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

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

# Accepted Manuscript


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

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

PII: S0304-4076(11)00002-9
DOI: 10.1016/j.jeconom.2010.07.009
Reference: ECONOM 3440

To appear in: *Journal of Econometrics*

Received date: 14 July 2010
Revised date: 14 July 2010
Accepted date: 28 July 2010

Please cite this article as: Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* (2011), doi:10.1016/j.jeconom.2010.07.009

# 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†
*School of Economics and Management, Aarhus University,*
*Bartholins Allé 10, DK-8000 Aarhus C, Denmark*
*& CREATES, University of Aarhus*
alunde@asb.dk

NEIL SHEPHARD
*Oxford-Man Institute, University of Oxford,*
*Eagle House, Walton Well Road, Oxford OX2 6ED, UK*
*& Department of Economics, University of Oxford*
neil.shephard@economics.ox.ac.uk

Version: July 14, 2010; First circulated in July 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 error 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 that 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 (2006) was used to perform the calculations reported here. We thank Torben Andersen, Tim Bollerslev, Ron Gallant, Xin Huang, Oliver Linton and Kevin Sheppard for comments on a previous version of this paper.

†Corresponding author

## 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 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, $\sigma$ 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 (psd) 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 (2010) 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 papers by Drechsler & Yaron (2008), Fleming, Kirby & Ostdiek (2003) and de Pooter, Martens & van Dijk (n.d.).

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 (2008), Zhang, Mykland & Aït-Sahalia (2005), Hansen & Lunde (2006), Hansen, Large & Lunde (2008), Kalnina & Linton (2008), Zhang (2006), Barndorff-Nielsen, Hansen, Lunde & Shephard (2008), Renault & Werker (2010), Hansen & Horel (2009), Jacod, Li, Mykland, Podolskij & Vetter (2009) and Andersen, Bollerslev & Meddahi (2010). 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), Goncalves & Meddahi (2009), Mykland & Zhang (2009*b*) 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), Bandi & Russell (2005), Voev & Lunde (2007), Griffin & Oomen (2009) and Large (2007). We will draw ideas from this work.

Our estimator, the *multivariate realised kernel*, differs from the univariate realised kernel estimator by Barndorff-Nielsen et al. (2008) in important ways. The latter converges a rate $n^{1/4}$ but critically relies on the assumption that the noise is a white noise process, and Barndorff-Nielsen et al. (2008) stress that their estimator cannot be applied to tick-by-tick data. In order not to be in obvious violation of the iid assumption, Barndorff-Nielsen et al. (2008) apply their estimator to prices that are (on average) sampled every minute or so. Here, in the present paper, we allow for a general form of noise that is consistent with the empirical features of tick-by-tick data. For this reason we adopt a larger bandwidth that has the implication that our multivariate realized kernel estimator converges at rate $n^{1/5}$. Although this rate is slower than $n^{1/4}$ it is, from a practical viewpoint, important to acknowledge that there are only 390 one-minute returns in a typical trading day, while many shares trade several thousand times, and $390^{1/4} < 2000^{1/5}$. So the rates of convergence will not (alone) tell us which estimators will be most accurate in practice – even for univariate estimation problem. In addition to being robust to noise with a general form of dependence, the $n^{1/5}$ convergence rate enables us to construct an estimator that is guaranteed to psd, which is not the case for the estimator by Barndorff-Nielsen et al. (2008). Moreover, our analysis of irregularly spaced and non-synchronous observations cause the asymptotic distribution of our estimator to be quite different from that in Barndorff-Nielsen et al. (2008). We discuss the differences between these estimators in greater details in Section 6.1.

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. We analyse up to 30 dimensional covariance matrices, and demonstrate efficiency gains that are around 20 fold compared to using daily data. 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://mit.econ.au.dk/vip_htm/alunde/BNHLS/BNHLS.htm`.

## 2. Defining the multivariate realised kernel

### 2.1. Synchronizing data: 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 induce 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 $t_1^{(i)}, t_2^{(i)}, \ldots$. We now define *refresh time* which will be key to the construction of multivariate realised kernels. This time scale was used in a cointegration study of price discovery by Harris, McInish, Shoesmith & Wood (1995), and Martens (2003) has used the same idea in the context of realised covariances.

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

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

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

The $\tau_1$ is the time it has taken for all the assets to trade, i.e. all their posted price have been updated. $\tau_2$ 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[1].

[1] 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 6.5.

- Show these stale pricing errors have no impact on the asymptotic distribution of the realised kernels.

[Figure 1 about here.]

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 $n^{(1)}, n^{(2)}, \ldots, n^{(d)}$. 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, $p = 21/27 \simeq 0.78$.

### 2.2. **Jittering end conditions**

It turns out that our asymptotic theory dictates we need to average $m$ prices at the very beginning and end of the day to obtain a consistent estimator.[2] The theory behind this will be explained in Section 6.4, where experimentation suggests the best choice for $m$ is around two for the kind of data we see in this paper. Now we define what we mean by jittering. Let $n, m \in \mathbb{N}$, with $n - 1 + 2m = N$, then set the vector observations $X_0, X_1, \ldots, X_n$ as $X_j = X(\tau_{N,j+m})$, $j = 1, 2, \ldots, n-1$, and

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

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. These prices allow us to define the high frequency vector returns: $x_j = X_j - X_{j-1}$, $j = 1, 2, \ldots, n$, that the realised kernels are built out of.

### 2.3. **Realised kernel**

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

$$K(X) = \sum_{h=-n}^{n} k(\tfrac{h}{H})\Gamma_h, \qquad \text{where } \Gamma_h = \sum_{j=h+1}^{n} x_j x_{j-h}', \text{ for } h \geq 0, \tag{1}$$

and $\Gamma_h = \Gamma_{-h}'$ for $h < 0$. Here $\Gamma_h$ is the $h$-th realised autocovariance and $k : \mathbb{R} \curvearrowright \mathbb{R}$ is a non-stochastic weight function. We focus on the class of functions, $\mathcal{K}$, that is characterized by:

[2] This kind of averaging appears in, for example, Jacod et al. (2009).

**Assumption K.** (i) $k(0) = 1$, $k'(0) = 0$; (ii) $k$ is twice differentiable with continuous derivatives; (iii) define $k_\bullet^{0,0} = \int_0^\infty k(x)^2 \mathrm{d}x$, $k_\bullet^{1,1} = \int_0^\infty k'(x)^2 \mathrm{d}x$, and $k_\bullet^{2,2} = \int_0^\infty k''(x)^2 \mathrm{d}x$ then $k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2} < \infty$; (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 $\Gamma_0$ gets unit weight, while $k'(0) = 0$ means the kernel gives close to unit weight to $\Gamma_h$ 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 define the realised autocovariances are not divided by the sample size and $k'(0) = 0$ 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.

