates roughly the same level as $\text{Cov}_s^{K_{2\times2}}$ but is discernibly noisier.

5.5 Analysis of the correlation estimates

In this subsection we will focus on estimating $\rho_s^{(i,j)} = [Y^{(i)}, Y^{(j)}]_s / \sqrt{[Y^{(i)}]_s [Y^{(j)}]_s}$ by the realised kernel correlation $\hat{\rho}_s^{(i,j)K} = K_s^{(i,j)} / \sqrt{K_s^{(i,i)} K_s^{(j,j)}}$ and the corresponding realised correlation $\hat{\rho}_s^{X\text{m}}$.

A table in our web Appendix average estimates for $\hat{\rho}_s^{K_{2\times2}}$, $\hat{\rho}_s^{K_{10\times10}}$ and $\hat{\rho}_s^{5\text{m}}$. It shows the expected result that $\hat{\rho}_s^{K_{2\times2}}$ is more precise than $\hat{\rho}_s^{K_{10\times10}}$. Both have average values which are quite a bit below the unconditional correlation of the daily open-to-close returns. This is not surprising. All the three

ingredients of the $\hat{\rho}_s^{K_{2\times2}}$ are measured with noise and so when we form $\hat{\rho}_s^{(i,j)K}$ it will be downward bias.

Table 6: Summary statistics across all asset pairs

Transaction prices

Estimator	Average HAC	Stdev	Bias	cor(.,K)	acf_1	acf_2	acf_3	acf_4	acf_5	acf_{10}
Summary stats for covariances										
$Cov^{K_{2\times2}}$	0.3180 [0.023]	0.334	-0.026	1.000	0.40	0.37	0.27	0.24	0.20	0.23
$Cov^{K_{10\times10}}$	0.3148 [0.026]	0.447	-0.029	0.787	0.23	0.20	0.16	0.16	0.11	0.13
Cov^{HY}	0.1026 [0.008]	0.099	-0.242	0.706	0.58	0.50	0.42	0.32	0.30	0.32
$Cov^{1/4m}$	0.1864 [0.013]	0.167	-0.158	0.764	0.60	0.52	0.41	0.33	0.29	0.28
Cov^{5m}	0.3082 [0.022]	0.334	-0.036	0.924	0.35	0.36	0.24	0.22	0.18	0.20
Cov^{30m}	0.2930 [0.025]	0.471	-0.051	0.646	0.15	0.11	0.10	0.10	0.10	0.10
Cov^{OtoC}	0.3435 [0.046]	1.067		0.288	0.03	0.01	0.02	0.03	0.02	0.05
Summary stats for correlations										
$Corr^{K_{2\times2}}$	0.3273 [0.010]	0.155		1.000	0.30	0.26	0.21	0.19	0.18	0.14
$Corr^{K_{10\times10}}$	0.3438 [0.013]	0.264		0.653	0.11	0.10	0.08	0.08	0.07	0.07
$Corr^{1/4m}$	0.1758 [0.007]	0.084		0.528	0.58	0.52	0.47	0.43	0.41	0.34
$Corr^{5m}$	0.3177 [0.010]	0.165		0.851	0.24	0.21	0.17	0.15	0.14	0.12
$Corr^{30m}$	0.3358 [0.015]	0.315		0.517	0.07	0.06	0.05	0.05	0.06	0.04
Average unconditional Open-to-Close correlation = 0.3974										

Mid-quotes

Estimator	Average HAC	Stdev	Bias	cor(.,K)	acf_1	acf_2	acf_3	acf_4	acf_5	acf_{10}
Summary stats for covariances										
$Cov^{K_{2\times2}}$	0.3183 [0.023]	0.347	-0.026	1.000	0.37	0.36	0.26	0.23	0.19	0.22
$Cov^{K_{10\times10}}$	0.3171 [0.026]	0.463	-0.027	0.767	0.19	0.17	0.14	0.15	0.10	0.13
Cov^{HY}	0.1628 [0.010]	0.136	-0.181	0.743	0.57	0.50	0.41	0.33	0.30	0.31
$Cov^{1/4m}$	0.1829 [0.013]	0.162	-0.161	0.733	0.62	0.53	0.42	0.34	0.30	0.29
Cov^{5m}	0.3080 [0.022]	0.333	-0.036	0.921	0.36	0.36	0.25	0.22	0.18	0.20
Cov^{30m}	0.2918 [0.024]	0.467	-0.052	0.668	0.16	0.12	0.10	0.10	0.10	0.10
Cov^{OtoC}	0.3447 [0.046]	1.067		0.299	0.03	0.02	0.02	0.03	0.02	0.05
Summary stats for correlations										
$Corr^{K_{2\times2}}$	0.3330 [0.010]	0.170		1.000	0.26	0.22	0.18	0.16	0.15	0.13
$Corr^{K_{10\times10}}$	0.3460 [0.014]	0.297		0.653	0.09	0.08	0.06	0.07	0.06	0.06
$Corr^{1/4m}$	0.1735 [0.006]	0.078		0.519	0.56	0.49	0.44	0.40	0.37	0.30
$Corr^{5m}$	0.3194 [0.010]	0.165		0.838	0.24	0.20	0.17	0.15	0.13	0.11
$Corr^{30m}$	0.3351 [0.015]	0.317		0.571	0.07	0.06	0.05	0.05	0.06	0.04
Average unconditional Open-to-Close correlation = 0.4035										

Summary statistics across all asset pairs. The first column identify the estimator, and the second gives the average estimate across all asset combinations, followed by the average Newey-West type standard error. The fourth gives the average standard deviation of the estimator. The fifth is the average bias. Next is average sample correlation with our realised kernel. The remaining columns give average autocorrelations. The upper panel is based on transaction prices, whereas the lower panel is based on mid-quotes. The sub panels give first results for covariance estimates followed by correlation results.

5.6 Analysis of the beta estimates

Here we will focus on estimating $\beta_s^{(i,j)} = [Y^{(i)}, Y^{(j)}]_s / [Y^{(j)}]_s$, by the realised kernel beta $\beta_s^{(i,j)K} = K_s^{(i,j)} / K_s^{(j,j)}$. Figure 2 presents scatter plots of beta estimates based on transaction prices (vertical axis) against the same estimate based on mid-quotes (horizontal axis). The two estimators are $\beta_s^{K_{2\times2}}$ to $\beta_s^{5\text{m}}$. The results are not very different in these two cases.

Figure 3 compares the fitted values from ARMA models for the kernel and 5 minute estimates of realised betas for the AA-SPY assets combination. These are based on the model estimates for the daily kernel based realised betas

$$\beta_s^K = \underset{(0.06)}{1.20} + \underset{(0.027)}{0.923}\beta_{s-1}^K + u_s - \underset{(0.048)}{0.726}u_{s-1}, \quad \text{adj}-R^2 = 0.213,$$

and for 5 minute based realised betas

$$\beta_s^{5\,\text{min}} = \underset{(0.06)}{1.16} + \underset{(0.024)}{0.950}\beta_{s-1}^{5\,\text{min}} + u_s - \underset{(0.039)}{0.821}u_{s-1}, \quad \text{adj}-R^2 = 0.145.$$

Both models have a significant memory, with autoregressive roots well above 0.9 and with large moving average roots. The fit of the realised kernel beta is a little bit better than that for the realised beta.

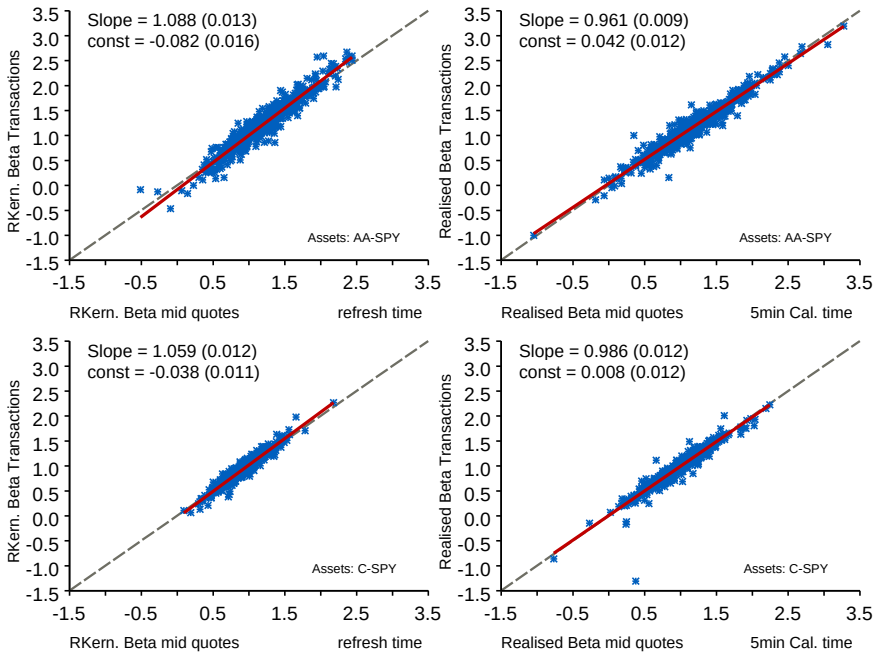

Figure 2: Scatter plots realised betas

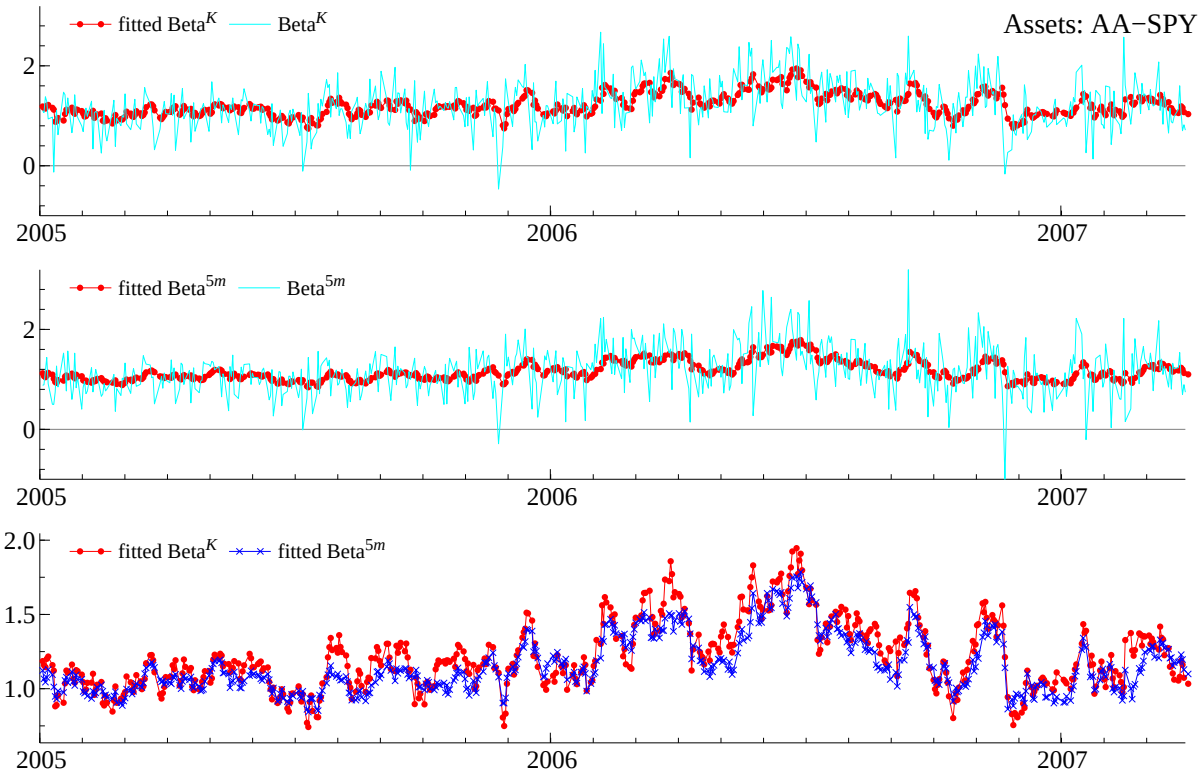

Figure 3: ARMA(1,1) model for transaction based betas

We also calculate the encompassing regressions. The estimates are for the realised kernel betas are

$$\beta_s^K = \underset{(0.031)}{0.084} + \underset{(0.053)}{0.858}\beta_{s-1}^K + \underset{(0.043)}{0.074}\beta_{s-1}^{5\,\text{min}} + u_s - \underset{(0.044)}{0.726}u_{s-1}, \quad \text{adj}-R^2 = 0.215,$$

with the corresponding 5 minute based realised betas

$$\beta_s^{5\,\text{min}} = \underset{(0.026)}{0.056} + \underset{(0.047)}{0.879}\beta_{s-1}^{5\,\text{min}} + \underset{(0.035)}{0.069}\beta_{s-1}^K + u_s - \underset{(0.040)}{0.822}u_{s-1}, \quad \text{adj}-R^2 = 0.150.$$

This shows that either estimator dominates the other in terms of encompassing, although the realised kernel has a slightly stronger t-statistic.

5.7 A scalar BEKK

5.7.1 Econometric framework

An important use of realised quantities is to forecast future volatilities and correlations of daily returns. The use of reduced form has been pioneered by Andersen et al. (2001) and Andersen et al. (2003). One useful way of thinking about the forecasting problem is to fit a GARCH type problem with lagged realised

quantities as explanatory variables, e.g. Engle & Gallo (2006). Here we follow this route, fitting a bivariate GARCH model with $\mathrm{E}(r_s|\mathcal{F}_{s-1}) = 0$, $\mathrm{Cov}(r_s|\mathcal{F}_{s-1}) = H_s$, where r_s is the 2×1 vector of daily close to close returns on the i-th and j-th asset, $\mathcal{F}_{s-1}$ is the information available at time $s-1$ to predict r_s. A standard Gaussian quasi-likelihood $-\frac{1}{2}\sum_{s=1}^{T}\left(\log|H_s| + r_s' H_s^{-1} r_s\right)$ is used to make inference. The model we fit is a variant on the scalar BEKK (e.g. Engle & Kroner (1995))

$$H_s = C'C + \alpha r_{s-1} r_{s-1}' + \beta H_{s-1} + \gamma \mathrm{K}_{s-1}^{K_{2\times2}}, \quad \alpha, \beta, \gamma \geq 0.$$

The question will be if γ is estimated to be statistically different from zero, for if it is not then the high frequency data enhances the forecast of future covariation.

5.7.2 Empirical results

Our results will be based on a relatively short time series of 2.5 years of daily measures, which is a challenging environment for GARCH type models.

The results in Table 7 suggest that lagged daily returns are no longer significant for this multivariate GARCH models once we have the realised kernel covariance. This is even though the realised kernel covariance misses out the overnight effect — the information in the close-to-open returns. An interesting feature of the series is that in most cases including $\mathrm{K}_{s-1}^{K_{2\times2}}$ reduces the size of the estimated H_{s-1} term.

6 Additional remarks

6.1 Relating $K(X)$ to the flat-top realised kernel $K^F(X)$

6.1.1 Positivity

In the univariate case the realised kernel

$$K(X) = \sum_{h=-n}^{n} k(\tfrac{h}{H+1})\Gamma_h, \quad H = c_0 n^{3/5}, \quad \Gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-h}$$

is at first sight very similar to the unbiased flat-top realised kernel of Barndorff-Nielsen et al. (2008*a*)

$$K^F(X) = \Gamma_0 + \sum_{h=1}^{n} k(\tfrac{h-1}{H+1})\left(\Gamma_h^F + \Gamma_{-h}^F\right), \quad H = d_0 n^{1/2}, \quad \Gamma_h^F = \sum_{j=1}^{n} x_j x_{j-h}.$$

Here the Γ_h and Γ_h^F are not divided by the sample size. This means that the end conditions, the observations at the start and end of the sample, can have influential effects on Γ_h. With Γ_h^F we removed this effect by starting the sum not at $h+1$ but at 1. However, an implication of this is that the resulting estimator is not guaranteed to be positive semi-definite whatever the choice of the weight function.

The alternative $K^F(X)$ has the advantage that it converges at a $n^{1/4}$ rate and is close to the parametric efficiency bound. It has the disadvantage that it can go negative, while we see in the next subsection that it is more sensitive to serial dependence in the noise and endogenous noise than $K(X)$.