In some of our results we use the following additional assumption on the Brownian semimartingale.

**Assumption SH.** *Assume* $\mu$ *and* $\sigma$ are *bounded and we will write* $\sigma_+ = \sup_{t\in[0,1]} |\sigma(t)|$.

This can be relaxed to locally bounded if $(\mu, \sigma)$ is an Ito process — e.g. Mykland & Zhang (2009*a*).

## 2.4. **Some assumptions about refresh time and noise**

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

### 2.4.1. *Assumptions about the refresh time*

We use the subscript-$N$ to make the dependence on $N$ explicit. Note that $N$ is random and we write the durations as $\tau_{N,i} - \tau_{N,i-1} = \Delta_{N,i} = \frac{D_{N,i}}{N}$ for all $i$.

We make the following assumptions about the durations between observation times.

**Assumption D.** (i) *That* $\mathrm{E}(D^r_{N,\lfloor tN\rfloor}|\mathcal{F}_{\tau_{N,\lfloor tN\rfloor -1}}) \overset{p}{\to} \varkappa_r(t)$, $0 < r \leq 2$, *as* $N \to \infty$. *Here we assume* $\varkappa_r(t)$ *are strictly positive càdlàg processes adapted to* $\{\mathcal{F}_t\}$; (ii) $\max_{i\in\{j+1,\dots,j+R\}} D_{N,i} = o_p(R^{1/2})$ for any $j$; (iii) $\tau_{N,0} \leq 0$ *and* $\tau_{N,N+1} \geq 1$.

**Remark 1.** If we have Poisson sampling the durations are exponential and the $\max_{i\in\{1,\dots,N\}} \Delta_{N,i} = O_p(\log(N)/N)$, so $\max_{i\in\{1,...,N\}} D_{N,i} = O_p(\log(N))$. Note both Barndorff-Nielsen et al. (2008) and Mykland & Zhang (2006) assume that $\max_{i\in\{1,...,N\}} D_{N,i} = O_p(1)$. Phillips & Yu (2008) provided a novel analysis of realised volatility under random times of trades. We use their assumptions **D** here, applied to the realised kernel. Deriving results for realised volatility under random times of trades is an active research area.[3]

**Example 1.** ***Refresh time.*** *If each individual series has trade times which arrive as independent Poisson processes with the same intensity* $\lambda N$*, then their scaled durations are* $D^{(j)} \overset{i.i.d.}{\sim} \exp(\lambda)$, $j = 1, 2, ..., d$*, so the refresh time durations are* $D_{N,i} \overset{L}{=} \max\left\{D^{(1)}, ..., D^{(d)}\right\}$*, and so (e.g. Embrechts, Klüppelberg & Mikosch (1997, p. 189)) the refresh times have the form of a renewal process* $\tau_{N,i} - \tau_{N,i-1} = \frac{1}{N} D_{N,i}$, $D_{N,i} \overset{L}{=} \sum_{j=1}^{d} \frac{1}{j} D^{(j)}$*. In particular* $\varkappa_1(t) = \lambda^{-1} \sum_{j=1}^{d} j^{-1}$, $\varkappa_2(t) = \lambda^{-2} \left\{\sum_{j=1}^{d} j^{-2} + \left(\sum_{j=1}^{d} j^{-1}\right)^2\right\}$*. Of interest is how these terms change as* $d$ *increases. The former is the harmonic series and divergent at the slow rate* $\log(d)$*. The conditional variance converges to* $\frac{\pi^2}{6\lambda^2}$ *as* $d \to \infty$*, so* $\lim_{D\to\infty} \frac{\varkappa_2(t)}{\varkappa_1(t)} \to 1$*. For* $d = 1$*,* $\varkappa_1(t) = \lambda^{-1}$ *and* $\varkappa_2(t) = 2\lambda^{-2}$*, so* $\varkappa_2(t)/\varkappa_1(t) = 2\lambda^{-1}$*.*

#### 2.4.2. *Assumptions about the noise*

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 follows, for example, Zhou (1998), Bandi & Russell (2005), Zhang et al. (2005), Barndorff-Nielsen et al. (2008) and Hansen & Lunde (2006).

We define the noise associated with $X(\tau_{N,j})$ at the observation time $\tau_{N,j}$ as $U_{N,j} = X(\tau_{N,j}) - Y(\tau_{N,j})$.

**Assumption U.** *Assume the component model*

$$U_{N,i} = \upsilon_{N,i} + \zeta_{N,i}, \quad \textit{where} \quad \upsilon_{N,i} = \sum_{h=0}^{\infty} \psi_h(\tau_{N,i-1-h})\epsilon_{N,i-h}, \quad \textit{with} \quad \epsilon_{N,i} = \Delta_{N,i}^{-1/2}[\tilde{W}(\tau_{N,i}) - \tilde{W}(\tau_{N,i-1})].$$

*Here* $\tilde{W}$ *is a standard Brownian motion and* $\{\zeta_{N,i}\}$ *is a sequence of independent random variables, with* $\mathrm{E}(\zeta_{N,i}|\mathcal{F}_{\tau_{N,i-1}}) = 0$ *and* $\mathrm{var}(\zeta_{N,i}|\mathcal{F}_{\tau_{N,i-1}}) = \Sigma_\zeta(\tau_{N,i-1})$. Further, $\epsilon_{N,i}$ and $\zeta_{N,i}$ *are assumed to be independent, while* $\left(\psi_h, \Sigma_\zeta\right)$ *are bounded and adapted to* $\{\mathcal{F}_t\}$, with $\sum_{j=0}^{\infty} |\psi_j(t)| < \infty$ a.s. uniformly in $t$. We also assume that $D \perp\!\!\!\perp \zeta$.

**Remark 2.** The auxiliary Brownian motion $\tilde{W}$ facilitates a general form of endogenous noise through correlation between $\tilde{W}$ and the Brownian motion, $W$, that drives the underlying process, $Y$. In fact, the case

[3] The earliest research on this that we know of is Jacod (1994). Mykland & Zhang (2006) and Barndorff-Nielsen et al. (2008) provided an analysis based on the assumption that $D_{N,i} = O_p(1)$, which is perhaps too strong an assumption here (see remark 1). Barndorff-Nielsen & Shephard (2005) allowed very general spacing, but assumed times and prices were independent. More recent important contributions include Hayashi, Jacod & Yoshida (2008) and Li, Mykland, Renault, Zhang & Zheng (2009). Mykland & Zhang (2009*a*) and Jacod (2008) provide insightful analysis.

$\tilde{W} = W$ is permitted under our assumptions.

**Remark 3.** The standard assumption in this literature is that $\psi_h(t)$ is zero for all $t$ and $h$, but this assumption is known to be shallow empirically. A $\psi_0(t)$ type term appears in Hansen & Lunde (2006, example 1) and Kalnina & Linton (2008) in their analysis of endogeneity and a two scale estimator.

The "local" long run variance of $v$ is given by $\Sigma_v(t) = \sum_{h=-\infty}^{\infty} \gamma_h(t)$, where $\gamma_h(t) = \sum_{j=1}^{\infty} \psi_{j+h}(t)\psi_j'(t)$ for $h \geq 0$ and $\gamma_h(t) = \gamma_{-h}(t)'$ for $h < 0$, so that the local long run variance of $U$ is given by

$$\Sigma_U(t) = \Sigma_v(t) + \Sigma_\zeta(t).$$

It is convenient to define the average long run variance of $U$ by

$$\Omega = \int_0^1 \Sigma_U(u)\mathrm{d}u,$$

which is a $d \times d$ matrix. When $d = 1$ we write $\omega^2$ in place of $\Omega$, $\sigma_U^2(t)$ in place of $\Sigma_U(t)$, etc. $\omega^2$ appears frequently later. It reflects the variance of the average noise a frequent trader would be exposed to.

## 3. Asymptotic results

### 3.1. Consistency

We note that the multivariate realised kernel can be written as

$$K(X) = K(Y) + K(Y,U) + K(U,Y) + K(U), \quad \text{where} \quad K(Y,U) = \sum_{h=-n+1}^{n-1} k(\tfrac{h}{H}) \sum_j y_j u_{j-h}',$$

with $y_j$ and $u_j$ defined analogous to the definition of $x_j$. This implies immediately that

**Theorem 1.** *Let* **K** *hold and suppose that* $K(Y) = O_p(1)$. *Then* $K(X) - K(Y) = K(U) + O_p(\sqrt{K(U)})$.

Theorem 1 is a very powerful result for dealing with endogenous noise. Note whatever the relationship between $Y$ and $U$, if $K(U) \xrightarrow{p} 0$ then $K(X) - K(Y) = o_p(1)$, so if also $K(Y) \xrightarrow{p} [Y]$ 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 means endogeneity does not matter for consistency. What matters is that the realised kernel applied to the noise process vanishes in probability.

Because realised kernels are built out of these $n$ high frequency returns, it is natural to state asymptotic results in terms of $n$ (rather than $N$).

**Lemma 1.** *Let* **K**, **SH**, **D**, *and* **U** *hold. Then as* $H, n, m \to \infty$ *with* $m/n \to 0$, $H = c_0 n^{\eta}$, $c_0 > 0$, *and* $\eta \in (0, 1)$

$$\frac{H^2}{n} K(X) \xrightarrow{p} \left| k''(0) \right| \Omega, \qquad \text{if } \eta < 1/2, \tag{2}$$

$$K(X) = \int_0^1 \Sigma(u) \mathrm{d}u + c_0^{-2} \left| k''(0) \right| \Omega + O_p(1), \qquad \text{if } \eta = 1/2, \tag{3}$$

$$K(X) \xrightarrow{p} \int_0^1 \Sigma(u) \mathrm{d}u, \qquad \text{if } \eta > 1/2. \tag{4}$$

This shows the crucial role of $H$. $H$ needs to increase with $n$ quite quickly to remove the influence on the estimator of the noise. For very slow rates of increase in $H$ the realized kernel will actually estimate a scaled version of the integrated long-run noise. To estimate the integrated variance our preferred approach is to set $H = c_0 n^{3/5}$, because $n^{3/5}$ is the optimal rate in the trade-off between (squared) bias and variance. In the next section we give a rule for selecting the best possible $c_0$. This approach delivers a positive semi-definite consistent estimator which is robust to endogeneity and semi-staleness of prices.

The results for $H \propto n^{1/2}$ achieve the best available rate of convergence (see below), but is inconsistent. The $O_p(1)$ remainder in (3) can, in some cases, be shown to be $o_p(1)$ in which case the bias, $c_0^{-2} \left| k''(0) \right| \Omega$, can be removed by using (2) with $\eta = 1/5$, for example. However, the resulting estimator is not necessarily positive semi-definite and we do not recommend this in applications.

To study these results in more detail we will develop a central limit theory for the realised kernels, which allows us to select a sensible $H$. Before introducing the multivariate results, it is helpful to consider the univariate case.

## 3.2. Univariate asymptotic distribution

### 3.2.1. *Core points*

The univariate version of the main results in our paper is the following.

**Theorem 2.** *Let* **K**, **SH**, **D**, *and* **U** *hold. If* $n \to \infty$, $H = c_0 n^{3/5}$ *and* $m^{-1} = o(\sqrt{H/n}) = o(n^{-1/5})$ *we have that*

$$n^{1/5} \left( K(X) - \int_0^1 \sigma^2(u) \mathrm{d}u \right) \xrightarrow{Ls} \mathrm{MN} \left( c_0^{-2} |k''(0)| \omega^2, 4 c_0 k_{\bullet}^{0,0} \int_0^1 \sigma^4(u) \frac{\varkappa_2(u)}{\varkappa_1(u)} \mathrm{d}u \right).$$

The notation $\xrightarrow{Ls}$ MN means stable convergence to a mixed Gaussian distribution. This notion is important for the construction of confidence intervals and the use of the delta method. The reason is that

$\int_0^1 \sigma^4(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u$ is random, and stable convergence guarantees joint convergence that is needed here. Stable convergence is discussed, for example, in Mykland & Zhang (2006), who also provide extensive references.