Table 7: Scalar BEKK models for close-to-close bivariate returns

Series	H_{t-1}	$r_{t-1}^{CtoC}\left(r_{t-1}^{CtoC}\right)'$	$\mathrm{K}_{t-1}^{K_{2\times2}}$	$\mathrm{RV}_{t-1}^{5m_{2\times2}}$	$\log L$
AA-BA	0.6254 (0.1530)	0.0000 –	0.2409 (0.0912)	–	-1057.3
	0.6404 (0.1387)	0.0000 –	–	0.2222 (0.0843)	-1059.0
	0.8486 (0.1043)	0.0250 (0.0141)	–	–	-1070.4
	0.6254 (0.1528)	–	0.2409 (0.1703)	0.0000 –	-1057.3
AA-SPY	0.6259 (0.1556)	0.0000 –	0.1774 (0.0720)	–	-547.67
	0.6430 (0.1389)	0.0000 –	–	0.1606 (0.0665)	-549.11
	0.8746 (0.0459)	0.0239 (0.0099)	–	–	-555.70
	0.6259 (0.1583)	–	0.1774 (0.1873)	0.0000 –	-547.67
BA-SPY	0.6507 (0.1155)	0.0000 –	0.2975 (0.0938)	–	-435.82
	0.6392 (0.0968)	0.0000 –	–	0.3162 (0.0861)	-433.44
	0.8554 (0.0573)	0.0345 (0.0130)	–	–	-456.06
	0.6392 (0.0926)	–	0.0000 –	0.3161 (0.2327)	-433.44
GE-SPY	0.2831 (0.2114)	0.0063 (0.0249)	0.4555 (0.1273)	–	-91.093
	0.2847 (0.2197)	0.0115 (0.0250)	–	0.4426 (0.1231)	-91.502
	0.8468 (0.0690)	0.0368 (0.0145)	–	–	-104.04
	0.2872 (0.2012)	–	0.3122 (0.2689)	0.1624 (0.2662)	-90.937

Likelihood ratio summary

	Mean	Std	5%	25%	Median	75%	95%
$\mathrm{K}_{t-1}^{K_{2\times2}}$	24.94	16.41	2.068	11.98	23.24	34.49	53.39
$\mathrm{RV}_{t-1}^{5m_{2\times2}}$	22.23	15.88	3.254	10.80	17.13	29.00	57.17

Estimation results for scalar BEKK models for close-to-close bivariate returns.

There are three reasons that $K^F(X)$ can go negative. The most obvious is the use of a kernel function that does not satisfy, $\int_{-\infty}^{\infty} k(x)\exp(ix\lambda)\mathrm{d}x \geq 0$ for all $\lambda \in \mathbb{R}$, such as the Tukey-Hanning kernel or the cubic kernel, $k(x) = 1 - 3x^2 + 2x^3$. The flat-top kernels give unit weight to γ_1 and γ_{-1}, which can mean $K^F(X)$ may be negative. This can be verified by rewriting the estimator as a quadratic form estimator, $x'Mx$, where M is a symmetric band matrix $M = \text{band}(1, 1, k(\frac{1}{H}), k(\frac{2}{H}), \ldots,)$. The determinant of the upper-left matrix is given by $-\left\{k(\frac{1}{H}) - 1\right\}^2$, so that $k(\frac{1}{H}) = 1$ is needed to avoid negative eigenvalues. Repeating this argument leads to $k(\frac{h}{H}) = 1$ for all h, which violates the condition that $k(\frac{h}{H}) \to 0$, as $h \to \infty$. Finally, the third reason that the flat-top kernel could produce a negative estimate was due to the construction of realized autocovariances, $\gamma_h = \sum_{j=1}^{n} x_j x_{j-h}$. This requires the use of "out-of-period"

intraday returns, such as x_{1-H}. This formulation was chosen because it makes $\mathrm{E}\{K(U)\} = 0$ when $U \in \mathcal{WN}$. However, since x_{-H} only appears once in this estimator, with the term $x_1 x_{1-H}$, it is evident that a sufficiently large value of x_{1-H} (positive or negative, depending on the sign of x_1) will cause the estimator to be negative. We have overcome the last obstacle by jittering the end-points, which makes the use of "out-of-period" redundant. They can be dropped at the expense of a $O(m^{-1})$ bias.

6.1.2 Efficiency

An important question is how inefficient is $K(X)$ in practice compared to the flat-top realised kernel, $K^F(X)$? The answer is quite a bit when $U \in \mathcal{WN}$. Table 8 gives $\mathrm{E}\left[n^{1/4}\{K(X) - [Y]\}\right]^2/\omega$ and $\mathrm{E}\left[n^{1/4}\left\{K^F(X) - [Y]\right\}\right]^2/\omega$, the mean square normalised by the rate of convergence of $K_P^F(X)$ (which is the flat-top realised kernel using the Parzen weight function. An implication is that the scaled MSE for the $K(X)$ and K_B^F will increase without bound as $n \to \infty$ because these estimators converge at a rate that is slower than $n^{1/4}$). The results are given in the case of Brownian motion observed with different types of noise. Results for two flat-tops are given, the Bartlett ($K_B^F(X)$) and Parzen ($K_P^F(X)$) weight functions. Similar types of results hold for other weight functions.

Consider first the case with Gaussian $U \in \mathcal{WN}$ with variance of ω^2. The results show that the variance of $K(X)$ is much bigger than its squared bias. For small n there is not much difference between the three estimators, but by the time $n = 4,096$ (which is realistic for our applications) the flat-top $K^F(X)$ has roughly half the MSE of $K(X)$ in the univariate case. Hence in ideal circumstances $K^F(X)$ has advantages over $K(X)$, but we are attracted to the positivity and robustness of $K(X)$.

The robustness advantage of $K(X)$ can be seen from for the four simulation designs where U_j is modelled as a dependent process. We consider the moving average specification, $U_j = \epsilon_j - \theta\epsilon_{j-1}$, with $\theta = \pm 0.5$ and the autoregressive specification, $U_j = \varphi U_{j-1} + \epsilon_j$, with $\varphi = \pm 0.5$, where ϵ_j is Gaussian white noise. The bandwidth for all estimators were to be "optimal" under $U \in \mathcal{WN}$, which is the default in the literature, so $H_B^F = 2.28\omega^{4/3}n^{2/3}$, $H_P^F = 4.77\omega n^{1/2}$, and $H_P = 3.51\omega^{4/5}n^{3/5}$ where $\omega^2 = \sum_{h=-\infty}^{\infty} \mathrm{cov}(U_j, U_{j-h})$. The results show the robustness of $K(X)$ and the strong asymptotic bias of K_P^F and K_B^F under the non-white noise assumption. The specifications, $\theta = 0.5$ and $\varphi = -0.5$ induce a negative first-order autocorrelation while $\theta = -0.5$ and $\varphi = 0.5$ induce positive autocorrelation. Negative first-order autocorrelation can be the product of bid-ask bounce effects, this is particularly the case if sampling only occurs when the price changes. Positive first-order autocorrelation would, for example, be relevant for the noise in bid prices because variation in the bid-ask spread would induce such dependence.

Table 8: Relative efficiency of the realised kernel $K(X)$

	$\omega^2 = 0.001$								
	normalised bias^2			normalised variance			normalised mse		
n	$K_B^F(X)$	$K_P^F(X)$	$K(X)$	$K_B^F(X)$	$K_P^F(X)$	$K(X)$	$K_B^F(X)$	$K_P^F(X)$	$K(X)$
$U \in WN$									
250	0.0	0.0	0.8	16.2	16.3	18.0	16.2	16.3	18.8
1,000	0.0	0.0	2.5	11.7	12.1	16.9	11.7	12.1	19.4
4,000	0.0	0.0	3.1	10.4	10.4	19.0	10.4	10.4	22.1
16,000	0.0	0.0	4.6	10.5	9.5	20.8	10.5	9.5	25.4
$U_j = \epsilon_j + 0.5\epsilon_{j-1}$									
250	1.5	1.2	0.6	15.3	15.7	17.6	16.9	16.9	18.2
1,000	22.1	7.3	2.2	11.0	11.9	16.9	33.0	19.2	19.1
4,000	175.7	18.5	3.2	9.3	10.2	19.0	185.0	28.8	22.2
16,000	898.5	41.0	4.4	9.0	9.4	20.9	907.6	50.4	25.4
$U_j = \epsilon_j - 0.5\epsilon_{j-1}$									
250	122.7	96.9	3.9	27.5	24.2	18.3	150.2	121.1	22.2
1,000	1,769.1	588.0	6.1	44.8	20.4	16.9	1,813.9	608.3	23.0
4,000	14,195.1	1,490.4	5.0	73.1	13.9	19.3	14,268.2	1,504.4	24.3
16,000	72,797.6	3,326.8	5.5	88.6	10.9	20.8	72,886.2	3,337.7	26.3
$U_j = -0.5U_{j-1} + \epsilon_j$									
250	39.1	30.9	1.3	18.9	18.1	17.9	58.0	49.0	19.2
1,000	1,261.0	74.9	3.3	35.9	13.2	16.8	1,296.9	88.1	20.0
4,000	7,751.7	141.1	3.5	40.8	10.8	18.8	7,792.5	151.9	22.4
16,000	40,973.1	253.8	4.8	52.0	9.7	20.9	41,025.2	263.5	25.7
$U_j = 0.5U_{j-1} + \epsilon_j$									
250	0.5	0.4	0.3	14.8	15.3	17.7	15.3	15.7	18.0
1,000	9.6	6.3	1.5	9.8	10.8	16.6	19.4	17.1	18.2
4,000	96.0	39.6	2.7	8.5	9.7	19.1	104.4	49.2	21.8
16,000	505.8	141.5	4.2	8.5	9.2	21.1	514.3	150.7	25.3

Relative efficiency of the realised kernel $K(X)$ and $K^F(X)$ when estimating $[Y]$ where Y is standard Brownian motion with independent which observed with noise U with variance ω^2. MSE, Var and Bias^2 *are all scaled by $n^{1/2}/\omega$. In the special case with Gaussian white noise the asymptotic lower bound for the normalized mse is 8.00 (the normalized mse for $K_P^F(X)$ converges to 8.54 as $n \to \infty$ in this special case).*

6.2 Preaveraging without bias correction

6.2.1 Estimating multivariate QV

In independent and concurrent work Vetter (2008, p. 29 and Section 3.2.4) has studied a univariate suboptimal preaveraging estimator of $[Y]$ whose bias is sufficiently small that the estimator does not need to be explicitly bias corrected to be consistent (the bias corrected version can be negative). Its rate of convergence does not achieve the optimal $n^{-1/4}$ rate. Hence his suboptimal preaveraging estimator has some similarities to our non-negative realised kernel. Implicit in his work is that his non-corrected

preaveraging estimator is non-negative. However, this is not remarked upon explicitly nor developed into the multivariate case where non-synchronously spaced data is crucial.

Here we outline what a simple multivariate uncorrected preaveraging estimator based on refresh time would look like. We define it as $\hat{V} = \sum_{j=1}^{n-H} \overline{x}_j \overline{x}_j'$, where $\overline{x}_j = (\psi_2 H)^{-1/2} \sum_{h=1}^{H} g\left(\frac{h}{H}\right) x_{j+h}$, $\psi_2 = \int_0^1 g^2(u)\mathrm{d}u$. Here $g(u)$, $u \in [0,1]$ is a non-negative, continuously differentiable weight function, with the properties that $g(0) = g(1) = 0$ and $\psi_2 > 0$. Now if we set $H = \theta n^{3/5}$, then the univariate result in Vetter (2008) would suggest that $\hat{V}$ converges at rate $n^{-1/5}$, like the univariate version of our multivariate realised kernel. There is no simple guidance, even in the univariate, as to how to choose θ.

In the univariate bias corrected form, Jacod et al. (2007) show that $\hat{V}$ is asymptotically equivalent to using a $K^F(X)$ with $k(x) = \psi_2^{-1} \int_x^1 g(u)g(u-x)\mathrm{d}u$ and $H \propto n^{1/2}$. It is clear the same result will hold for the relationship between $\hat{V}$ and $K(X)$ in the multivariate case when $H = \theta n^{3/5}$. A natural choice of g is $g(x) = (1-x) \wedge x$, which delivers $\int_0^1 g^2(u)\mathrm{d}u = 1/12$ and a k function which is the Parzen weight function. Hence one might investigate using $\theta = c_0$ as in our paper, to drive the choice of H for $\hat{V}$ when applied to refresh time based high frequency returns.

Kinnebrock & Podolskij (2008*a*) have defined a bias corrected preaveraging estimator of the multivariate $[Y]$ with $H = \theta n^{1/2}$, for which they derive limit theory. To define their high frequency returns they use the Refresh Time idea — taken from an early draft of this paper. Their estimator has the disadvantage that it it is not guaranteed to be positive semi-definite.

6.2.2 Estimating integrated quarticity

In order to construct feasible confidence intervals for our realised quantities (see Barndorff-Nielsen & Shephard (2002)) we have to estimate Ψ. Our approach is based on the no-noise Barndorff-Nielsen & Shephard (2004) bipower type estimator applied to suboptimal preaveraged data taking $H = \theta n^{3/5}$. This is not an optimal estimator, it will converge at rate $n^{1/5}$, but it will be positive semidefinite. The proposed (positive semi-definite) estimator of $\mathrm{vec}(\Psi)$ is $\hat{Q} = n \sum_{j=1}^{n-H-1} \left\{ c_j c_j' - \frac{1}{2}\left(c_j c_{j+H}' + c_{j+H} c_j \right) \right\}$, where $c_j = \mathrm{vec}(\bar{x}_j \bar{x}_j')$. That the elements of $\hat{Q}$ is consistent using this choice of bandwidth is implicit in the thesis of Vetter (2008, p. 29 and Section 3.2.4).

6.3 The case with $[0, T]$

Throughout the paper we have discussed estimating QV over a unit interval, now we extend this to the interval $[0, T]$. Technically this is trivial, it is just a time-change argument. The results are that the QV target is $\int_0^T \Sigma(u)\mathrm{d}u$, while $\Psi = T \int_0^T \{\Sigma(u) \otimes \Sigma(u)\} \circ T(u)\mathrm{d}u$. Finally, $\xi_{ii} = \Omega_{ii} / \sqrt{T \int_0^T \Sigma_{ii}^2(u) \circ T(u)\mathrm{d}u}$.

6.4 Finite sample improvements

The realised kernel is non-negative so we can use log-transform to improve its finite sample performance. In particular

$$n^{1/5}\left\{\log\left(K(X)\right)-\log\left(\int_0^1\sigma^2(u)\mathrm{d}u\right)\right\}\xrightarrow{Ls}\mathrm{MN}\left\{\frac{\kappa}{\int_0^1\sigma^2(u)\mathrm{d}u},4\left(\frac{\kappa}{\int_0^1\sigma^2(u)\mathrm{d}u}\right)^2\right\}.$$

When the data is regularly spaced and the volatility is constant then $\kappa\sigma^{-2}=(\omega/\sigma)^{2/5}\left|k''(0)\right|^{1/5}\left(k_\bullet^{0.0}\right)^{2/5}$, which is slightly less dependent on σ^2 than the non-transformed version.

6.5 Subtlety of end effects

We have introduced jittering to eliminate end-effects. The larger is m the smaller is the end-effects, however increasing m has the drawback that is reduces the sample size, n, that can be used to compute the realised autocovariances. Given N observations, the sample size available after jittering is $n=N-2(m-1)$, so extensive jittering will increase the variance of the estimator. In this subsection we study this trade-off.

We focus on the univariate case where $U\in\mathcal{WN}$. The mean square error caused by end-effects is simply the squared bias plus the variance of $U_0U_0'+U_nU_n'$, which is given by $4m^{-2}\omega^4+4m^{-2}\omega^4=8\omega^4m^{-2}$, as shown in Appendix A, see the proof of Lemma A.4. The asymptotic variance (abstracting from end-effects) is $5\kappa^2n^{-2/5}=5\left|k''(0)\omega^2\right|^{2/5}\left\{k_\bullet^{0.0}\mathrm{IQ}\right\}^{4/5}n^{-2/5}$. So the trade-off between contributions from end-effects and asymptotic variance is given by

$$g_{N,\omega^2,\mathrm{IQ}}(m)=m^{-2}8\omega^4+5\left|k''(0)\omega^2\right|^{2/5}\left\{k_\bullet^{0.0}\mathrm{IQ}\right\}^{4/5}(N-m)^{-2/5}.$$

This function is plotted in Figure 4 for the case where $N=1,000$ and $\mathrm{IQ}=1$ and $\omega^2=0.0025$ and 0.001. The optimal value of m ranges from 1 to 2. The effect of increasing n on optimal m can be seen from Figure 4, where the optimal value of m has increased a little from Figure 4 as n has increased to $5,000$. However, the optimal amount of jittering is still rather modest.

6.6 Finite lag refresh time

In this paper we roughly synchronise our return data using the concept of Refresh Time. Refresh Time guarantees that our returns our not stale by more than one lag in Refresh Time. Our proofs need a somewhat less tight condition, that returns are not stale by more than a finite number of lags. This suggests it may be possible to find a different way of synchronising data which throws information away less readily than Refresh Time. We leave this problem to further research.