The minimum mean square error of the $H = c_0 n^{3/5}$ estimator is achieved by setting $c_0 = c^* \xi^{4/5}$ so $H = c^* \xi^{4/5} n^{3/5}$ where

$$c^* = \left\{ \frac{k''(0)^2}{k_\bullet^{0,0}} \right\}^{1/5}, \quad \xi^2 = \frac{\omega^2}{\sqrt{\mathrm{IQ}}}, \quad \mathrm{IQ} = \int_0^1 \sigma^4(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u.$$

Notice that the serial dependence in the noise will impact the choice of $c_0$ with ceteris paribus increasing dependence leading to larger values of $H$. Then

$$c_0 k_\bullet^{0,0}\mathrm{IQ} = \kappa^2, \quad \frac{|k''(0)|}{c_0^2}\omega^2 = \kappa, \quad \text{where} \quad \kappa = \kappa_0 \left\{\mathrm{IQ}\omega\right\}^{2/5}, \quad \kappa_0 = \left( |k''(0)| \left(k_\bullet^{0,0}\right)^2 \right)^{1/5}.$$

Then

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

This shows both the bias and variance of the realised kernel will increase with the value of the long-run variance of the noise. Interestingly time-variation in the noise does not, in itself, change the precision of the realised kernel — all that matters is the average level of the long-run variance of the noise. For the Parzen kernel we have $\kappa_0 = 0.97$.

### 3.2.2. *Some additional comments*

The conditions on $m$ is caused by end effects, as these induce a bias in $K(U)$ that is of the order $2m^{-1}\omega^2$. Empirically $\omega^2$ is tiny so $2m^{-1}\omega^2$ will be small even with $m = 1$, but theoretically this is an important observation. Assumption **D**.i implies $\mathrm{Var}(D_{n,\lfloor tn \rfloor}|\mathcal{F}_{\tau_{\lfloor tn \rfloor -1}}) \xrightarrow{p} \varkappa_2(t) - \varkappa_1^2(t)$, which is non-negative. Thus we have the inequality, $\varkappa_2(t)/\varkappa_1(t) \geq \varkappa_1(t)$, which means that $\int_0^1 \sigma^4(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u \geq \int_0^1 \sigma^4(u)\varkappa_1(u)\mathrm{d}u$. So the asymptotic variance above is higher than a process with time-varying but non-stochastic durations. The random nature of the durations inflates the asymptotic variance.

The result looks weak compared to the corresponding result for the flat-top kernel $K^F(X)$ introduced by Barndorff-Nielsen et al. (2008) with $k'(0) = 0$. They had the nicer result that[4]

$$n^{1/4}\left\{ K^F(X) - \int_0^1 \sigma^2(u)\mathrm{d}u \right\} \xrightarrow{Ls} \mathrm{MN}\left\{ 0, 4c_0 k_\bullet^{0,0}\mathrm{IQ} + \tfrac{8}{c_0}k_\bullet^{1,1}\omega^2 \int_0^1 \sigma^2(u)\mathrm{d}u + \tfrac{4}{c_0^3}k_\bullet^{2,2}\omega^4 \right\},$$

when $H = c_0 n^{1/2}$, under the (far more restrictive) assumption that $U$ is white noise. Hence, the implication is that the kernel estimators proposed in this paper will be (asymptotically) inferior to $K^F(X)$ in the special

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

case where $U$ is white noise. The advantage of our estimator, which has $H = c_0 n^{3/5}$, is that it is based on far more realistic assumptions about the noise. This has the practical implication that $K(X)$ can be applied to prices that are sampled at the highest possible frequency. This point is forcefully illustrated in Section 6.1.2 where we compare the two estimators, $K(X)$ and $K^F(X)$, and show the importance of being robust to endogeneity and serial dependence. A simulation design shows that $K(X)$ is far more accurate than $K^F(X)$ when the noise is serially dependent. Moreover, as an extra benefit of constructing our estimator from $\mathcal{K}$ is that it ensures positive semi-definiteness. Naturally, one can always truncate an estimator to be psd, for instance by replacing negative eigenvalues with zeros. Still, we find it convenient that the estimator is guaranteed to be psd, because it makes a check for positive definiteness and correction for lack thereof, entirely redundant.

Having an asymptotic bias term in the asymptotic distribution is familiar from kernel density estimation with the optimal bandwidth. The bias is modest so long as $H$ increases at a faster rate than $\sqrt{n}$. If $k''(0) = 0$ 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).

The following theorem rules out an important class of estimators which seems to be attractive to empirical researchers.

**Lemma 2.** *Given* $\boldsymbol{U}$ *and a kernel function with* $k'(0) \neq 0$ *but otherwise satisfies* $\boldsymbol{K}$*. Then, as* $n, H, m \to \infty$ *we have that* $\frac{H}{n} K(U) \xrightarrow{p} 2\left|k'(0)\right| \int_0^1 \left\{\Sigma_\zeta(u) + \gamma_0(u)\right\} \mathrm{d}u$.

**Remark 4.** If $k'(0) \neq 0$ then there does not exist a consistent $K(X)$. This rules out, for example, the well known Bartlett type estimator in this context.

### 3.2.3. *Choosing the bandwidth $H$ and weight function*

The relative efficiency of different realised kernels in this class are determined solely by the constant $|k''(0)(k_\bullet^{0,0})^2|^{1/5}$ 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, $|k''(0)(k_\bullet^{0,0})^2|^{1/5}$, 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 $\xi^2$ in Section 3.4.

[Table 1 about here.]

### 3.3. Multivariate asymptotic distribution

To start we extend the definition of the integrated quarticity to the multivariate context

$$\mathrm{IQ} = \int_0^1 \{\Sigma(u) \otimes \Sigma(u)\} \frac{\varkappa_2(u)}{\varkappa_1(u)} \mathrm{d}u,$$

which is a $d^2 \times d^2$ random matrix.

**Theorem 3.** *Suppose $H = c_0 n^{3/5}$, $m^{-1} = o(n^{-1/5})$,* **K**, **SH**, **D**, *and* **U** *then*

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

This is the multivariate extension of Theorem 2, yielding a limit theorem for the consistent multivariate estimator in the presence of noise. The bias is determined by the long-run variance $\Omega$, whereas the variance depends solely on the integrated quarticity.