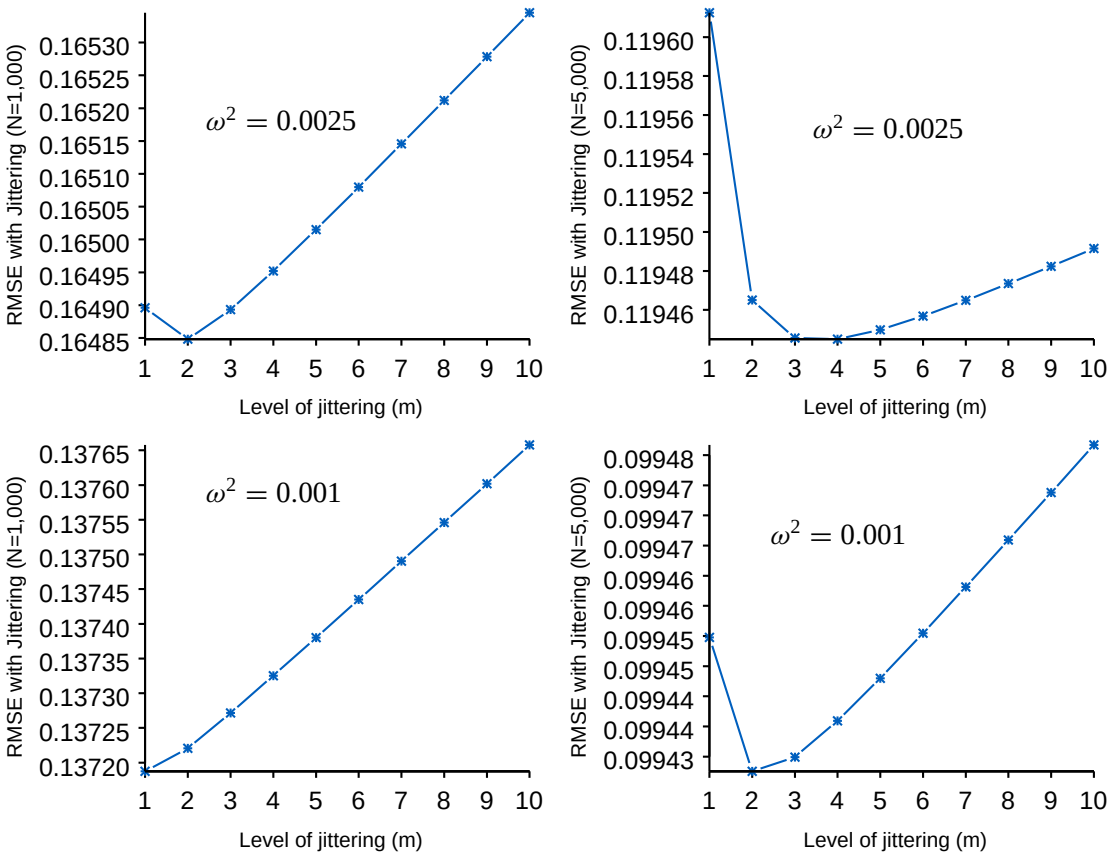

Figure 4: Sensitivity to the the choice of m. The Figure shows the RMSE as a function of m for the sample sizes $N = 1,000$ and $N = 5,000$, and $\omega^2 = 0.001$ and $\omega^2 = 0.0025$.

6.7 Jumps

In this paper we have assumed that Y is a pure $\mathcal{BSM}$. The analysis could be extended to the situation where Y is a pure $\mathcal{BSM}$ a finite activity jump process. The analysis in Barndorff-Nielsen et al. (2008*a*, section 5.6) suggests that the realised kernel is consistent for the quadratic variation, $[Y]$, at the same rate of convergence as before, but with a different asymptotic distribution.

7 Conclusions

In this paper we have proposed the multivariate realised kernel, which is a non-normalised HAC type estimator applied to high frequency financial returns, as an estimator of the ex-post variation of asset prices in the presence of noise and non-synchronous trading. The choice of kernel weight function is important here — for example the Bartlett weight function yields inconsistent estimation in our case.

Our analysis is based on three innovations: (i) we used a weight function which delivers biased kernels, however allowing us to use positive semi-definite estimators, (ii) we coordinate the collection

of data through the idea of refresh time, (iii) we show the estimator is robust to the remaining staleness in the data. Using this setup we are able to show consistency and asymptotic mixed Gaussianity of our estimator.

Our simulation study indicates our estimation procedure is close to being unbiased for covariances under realistic situations. Not surprisingly the estimators of correlations is downward biased due to the sampling variance of our estimators of variance. The empirical results based on our new estimator are striking, providing much sharper estimates of dependence amongst assets than has previously been available.

Multivariate realised kernels have potentially many areas of application, improving our ability to estimate covariances — allowing high frequency data to significantly improve our predictive models as well as better understand the pricing and management of risk in financial markets.

References

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2008), Parametric and nonparametric measurement of volatility, *in* Y. Aït-Sahalia & L. P. Hansen, eds, 'Handbook of Financial Econometrics', North Holland, Amsterdam. Forthcoming.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2000), 'Great realizations', *Risk* **13**, 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55. Correction published in 2003, volume 98, page 501.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2006), Market microstructure noise and realized volatility forecasting. Unpublished paper: Department of Economics, Duke University.

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

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

Bandi, F. M. & Russell, J. R. (2006), 'Seperating microstructure noise from volatility', *Journal of Financial Economics* **79**, 655–692.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008*a*), 'Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* . Forthcoming.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008*b*), 'Realised kernels in practice', *forthcoming in Econometrics Journal* .

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

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

Barndorff-Nielsen, O. E. & Shephard, N. (2007), Variation, jumps and high frequency data in financial econometrics, *in* R. Blundell, T. Persson & W. K. Newey, eds, 'Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress', Econometric Society Monographs, Cambridge University Press, pp. 328–372.

Bollerslev, T., Tauchen, G. & Zhou, H. (2008), Expected stock returns and variance risk premia. Unpublished paper: Department of Economics, Duke University.

Brownless, C. T. & Gallo, G. M. (2006), 'Financial econometric analysis at ultra-high frequency: Data handling concerns', *Computational Statistics & Data Analysis* **51**, 2232–2245.

Campbell, J. Y., Lo, A. W. & MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Dovonon, P., Goncalves, S. & Meddahi, N. (2007), Bootstrapping realized multivariate volatility measures. Unpublished paper: Tanaka Business School, Imperial College, London.

Drechsler, I. & Yaron, A. (2008), What's vol got to do with it. Unpublished paper: The Wharton School, University of Pennsylvania.

Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Engle, R. F. & Gallo, J. P. (2006), 'A multiple indicator model for volatility using intra daily data', *Journal of Econometrics* **131**, 3–27.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**, 122–150.

Epps, T. W. (1979), 'Comovements in stock prices in the very short run', *Journal of the American Statistical Association* **74**, 291–296.

Falkenberry, T. N. (2001), High frequency data filtering, Technical report, Tick Data.

Fisher, L. (1966), 'Some new stock-market indexes', *Journal of Business* **39**, 191–225.

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

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

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Springer-Verlag New York, Inc.

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

Guillaume, D. M., Dacorogna, M. M., Dave, R. R., Müller, U. A., Olsen, R. B. & Pictet, O. V. (1997), 'From the bird's eye view to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets', *Finance and Stochastics* **2**, 95–130.

Hansen, P. R. & Lunde, A. (2005), 'A realized variance for the whole day based on intermittent high-frequency data', *Journal of Financial Econometrics* **3**, 525–544.

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

Harris, F., McInish, T., Shoesmith, G. & Wood, R. (1995), 'Cointegration, error correction and price discovery on informationally-linked security markets', *Journal of Financial and Quantitative Analysis* **30**, 563–581.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Jacod, J. (2007), Statistics and high frequency data. Unpublished paper.

Jacod, J. (2008), A new type of limit theorems for discretized processes. Unpublished paper: Laboratoire de Probabilites, Universite Paris VI.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2007), Microstructure noise in the continuous case: the pre-averaging approach. Unpublished paper: Department of Statistics, University of Chicago.

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

Jacod, J. & Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, 2 edn, Berlin, Springer.

Kalnina, I. & Linton, O. (2008), 'Estimating quadratic variation consistently in the presence of correlated measurement error', *Journal of Econometrics* . Forthcoming.

Kinnebrock, S. & Podolskij, M. (2008*a*), An econometric analysis of modulated realised covariance, regression and correlation in noisy diffusion models. Unpublished paper: Oxford Financial Research Centre, University of Oxford.

Kinnebrock, S. & Podolskij, M. (2008*b*), 'A note on the central limit theorem for bipower variation of general functions', *Stochastic Processes and Their Applications* **118**, 1056–1070.

Large, J. (2007), Accounting for the Epps effect: Realized covariation, cointegration and common factors. Unpublished paper: Oxford-Man Institute, University of Oxford.

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

Martens, M. (2003), Estimating unbiased and precise realized covariances. Unpublished paper: Department of Finance, Erasmus School of Economics, Rotterdam.

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

Mykland, P. A. & Zhang, L. (2008), Inference for continuous semimartingales observed at high frequency: A general approach. Unpublished paper: Department of Statistics, University of Chicago.

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

Phillips, P. C. B. & Yu, J. (2008), Information loss in volatility measurement with flat price trading. Unpublished paper: Cowles Foundation for Research in Economics, Yale University.

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

Renault, E. & Werker, B. (2008), Causality effects in return volatility measures with random times. Unpublished paper: Department of Economics, University of North Carolina.

Reno, R. (2003), 'A closer look at the Epps effect', *International Journal of Theoretical and Applied Finance* **6**, 87–102.

Vetter, M. (2008), Estimation methods in noisy diffusion models. Unpublished Ph.D. thesis, Institute of Mathematics, Ruhr University Bochum.

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

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

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

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

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

Zhou, B. (1998), Parametric and nonparametric volatility measurement, *in* C. L. Dunis & B. Zhou, eds, 'Nonlinear Modelling of High Frequency Financial Time Series', John Wiley Sons Ltd, New York, chapter 6, pp. 109–123.

Appendices

Under the assumptions given in this paper, our line of argument will be as follows.

- Assume the data is synchronized and then a time-change allows us to think of the data as being regularly spaced. This is clear from the arguments in Barndorff-Nielsen et al. (2008*a*).
- Show the staleness left by the definition of refresh time has no impact on the asymptotic distribution of the equally spaced realised kernel. This is shown in Appendix B.
- Show the realised kernel is consistent and work out its limit theory for synchronized and equally spaced data. This is shown in Appendix A.

Appendix A: Proofs

Proof of Theorem 1. We note that for all i, j,

$$K\begin{pmatrix} Y^{(i)} \\ U^{(j)} \end{pmatrix} = \begin{pmatrix} K(Y^{(i)}) & K(Y^{(i)},U^{(j)}) \\ K(Y^{(i)},U^{(j)}) & K(U^{(j)}) \end{pmatrix} \geq 0,$$

which means that by taking the determinant of this matrix and rearranging we see that $K(Y^{(i)},U^{(j)})^2 \leq K(Y^{(i)})K(U^{(j)})$. Consequently, provided $K(Y) \xrightarrow{p} [Y]$,

$$\begin{aligned} K(X) - [Y] &= K(Y) - [Y] + O_p\left(\sqrt{\max_i K\left(Y^{(i)}\right)}\sqrt{\max_j K\left(U^{(j)}\right)}\right) + K(U) \\ &= K(Y) - [Y] + O_p\left(\sqrt{K(U)}\right). \end{aligned}$$

From this, together with the results of Lemmas A.1 and A.5, the conclusions of the Theorem follow directly. □

Proof of Proposition 1. This is a special case of Theorem 2. □

Proof of Proposition 2. The asymptotic variance of $K(Y)$ is given in Barndorff-Nielsen et al. (2008*a*) and the results concerning $K(U)$ follow from Lemma **??**, given below. □

Proof of Proposition 3. The problem is simply to minimize the squared bias plus the contribution from the asymptotic variance with respect to c_0. Set $\text{IQ} = \int_0^1 \sigma^4(u)\text{d}u$. The first order conditions of $\min_{c_0}\left\{-c_0^{-4}k''(0)^2\omega^4 + c_0 4k_\bullet^{0,0}\text{IQ}\right\}$ yield the optimal value for c_0

$$c_0^* = \left(\frac{k''(0)^2\omega^4}{k_\bullet^{0,0}\text{IQ}}\right)^{1/5} = c^*\xi^{4/5}, \qquad \text{with} \quad c^* = \left\{k''(0)^2/k_\bullet^{0,0}\right\}^{1/5}.$$

With $H^* = c^*\xi^{4/5}n^{3/5}$ the asymptotic bias is given by

$$-\left(\frac{k''(0)^2\omega^4}{k_\bullet^{0,0}\text{IQ}}\right)^{-2/5} k''(0)\omega^2 n^{-1/5} = \left|k''(0)\omega^2\right|^{1/5}\left\{k_\bullet^{0,0}\text{IQ}\right\}^{2/5} n^{-1/5},$$

and the asymptotic variance is

$$\left(\frac{k''(0)^2\omega^4}{k_\bullet^{0,0}\text{IQ}}\right)^{1/5} 4k_\bullet^{0,0}\text{IQ}n^{-2/5} = 4\left|k''(0)\omega^2\right|^{2/5}\left\{k_\bullet^{0,0}\text{IQ}\right\}^{4/5} n^{-2/5}.$$

□

A.1 Proof of Theorem 2.

We decompose the realised kernel into four terms that we will analyze separately.

$$K(X) - \int_0^1 \Sigma(u)\text{d}u = \left\{K(Y) - \int_0^1 \Sigma(u)\text{d}u\right\} + \text{E}\left\{K(U)\right\} + \left\{K(U) - \text{E}\left[K(U)\right]\right\} + \left\{K(Y,U) + K(U,Y)\right\}.$$

We start by deriving the asymptotic properties of $K(Y) - \int_0^1 \Sigma(u)\text{d}u$. The flat-top does not play a role in the asymptotic analysis of $K(Y)$, so the result for the univariate case follows from Barndorff-Nielsen et al. (2008*a*). The multivariate result is the following.

Lemma A.1 *$K(Y) = [Y] + O_p(\frac{H}{n})$, and with $H = c_0 n^{3/5}$, then*

$$n^{1/5}\left\{K(Y) - \int_0^1 \Sigma(u)\text{d}u\right\} \xrightarrow{Ls} \text{MN}\left\{0, 4c_0 k_\bullet^{0,0}\Psi\right\}.$$

Proof of Lemma A.1. First we note that, for any fixed $a \in \mathbb{R}^d$ the convergence of $a'\left\{K(Y) - \int_0^1 \Sigma(u)\text{d}u\right\}a$ follows directly by applying the univariate results in Barndorff-Nielsen et al. (2008*a*). Stable convergence for multivariate statistics, such as the realised autocovariances, Γ_h, are established in Kinnebrock & Podolskij (2008*b*), see also Jacod (2007). With $H \propto n^\gamma$, the consistency and stable convergence follow from Jacod (2008, theorems 2.1 and 2.2). What remains is to derive the asymptotic variance. For $a, b, c, d \in \mathbb{R}^d$ the asymptotic covariance between $a'K(Y)b$ and $c'K(Y)d$ is given as the plim of

$$c_0\frac{n}{H}\left(\sum_{h=-n+1}^{n-1} k(\tfrac{|h|}{H+1})\sum_i a'y_i y_{i-h}b\right)\left(\sum_{l=-n+1}^{n-1} k(\tfrac{|l|}{H+1})\sum_j c'y_j y_{j-l}d\right)$$
$$= c_0\frac{n}{H}\left(\sum_{\substack{h=l\\ i=j}} k(\tfrac{h}{H+1})^2 a'y_i y_i' cb'y_{i-h}y_{i-h}'d + \sum_{\substack{h=-l\\ i=j+h}} k(\tfrac{h}{H+1})^2 a'y_i y_i' db'y_{i-h}y_{i-h}'c\right) + O_p(\tfrac{H}{n})$$

$$\xrightarrow{p} c_0 \int_{-\infty}^{\infty} k(s)^2 \int_0^1 \left\{a'\Sigma(u)cb'\Sigma(u)d + a'\Sigma(u)db'\Sigma(u)c\right\} \mathrm{d}u\mathrm{d}s$$
$$= \mathrm{vec}(\tfrac{ab'+ba'}{2})' \left[4c_0 k_\bullet^{0,0} \int_0^1 \{\Sigma(u) \otimes \Sigma(u)\}\right] \mathrm{d}u \mathrm{vec}(\tfrac{cd'+dc'}{2}).$$