**Corollary 1.** *An implication of Theorem 3 is that for $a, b \in \mathbb{R}^d$ we have*

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

*where $v_{ab} = \mathrm{vec}(\frac{ab'+ba'}{2})$. For two different elements, $a'K(X)b$ and $c'K(X)d$ say, their asymptotic covariance is given by $4c_0 k_\bullet^{0,0} v_{ab}' \mathrm{IQ} v_{cd}$.*

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

**Example 2.** *In the bivariate case we can write the results as*

$$n^{1/5} \begin{pmatrix} K(X^{(i)}) - \int_0^1 \Sigma_{ii} \mathrm{d}u \\ K(X^{(i)}, X^{(j)}) - \int_0^1 \Sigma_{ij} \mathrm{d}u \\ K(X^{(j)}) - \int_0^1 \Sigma_{jj} \mathrm{d}u \end{pmatrix} \stackrel{Ls}{\to} \mathrm{MN}(A, B), \tag{6}$$

*where*

$$A = c_0^{-2} |k''(0)| \begin{pmatrix} \Omega_{ii} \\ \Omega_{ij} \\ \Omega_{jj} \end{pmatrix} \quad \text{and} \quad B = 2c_0 k_\bullet^{0,0} \int_0^1 \begin{pmatrix} 2\Sigma_{ii}^2 & 2\Sigma_{ii}\Sigma_{ij} & 2\Sigma_{ij}^2 \\ \bullet & \Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2 & 2\Sigma_{ii}\Sigma_{ji} \\ \bullet & \bullet & 2\Sigma_{jj}^2 \end{pmatrix} \frac{\varkappa_2}{\varkappa_1} \mathrm{d}u,$$

*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 (extending the work of, for example, Andersen, Bollerslev, Diebold & Labys (2003), Barndorff-Nielsen & Shephard (2004) and Dovonon, Goncalves & Meddahi (2007)). For example, with* $\beta^{(i,j)} = \int_0^1 \Sigma_{ij}\,\mathrm{d}u \Big/ \int_0^1 \Sigma_{jj}\,\mathrm{d}u$,

$$n^{1/5}\left(\frac{K(X^{(i)}, X^{(j)})}{K(X^{(j)})} - \beta^{(i,j)}\right) \xrightarrow{Ls} \mathrm{MN}\,(A, B)\,, \quad \textit{where} \quad A = \frac{c_0^{-2}|k''(0)|}{\int_0^1 \Sigma_{jj}\mathrm{d}u}\left(\Omega_{ij} - \Omega_{jj}\beta_{ij}\right),$$

*and*

$$B = \frac{2c_0 k_\bullet^{0,0}}{\left(\int_0^1 \Sigma_{jj}\mathrm{d}u\right)^2}\begin{pmatrix} 1, & -\beta^{(i,j)} \end{pmatrix}\left[\int_0^1 \begin{pmatrix} \Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2 & 2\Sigma_{ii}\Sigma_{ji} \\ 2\Sigma_{jj}\Sigma_{ij} & 2\Sigma_{jj}^2 \end{pmatrix}\frac{\varkappa_2}{\varkappa_1}\mathrm{d}u\right]\begin{pmatrix} 1 \\ -\beta^{(i,j)} \end{pmatrix}.$$

### 3.4. **Practical issues: Choice of $H$**

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 $\xi_i^2 = \Omega_{ii}/\sqrt{\mathrm{IQ}_{ii}}$, where $\Sigma_{ii}(u)$ is the spot variance for the $i$-th asset. In practice we usually approximate $\sqrt{\mathrm{IQ}_{ii}}$ 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 $\Omega_{ii}$ which use high frequency data. Then we could construct some ad hoc rules for choosing the global $H$, such as $H_{\min} = \min(H^{(1)}, ..., H^{(d)})$, $H_{\max} = \max(H^{(1)}, ..., H^{(d)})$, or $\bar{H} = d^{-1}\sum_{i=1}^d H^{(i)}$, 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 $\iota$ 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 2 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 $\beta^{(i,j)}$. For $\beta^{(1,2)}$ the trade-off is between $c_0^{-4}|k''(0)|^2\left(\Omega_{12}-\Omega_{22}\beta^{(1,2)}\right)^2$, and

$$2c_0k_\bullet^{0,0}\int_0^1\left(\Sigma_{11}\Sigma_{22}+\Sigma_{12}^2-4\beta^{(1,2)}\Sigma_{11}\Sigma_{22}+2\beta^{(1,2)2}\Sigma_{22}\right)\frac{\varkappa_2}{\varkappa_1}\mathrm{d}u.$$

## 4. Simulation Study

So far the analysis has been asymptotic as $n\to\infty$. Here we carry 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[5], and the fact that the OU-process have an exact discretization (e.g. Glasserman (2004, pp. 110)). Our simulations are based on the following configuration $(\mu^{(i)},\beta_0^{(i)},\beta_1^{(i)},\alpha^{(i)},\rho^{(i)})=(0.03,-5/16,1/8,-1/40,-0.3)$, 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 $\sigma^2$ 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

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

where the noise-to-signal ratio, $\xi^2$ 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)).

[5]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.

To model the non-synchronously spaced data we use two independent Poisson process sampling schemes to generate the times of the actual observations $\{t_j^{(i)}\}$ 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 $\lambda$ through 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.

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 $H_{\text{mean}}$ are presented in Table 2.

[Table 2 about here.]

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 $\xi^2$ 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 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.

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 variance dominates the root MSE. Taken together this implies the realised kernel estimators of the correlation and regression (beta) 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 in the correlation case.

## 5. Empirical illustration

We analyze high-frequency assets prices for thirty assets.[6] In the analysis the main focus will be on the empirical properties of $30 \times 30$ realised kernel estimates. To conserve space we will only present detailed results for a $10 \times 10$ submatrix of the full $30 \times 30$ matrix. The ten assets we will focus on are 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), El DuPont de Nemours & Co. (DD), 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, 2002 to July 31, 2008, delivering 1503 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 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 (2009). In this paper we follow the step-by-step cleaning procedure used in Barndorff-Nielsen et al. (2009) 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 INTC and MFST from NASDAQ and for SPY for which all retained observations are from Pacific). Delete other entries.

**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)

[6] The ticker symbols of these assets are AA, AIG, AXP, BA, BAC, C, CAT, CVX, DD, DIS, GE, GM, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PFE, PG, SPY, T, UTX, VZ, WMT, and XOM.

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 10 mean absolute deviations from a centered median (excluding the observation under consideration) of 50 observations.

**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. 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 estimator by Hayashi & Yoshida (2005) 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 returns, 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 is used. In our analysis we present estimates for the upper left $10 \times 10$ block of the full $30 \times 30$ integrated covariance matrix. The estimates are constructed using three different sampling schemes. a) Refresh Time sampling applied to full set of DJ stocks, b) Refresh Time sampling applied to only the 10 stock that we focus on and c) Refresh Time sampling applied to each unique pair of assets. So in our analysis we will present three sets of realized kernel estimates of the elements of the integrated covariance matrix. One set that comes from a $30 \times 1$ vector of returns, the same set estimated using only the required $10 \times 1$ vector of returns, and finally a set constructed from the 45 distinct $2 \times 2$ covariance matrix estimates. Note that the two first estimators are positive semidefinite by construction, while the latter is not guaranteed to be so. We compute these covariance matrix estimates for each day in our sample.