Here we used $\int_{-\infty}^{\infty} k(s)^2 \mathrm{d}s = 2k_\bullet^{0,0}$ and

$$a'\Sigma cb'\Sigma d + a'\Sigma db'\Sigma c = c'\Sigma ab'\Sigma d + d'\Sigma ab'\Sigma c = \mathrm{tr}\left\{\Sigma ab'\Sigma dc' + \Sigma ab'\Sigma cd'\right\}$$
$$= \mathrm{vec}(ab')' (\Sigma \otimes \Sigma) \mathrm{vec}(dc' + cd') = \tfrac{1}{2}\mathrm{vec}(ab' + ba')' (\Sigma \otimes \Sigma) \mathrm{vec}(dc' + cd').$$

□

A.1.1 Results concerning $K(U)$

We derive the asymptotic properties of $K(U)$ under the assumption that $U \in \mathcal{CS}$. The following definitions lead to a useful representation of $K(U)$. For $h = 0, 1, \ldots$ we define

$$V_h = \sum_{j=h+1}^{n-1} U_j U_{j-h}' + U_{j-h}U_j', \quad \text{and} \qquad Z_h = \left(U_0 U_h' + U_h U_0'\right) + (U_n U_{n-h}' + U_{n-h}U_n').$$

Lemma A.2 *The realised autocovariances of U can be written as*

$$\Gamma_0(U) = V_0 - V_1 + \tfrac{1}{2}Z_0 - Z_1 \tag{A.1}$$
$$\Gamma_h(U) + \Gamma_h(U)' = -V_{h-1} + 2V_h - V_{h+1} + Z_h - Z_{h+1}, \tag{A.2}$$

Proof. The first expression, (A.1), follows from

$$\Gamma_0(U) = \sum_{j=1}^{n}(U_j - U_{j-1})(U_j - U_{j-1})' = \sum_{j=1}^{n-1}(U_j U_j' + U_j U_j') + U_n U_n' + U_0 U_0'$$
$$- \sum_{j=2}^{n-1}(U_j U_{j-1}' + U_{j-1}U_j') - (U_n U_{n-1}' + U_{n-1}U_n' + U_0 U_1' + U_1 U_0'),$$

and (A.2) is proven similarly. □

Lemma A.3 *The realised kernel for U has the exact representation:*

$$K(U) = \left\{k(0) - k(\tfrac{1}{H+1})\right\} V_0 - \sum_{h=1}^{n-1} \left\{k(\tfrac{h+1}{H+1}) - 2k(\tfrac{h}{H+1}) + k(\tfrac{h-1}{H+1})\right\} V_h \tag{A.3}$$
$$+ \tfrac{1}{2}Z_0 - \sum_{h=1}^{n-1} \left\{k(\tfrac{h}{H+1}) - k(\tfrac{h-1}{H+1})\right\} Z_h.$$

Proof. Follows from the definition of $K(U)$ and Lemma A.2. □

Now we prove the result concerning the end-effects. We note that U_0 and U_n are absent from V_l, for all $l = 0, 1, \ldots$, so end-effects can only have an impact on $K(U)$ through Z_h, $h = 0, 1, \ldots$.

Lemma A.4 $\mathrm{E}(Z_0) = 4m^{-1}\sum_{h=-m+1}^{m-1} \frac{m-|h|}{m}\Omega_h$ *and* $\mathrm{E}(Z_h) = 2m^{-1}\sum_{j=h}^{m+h-1}(\Omega_j + \Omega_j')$.

Proof of Lemma A.4. Recall that $U_0 = m^{-1}\sum_{j=0}^{m-1} U(t_j)$. So that $\mathrm{E}(U_0) = 0$ and

$$\mathrm{E}(U_0 U_0') = m^{-2}\sum_{j=0}^{m-1}\sum_{i=0}^{m-1}\mathrm{E}\{U(t_j)U(t_i)'\} = m^{-1}\sum_{h=-m+1}^{m-1}\tfrac{m-|h|}{m}\Omega_h,$$

and similar for $\mathrm{E}(U_n U_n')$. So the first result follows from $\mathrm{E}\,(Z_0) = 2\{\mathrm{E}(U_0U_0') + \mathrm{E}(U_nU_n')\}$. Next, for $h > 0$,

$$\mathrm{E}(U_0 U_h') = m^{-1}\sum_{j=0}^{m-1}\mathrm{E}\{U(t_j)U(t_{m-1+h})'\} = m^{-1}\sum_{j=0}^{m-1}\Omega_{-j-h} = m^{-1}\sum_{j=0}^{m-1}\Omega_{j+h}',$$

and similarly we find $\mathrm{E}(U_n U_{n-h}') = m^{-1}\sum_{j=0}^{m-1}\Omega_{j+h}$, and the second and last result follows. □

It is a simple matter to compute the bias of the realised kernels caused by the noise.

Lemma A.5 *For large n, H and m we have.*

$$\mathrm{E}\{K(U)\} = -k''(0)\frac{n}{H^2}\Omega + O(m^{-1}) + o(\frac{n}{H^2})$$

and the asymptotic variance of $K(U)$ is $o(nH^{-3+\epsilon}) + O(H^{-1}m^{-1})$ for any $\epsilon > 0$.

Remark. So with $H \propto n^{3/5}$ and $m \to \infty$, we note that $\mathrm{Var}\{K(U)\} = o(n^{-3/5})$. This implies that $n^{1/5}[K(U) - \mathrm{E}\{K(U)\}] = o_p(1)$, so $K(U)$ only contributes to the bias term in the asymptotic distribution, not to the asymptotic variance when $H \propto n^{3/5}$.

Proof. First we prove the result concerning the bias. We have,

$$k(0) - k\left(\tfrac{1}{H+1}\right) = -\tfrac{k''(\varepsilon_H)}{2(H+1)^2}, \qquad \text{for some } 0 \le \varepsilon_H \le \tfrac{1}{H+1},$$

since $k'(0) = 0$. Now we define $a_0 = -k''(\varepsilon_H)$ and $a_h = H^2\{-k(\frac{|h|+1}{H+1}) + 2k(\frac{|h|}{H+1}) - k(\frac{|h|-1}{H+1})\}$, then

$$\begin{aligned}
&\left\{k(0) - k(\tfrac{1}{H+1})\right\}\mathrm{E}(V_0) - \sum_{h=1}^{n-1}\left\{k(\tfrac{h+1}{H+1}) - 2k(\tfrac{h}{H+1}) + k(\tfrac{h-1}{H+1})\right\}\mathrm{E}(V_h)\\
&= H^{-2}\sum_{h=-n+1}^{n-1} a_h(n-1-h)\mathrm{E}(U_j U_{j-h}') = H^{-2}\sum_{h=-n+1}^{n-1} a_h(n-1-h)\Omega_h\\
&= H^{-2}\sum_{|h|\le\sqrt{H}} a_h(n-1-h)\Omega_h + H^{-2}\sum_{|h|>\sqrt{H}} a_h(n-1-h)\Omega_h.
\end{aligned}$$

By the continuity of $k''(x)$ it follows that

$$\sup_{|h|\le\sqrt{H}}\left|\frac{H^2}{n}a_h(n-1-h) + k''(0)\right| \to 0, \qquad \text{as } H, n \to \infty \qquad \text{with } H/n = o(1),$$

so that the first term

$$H^{-2}\sum_{|h|\le\sqrt{H}} a_h(n-1-h)\Omega_h = -k''(0)\frac{n}{H^2}\Omega + o(\frac{n}{H^2}).$$

For the second term

$$H^{-2}\left|\sum_{|h|>\sqrt{H}} a_h(n-1-h)\Omega_h\right| \leq \frac{n}{H^2}\sum_{|h|>\sqrt{H}}|a_h\Omega_h| \leq \sup_{|h|\geq\sqrt{H}}|\Omega_h|\,\frac{n}{H^2}\sum_{|h|>\sqrt{H}}\left|H^2 a_h\right|,$$

which vanishes exponentially fast. For the Z-terms we have by Lemma A.4 that $\mathrm{E}(Z_0) = O(m^{-1})$, and

$$\begin{aligned}
\mathrm{E}\sum_{h=1}^{n-1}\left\{k(\tfrac{h}{H+1}) - k(\tfrac{h-1}{H+1})\right\}Z_h &= \sum_{h=1}^{n-1}\left\{k(\tfrac{h}{H+1}) - k(\tfrac{h-1}{H+1})\right\}2m^{-1}\sum_{j=1}^{m-1}(\Omega_{j+h}+\Omega'_{j+h}) \\
&\leq \sum_{h=1}^{n-1}\left\{k(\tfrac{h}{H+1}) - k(\tfrac{h-1}{H+1})\right\}2m^{-1}\sum_{j=1}^{\infty}|\Omega_{j+h}+\Omega'_{j+h}| \\
&\simeq m^{-1}2\int\left\{k'(\tfrac{h}{H+1}) + O(H^{-1})\right\}\sum_{j=1}^{\infty} j\,|\Omega_j+\Omega'_j| = O(m^{-1}).
\end{aligned}$$