The fraction of the data we retained by constructing Refresh Time is recorded in Table 3 for each of the 45 distinct $2 \times 2$ matrices. It records the average of the daily $p$ statistics defined in Section 2.1 for each pair. It emerges that we never 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.

For the $10 \times 1$ case the data loss is more pronounced. Still, on average more that 25 percent of the observations remain in the sample. For transaction data the average number of Refresh Time observations is $1,470$, whereas the corresponding number is $4,491$ for the quote data. So in most cases we have an observation on average more often than every 5 seconds for quote data and 15 seconds for trade data. We observed that the data loss levels off as the dimension increases. For the $30 \times 1$ case we have on average more that 17 percent of the observations remaining in the sample. For transaction data the average number of Refresh Time observations is 966 and $2,978$ for the quote data. This gives an observation on average more often than every 8 seconds for quote data and 24 seconds for trade data.

[Table 3 about here.]

### 5.3. Analysis of the covariance estimators: $\mathrm{Cov}_s^K$, $\mathrm{Cov}_s^{HY}$, $\mathrm{Cov}_s^{OtoC}$ and $\mathrm{Cov}_s^{\Delta\mathrm{m}}$

Throughout this subsection the target which we wish to estimate is $[Y^{(i)}, Y^{(j)}]_s$, $i, j = 1, 2, \ldots, d$, $s \in \mathbb{N}$. 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 the full 30-dimensional vector, the 10-dimensional sub-vector and (all possible) pairs of the ten assets. The resulting estimates of $[Y^{(i)}, Y^{(j)}]_s$ are denoted by $\mathrm{Cov}_s^{K_{30\times30}}$, $\mathrm{Cov}_s^{K_{10\times10}}$ and $\mathrm{Cov}_s^{K_{2\times2}}$, respectively. These 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: $\mathrm{Cov}_s^{HY}$, the Hayashi & Yoshida (2005) covariance estimator. $\mathrm{Cov}_s^{\Delta}$, the realised covariance based on intraday returns that span a interval of length $\Delta$, e.g. 5 or 30 minutes (the previous-tick method is used). $\mathrm{Cov}_s^{OtoC}$, 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 Hayashi & Yoshida (2005) estimator $\mathrm{Cov}_s^{HY}$ and the open to close estimator $\mathrm{Cov}_s^{OtoC}$. 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. It is interesting to note that the $\mathrm{Cov}_s^{HY}$ estimates are typically much lower than the corresponding $\mathrm{Cov}_s^{OtoC}$ estimate. Numbers in bold font indicate estimates that are significantly different from $\mathrm{Cov}_s^{OtoC}$ at the one percent level. This assessment is carried out in the following way.

[Table 4 about here.]

For a given estimator, e.g. $\text{Cov}_s^{K_{2\times 2}}$, we consider the difference $e_s = \text{Cov}_s^{K_{2\times 2}} - \text{Cov}_s^{OtoC}$, and compute the sample bias as $\bar{e}$ and robust (HAC) variance as $s_\eta^2 = \gamma_0 + 2\sum_{h=1}^{q}\left(1 - \frac{h}{q+1}\right)\gamma_h$, where $\gamma_h = \frac{1}{T-h}\sum_{s=1}^{n-h}\eta_s\eta_{s-h}$. Here $\eta_s = e_s - \bar{e}$ and $q = \text{int}\left\{4(T/100)^{2/9}\right\}$. The number is boldfaced if $\left|\sqrt{T}\bar{e}/s_\eta\right| > 2.326$. The results in Table 4 indicate that $\text{Cov}_s^{HY}$ is severely downward bias. Every covariance estimator for every pair of assets for both trades and quotes are statistically significantly biased.

### 5.4. Results for $\text{Cov}_s^{K_{30\times 30}}$, $\text{Cov}_s^{K_{10\times 10}}$ and $\text{Cov}_s^{K_{2\times 2}}$

We now move on to more successful estimators. The upper panel of Table 5 presents the time series average estimates for $\text{Cov}_s^{K_{30\times 30}}$, the middle panel for $\text{Cov}_s^{K_{10\times 10}}$, and the lower panel give results for $\text{Cov}_s^{K_{2\times 2}}$. The diagonal elements are the estimates based on transactions. Off-diagonal numbers are boldfaced if they are significantly biased (compared to $\text{Cov}_s^{OtoC}$) 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.

[Table 5 about here.]

[Table 6 about here.]

A much tougher comparison is to replace the noisy $e_s = \text{Cov}_s^{K} - \text{Cov}_s^{OtoC}$ with $e_s = \text{Cov}_s^{K_{d\times d}} - \text{Cov}_s^{K_{d'\times d'}}$, where the two estimates come from applying the realized kernel to price vectors of dimension $d$ and $d'$. 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 the three realised kernel estimators. When we compute the same test based on $e_s = \text{Cov}_s^{K_{d\times d}} - \text{Cov}_s^{5\text{m}}$ we find that the realized kernels and the realised covariances based on 5 minutes returns are also quite similar.

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 $\text{Cov}_s^{HY}$ which is sensitive to this.

The Table shows $\text{Cov}_s^{K_{30\times 30}}$, $\text{Cov}_s^{K_{10\times 10}}$ and $\text{Cov}_s^{K_{2\times 2}}$ have roughly the same average value, which is slightly below $\text{Cov}_s^{OtoC}$. $\text{Cov}_s^{K_{2\times 2}}$ has a seven times smaller variance than $\text{Cov}_s^{OtoC}$, which shows it is a lot more precise. Of course integrated variance is its self random so seven underestimates the efficiency gain of using $\text{Cov}_s^{K_{2\times 2}}$. If volatility is close to being persistent then $\text{Cov}_s^{K_{30\times 30}}$ is at least $\frac{4.255^2}{1.61^2(1-acf_1)} \simeq 20$ times more informative than the cross product of daily returns. The same observation holds for mid-quotes.

$\mathrm{Cov}_s^{15\mathrm{s}}$ and $\mathrm{Cov}_s^{HY}$ are very precise estimates of the wrong quantity. $\mathrm{Cov}_s^{5\mathrm{m}}$ is quite close to $\mathrm{Cov}_s^{K_{30\times30}}$, $\mathrm{Cov}_s^{K_{10\times10}}$ and $\mathrm{Cov}_s^{K_{2\times2}}$, with $\mathrm{Cov}_s^{5\mathrm{m}}$ and $\mathrm{Cov}_s^{K_{30\times30}}$ having a correlation of 0.942. We note that realised kernel results seem to show some bias compared to $\mathrm{Cov}_s^{OtoC}$, the difference is however statistically insignificantly different than zero, as $\mathrm{Cov}_s^{OtoC}$ turns out to be very noisy.

The corresponding results for correlations are interesting. Naturally, the computation of the correlation involves a non-linear transformation of roughly unbiased and noisy estimates. We should therefore (by a Jensen inequality argument) expect all these estimates to be biased. The most persistent estimator is $\mathrm{Corr}_s^{1/4m}$, but the high autocorrelation merely reflects the large distortion that noise has on this estimator, as is also evident from the sample average of this correlation estimator. The largest autocorrelation amongst the more reliable estimators is that of $\mathrm{Corr}_s^{K_{2\times2}}$, which suggest that this is most effective estimate of the correlation.

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 $\mathrm{Cov}_s^{K_{2\times2}}$ 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 $\mathrm{Cov}_s^{K_{2\times2}}$, $\mathrm{Cov}_s^{5\mathrm{m}}$ and $\mathrm{Cov}_s^{30\mathrm{m}}$, while once again $\mathrm{Cov}_s^{HY}$ struggles. Overall $\mathrm{Cov}_s^{K_{2\times2}}$ and $\mathrm{Cov}_s^{5\mathrm{m}}$ behave in a similar manner, with $\mathrm{Cov}_s^{K_{2\times2}}$ slightly stronger. $\mathrm{Cov}_s^{K_{10\times10}}$ estimates roughly the same level as $\mathrm{Cov}_s^{K_{2\times2}}$ 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 $\hat{\rho}_s^{X\mathrm{m}}$.

A table in our web Appendix reports the average estimates for $\hat{\rho}_s^{K_{2\times2}}$, $\hat{\rho}_s^{K_{10\times10}}$ and $\hat{\rho}_s^{5\mathrm{m}}$. 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.

### 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 $\beta_s^{K_{2\times 2}}$ 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 about here.]

[Figure 3 about here.]

We also calculate the encompassing regressions. The estimates 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**

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 multivariate GARCH models 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 $d \times 1$ vector of daily close to close returns, $\mathcal{F}_{s-1}$ 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))

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

Here we follow the literature and use $H_s$ to denote the conditional variance matrix (not to be confused with our bandwidth parameters).

Instead of estimating the $d(d+1)/2$ unique elements of $C$ we use a variant of variance targeting as suggested in Engle & Mezrich (1996). The general idea is to estimate the intercept matrix by an auxiliary estimator that is given by

$$\hat{C}'\hat{C} = \bar{S} \odot (1 - \alpha - \beta - \gamma\kappa), \qquad \bar{S} = \frac{1}{T}\sum_{s=1}^{T} r_s r_s', \tag{7}$$

where $\odot$ denotes the Hadamard product. There is a slight deviation from the situation considered by Engle & Mezrich (1996) because $\mathrm{K}_{s-1}^{K_{2\times 2}}$ is only estimated for the part of the day where the NYSE is open. To accommodate this we follow Shephard & Sheppard (2009) that introduce the scaling matrix, $\kappa$, in (7) which we estimate by

$$\hat{\kappa}_{ij} = \left(\frac{\hat{\mu}_{RK}}{\hat{\mu}}\right)_{ij}, \qquad \hat{\mu} = T^{-1}\sum_{s=1}^{T} r_s r_s' \quad \text{and} \quad \hat{\mu}_{RK} = T^{-1}\sum_{s=1}^{T} \mathrm{K}_s.$$

Having $\bar{S}$ and $\hat{\kappa}$ at hand the remaining parameters are simply estimated by maximizing the concentrated quasi-log-likehood, with

$$H_s = \bar{S} \odot (1 - \alpha - \beta - \gamma\hat{\kappa}) + \beta H_{s-1} + \alpha r_{s-1} r_{s-1}' + \gamma \mathrm{K}_{s-1}, \quad \alpha, \beta, \gamma \geq 0.$$

An interesting question is whether $\gamma$ is statistically different from zero, because this means that high frequency data enhances the forecast of future covariation. In our analysis we will also augment the model with $\mathrm{RV}_{s-1}^{5m}$

We estimate scalar BEKK models for the $30 \times 30$, $10 \times 10$, and the 45 $2 \times 2$ cases. In Table 7 we present estimates for the two larger dimensions and three selected $2 \times 2$ cases. The results in Table 7 suggest that lagged daily returns are no longer significant for this multivariate GARCH model 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 $\mathrm{K}_{s-1}$ reduces the size of the estimated $H_{s-1}$ term. It is also interesting to note that including $\mathrm{K}_{s-1}$ in general gives a higher log-likelihood than including $\mathrm{RV}_{s-1}^{5m}$. This holds for both the 30-dimensional and the 10-dimensional cases, and for 40 of the 45 2-dimensional cases. In our web appendix we report summary statistics of two likelihood ratio tests applied to all the 45 2-dimensional cases. The average LR statistic for removing $\mathrm{RV}_{s-1}^{5m}$ from our most general specification is 0.66, where as the corresponding average for removing $\mathrm{K}_{s-1}$ is 11.9. These tests can be interpreted as encompassing tests, and provide an overwhelming evidence that the information in $\mathrm{RV}_{s-1}^{5m}$is contained in $\mathrm{K}_{s-1}$.

[Table 7 about here.]

## 6. Additional remarks

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

In the univariate case the realised kernel $K(X) = \sum_{h=-n}^{n} k(\frac{h}{H})\Gamma_h$, with $\Gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-|h|}$, is at first sight very similar to the unbiased flat-top realised kernel of Barndorff-Nielsen et al. (2008)

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

Here the $\Gamma_h$ and $\Gamma_h^F$ 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. Jittering eliminates the end effects in $K(X)$, whereas the presence of $x_{-1}, x_{-2}, \ldots$ and $x_{n+1}, x_{n+2}, \ldots$ in the the definition of $\Gamma_h^F$ removes the end effects from $K^F(X)$. 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 (under the restrictive independent noise assumption) 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 sensitive to deviations from independent noise, such as serial dependence in the noise and endogenous noise, which $K(X)$ is robust to. The requirement that $K(X)$ be positive results in the bias-variance trade-off and reduces the best rate of convergence from $n^{1/4}$ to $n^{1/5}$. This resembles the effects seen in the literature on density estimation with kernel functions. The property, $\int u^2 k(u)\mathrm{d}u = 0$, reduces the order of the asymptotic bias, but kernel functions that satisfy $\int u^2 k(u)\mathrm{d}u = 0$ can result in negative density estimates, see Silverman (1986, sections 3.3 and 3.6).