Next, we turn to the asymptotic variance of $K(U)$. Consider for simplicity the univariate case. First we choose some $\eta > 0$. We see from expression (A.3) that the contribution from terms involving V_h, $h = 0, 1, \ldots$ has a total variance

$$\mathrm{var}(H^{-2}\sum_h a_h\sum_i 2U_iU_{i-h}) \leq 4H^{-4}\sum_{\substack{|i-j|<H^\eta\\|h-l|<H^\eta}} a_ha_l\mathrm{E}(U_iU_{i-h}U_jU_{j-l}) + 4n^4\sup_{M\geq H^\eta}\varrho_M,$$

Assumption 1 guarantees that the last term vanishes at exponential rate. For the first term we have

$$\begin{aligned}
4H^{-4}\sum_{\substack{|i-j|<H^\eta\\|h-l|<H^\eta}} a_ha_l\mathrm{E}(U_iU_{i-h}U_jU_{j-l}) &\leq 4H^{-4}2H^\eta n\mathrm{E}(U_i^4)\sum_{|h-l|<H^\eta} a_ha_l \\
&\leq 8H^{-3+\eta}n\mathrm{E}(U_i^4)2H^\eta\left\{\int k''(s)^2\mathrm{d}s + o(1)\right\} = O(nH^{-3+2\eta})\mathrm{E}(U_i^4).
\end{aligned}$$

The other terms involving Z_h, $h \geq 0$ have the form

$$\mathrm{var}\left[\sum_{h=1}^{n-1}\left\{k(\tfrac{h}{H+1}) - k(\tfrac{h-1}{H+1})\right\}U_0U_h\right] \leq H^{-2}\sum_{i=1}^{n}\left\{k'(\tfrac{h}{H+1}) + O(H^{-1})\right\}^2\mathrm{E}(U_0^2U_h^2) = O(H^{-1}m^{-1}).$$

□

A.1.2 Results concerning $K(Y,U)$ and $K(U,Y)$

Lemma A.6 *$K(Y,U) = O_p(H^{-1/2})$ so that $K(Y,U) = o_p(n^{-1/5})$ when $H \propto n^{3/5}$.*

Proof. For simplicity we prove the result for the case $d = 1$. We have

$$K(Y,U) = \sum_h k(\tfrac{h}{H+1})\sum_j y_j\left(U_{j-h} - U_{j-h-1}\right) \simeq \sum_h k(\tfrac{h}{H+1})\{W_h - W_{h-1}\},$$

where $W_h = \sum_{j=h+1}^{n} y_jU_{j-h}$, so that $K(Y,U) \simeq \sum_{h=-n+1}^{n-1} H^{-1}k'(\tfrac{|h|}{H+1})W_h$. Write $\mathrm{E}_Y(\cdot) = \mathrm{E}(\cdot|\{Y\})$. Then

$$\mathrm{E}_Y\left(\sum_{i=h+1}^{n}\sum_{j=l+1}^{n} y_iU_{i-h}y_jU_{j-l}\right) = \sum_{i=j} y_iy_j\mathrm{E}_Y(U_{i-h}U_{i-l}) + \sum_{i\neq j} y_iy_j\mathrm{E}_Y(U_{i-h}U_{j-l})$$

$$= \sum_{i=j} y_i y_j \Omega_{l-h} + \sum_{q\neq 0} \Omega_{q+l-h} \sum_i y_i y_{i-q}.$$

The first term converges to $\Omega_{h-l}\int_0^1 \sigma^2(u)\mathrm{d}u$ in probability while the second term vanishes. So that

$$\sum_{h=-n+1}^{n-1} H^{-1}k'(\tfrac{|h|}{H+1})W_h \sum_{l=-n+1}^{n-1} H^{-1}k'(\tfrac{|l|}{H+1})W_l \simeq H^{-2}\sum_{h,l} k'(\tfrac{|h|}{H+1})k'(\tfrac{|l|}{H+1})\Omega_{h-l}\int_0^1 \sigma^2(u)\mathrm{d}u$$
$$\simeq H^{-2}\sum_{\substack{h,l \\ |h-l|<H^\gamma}} k'(\tfrac{|h|}{H+1})k'(\tfrac{|l|}{H+1})\Omega_{h-l}\int_0^1 \sigma^2(u)\mathrm{d}u$$
$$\simeq H^{-2}\sum_h \left\{k'(\tfrac{|h|}{H+1})\right\}^2 \sum_{\substack{l \\ |h-l|<H^\gamma}} \Omega_{h-l}\int_0^1 \sigma^2(u)\mathrm{d}u \simeq H^{-1}\Omega\int_0^1 \sigma^2(u)\mathrm{d}u\int_{-\infty}^{\infty}\{k'(u)\}^2\mathrm{d}u.$$

Thus $K(Y,U) = O_p(H^{-1/2})$. With $H = n^{3/5}$ we have $n^{1/5}K(Y,U) = O_p(n^{-1/10})$ such that this term does not contribute to the asymptotic distribution when $H^{3/5}$. □

Appendix B: Errors induced by stale prices

The stale prices induce a particular form of noise with an endogenous component. The assumptions that we made about the noise were formulated for prices sampled with the Refresh Time. It may be more natural to formulate assumptions for the noise that is tied to actual observation times rather than the artificial refresh times. Here we show that the limit distribution for $K(X)$ is the same under both assumptions.

The price indexed by time τ_j is, in fact, the price recorded at time $t_j^{(i)} \leq \tau_j$, for $i = 1,\ldots,d$. With Refresh Time we have $\tau_j \geq t_j^{(i)} > \tau_{j-1}$ so that

$$X^{(i)}(\tau_j) = Y^{(i)}(t_j^{(i)}) + U^{(i)}(t_j^{(i)}) = Y^{(i)}(\tau_j) + \underbrace{U^{(i)}(t_j^{(i)}) - \{Y^{(i)}(\tau_j) - Y(t_j^{(i)})\}}_{\tilde{U}^{(i)}(\tau_j)}.$$

This shows that if the dependence in $\{U\}$ was specific to $U^{(i)}(t_j^{(i)})$, rather than $U^{(i)}(\tau_j)$, then the actual measurement error, $\tilde{U}^{(i)}(\tau_j) = X^{(i)}(\tau_j) - Y^{(i)}(t_j^{(i)})$, has an endogenous component given by $Z_j^{(i)} = Y^{(i)}(\tau_j) - Y^{(i)}(t_j^{(i)})$. The implication is that $K(X) = K(Y+U+Z)$, when we define the noise using the observation times $U_j^{(i)} = U^{(i)}(t_j^{(i)})$ for $i = 1,\ldots,d$.

Theorem B.1 *Suppose that $H \propto n^{3/5}$ and that Assumption 1 holds for $U_j = \left\{U^{(1)}(t_j^{(1)}),\ldots,U^{(d)}(t_j^{(d)})\right\}'$. Then $K(Y+U+Z) - K(Y+U) = o_p(n^{1/5})$.*

The implication is that the asymptotic distribution is unaffected by Refresh Time.

Proof. We prove the result by showing that $K(Z)$, $K(Y,Z)$, $K(U,Z)$ are all $o_p(n^{-1/5})$. First we note that Z_j, $j = 1,\ldots,n$ are increments in Y, computed over non-overlapping intervals. So $\{Z_j\}$ is effectively a heteroskedastic independent process, where $\mathrm{E}(Z_j^2)$ is bounded by a term that is of order

$$\delta_n = \sup_j \int_{\tau_{j-1}}^{\tau_j} \sigma^2(s)\mathrm{d}s = O(n^{-1}).$$

So we can apply our analysis of $K(U)$, and analogous to Lemma A.5 we find that

$$\mathrm{E}\{K(Z)\} = \frac{n}{H^2}\delta_n = O(n^{-6/5}),$$

and the asymptotic variance of $K(Z)$ is $o(nH^{-2})\delta_n^2 = o(n^{-11/5})$. Using the same argument as in the proof of Theorem 1, we have $K(Y, Z)^2 \leq K(Y)K(Z)$ and $K(U, Z)^2 \leq K(Y)K(Z)$, which proves that both $K(Y, Z)$ and $K(U, Z)$ are $o_p(n^{-1/5})$, as was needed. □