### 6.1.1. *Positivity*

There are three reasons that $K^F(X)$ can go negative.[7] 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 $\gamma_1$ and $\gamma_{-1}$, 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 $M = \mathrm{band}(1, 1, k(\frac{1}{H}), k(\frac{2}{H}), \ldots, )$. 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 $U$ is white noise. 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 $U$ is white noise. 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 $U$ white noise with variance of $\omega^2$. 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 $K^F(X)$

[7] The flat-top kernel is only rarely negative with modern data. However, if $[Y]$ is very small and the $\omega^2$ very large, which we saw on slow days on the NYSE when the tick size was \$1/8, then it can happen quite often when the flat-top realised kernel is used. We are grateful to Kevin Sheppard for pointing out these negative days.

has roughly half the MSE of $K(X)$ in the univariate case. Hence in ideal (but unrealistic) 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 using 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 $\epsilon_j$ is Gaussian white noise. The bandwidth for all estimators were to be "optimal" under $U$ being white noise, 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 $\omega^2 = \sum_{h=-\infty}^{\infty} \text{cov}(U_j, U_{j-h})$. 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 about here.]

### 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 case, as to how to choose $\theta$.

In the univariate bias corrected form, Jacod et al. (2009) 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.

Following the initial draft of this paper Christensen, Kinnebrock & Podolskij (2009) have defined a bias corrected preaveraging estimator of the multivariate $[Y]$ with $H = \theta n^{1/2}$, for which they derive limit theory. 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 the stochastic $d^2 \times d^2$ matrix, IQ. 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 vec (IQ) 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. **Finite sample improvements**

The realised kernel is non-negative so we can use log-transform

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

to improve its finite sample performance. 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 depends less on $\sigma^2$ than the non-transformed version.

### 6.4. **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 $U$ is white noise. 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 $4m^{-2}\omega^4 + 4m^{-2}\omega^4 = 8\omega^4 m^{-2}$, as shown in the proof of Lemma A.2. The asymptotic variance (abstracting from end-effects) is $5\kappa^2 n^{-2/5} = 5\left|k''(0)\omega^2\right|^{2/5}\left\{k_\bullet^{0,0}\mathrm{IQ}\right\}^{4/5} n^{-2/5}$. 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.

[Figure 4 about here.]

### 6.5. **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 are 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.

### 6.6. **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 $\mathcal{BSM}$ plus a finite activity jump process. The analysis in Barndorff-Nielsen et al. (2008, 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 an inconsistent estimator in this context.

Our analysis is based on three innovations: (i) we used a weight function which delivers biased kernels, 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. We are able to show consistency and asymptotic mixed Gaussianity of our estimator.

Our simulation study indicates our estimator is close to being unbiased for covariances under realistic situations. Not surprisingly the estimators of correlations are 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. We have analysed problems of up to 30 dimensions and have found that efficiency gains of using the high frequency data are around 20 fold.

Multivariate realised kernels have potentially many areas of application, improving our ability to estimate covariances. In particular, this allows us to utilize high frequency data to significantly improve our predictive models as well as providing a better understand of asset pricing and management of risk in financial markets.

## References


Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2010), Parametric and nonparametric measurement of volatility, *in* Y. Aït-Sahalia & L. P. Hansen, eds, 'Handbook of Financial Econometrics', North Holland, Amsterdam, pp. 67–138.

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. (2010), 'Market microstructure noise and realized volatility forecasting', *Journal of Econometrics* .

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.

Bandi, F. M. & Russell, J. R. (2008), 'Microstructure noise, realized variance, and optimal sampling', *Review of Economic Studies* **75**, 339–369.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008), 'Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realised kernels in practice: Trades and quotes', *Econometrics Journal* **12**, 1–33.

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. (2005), 'Power variation and time change', *Theory of Probability and Its Applications* **50**, 1–15.

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.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2009), Pre-averaging estimators of the ex-post covariance matrix. Research paper 2009-45, CREATES, Aarhus University.

de Pooter, M., Martens, M. & van Dijk, D. (n.d.), 'Predicting the daily covariance matrix for s&p 100 stocks using intraday data — but which frequency to use?', *Econometric Reviews* **27**, 199–229.

Doornik, J. A. (2006), *Ox: An Object-Orientated Matrix Programming Langguage*, 5 edn, Timberlake Consultants Ltd., London.

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., Klüppelberg, 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.

Engle, R. F. & Mezrich, J. (1996), 'GARCH for groups', *Risk* pp. 36–40.

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.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic values of volatility timing using 'realized' volatility', *Journal of Financial Economics* **67**, 473–509.

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.

Goncalves, S. & Meddahi, N. (2009), 'Bootstrapping realized volatility', *Econometrica* **77**, 283–306.

Griffin, J. E. & Oomen, R. C. A. (2009), 'Covariance measurement in the presence of non-synchronous trading and market microstructure noise', *Journal of Econometrics* . Forthcoming.

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. & Horel, G. (2009), 'Quadratic variation by markov chains', *working paper* . `http://www.stanford.edu/people/peter.hansen`.

Hansen, P. R., Large, J. & Lunde, A. (2008), 'Moving average-based estimators of integrated variance', *Econometric Reviews* **27**, 79–111.

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

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., Jacod, J. & Yoshida, N. (2008), Irregular sampling and central limit theorems for power variations: the continuous case. Unpublished paper, Keio University.

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

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

Jacod, J. (2008), Statistics and high frequency data. Unpublished paper.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: the pre-averaging approach', *Stochastic Processes and Their Applications* **119**, 2249–2276.

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* **147**, 47–59.

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

Li, Y., Mykland, P., Renault, E., Zhang, L. & Zheng, X. (2009), Realized volatility when endogeniety of time matters. Working Paper, Department of Statistics, University of Chicago.

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. (2009*a*), The econometrics of high frequency data, *in* A. L. M. Kessler and & M. S. rensen, eds, 'Statistical Methods for Stochastic Differential Equations', Forthcoming, Chapman & Hall/CRC Press.

Mykland, P. A. & Zhang, L. (2009*b*), 'Inference for continuous semimartingales observed at high frequency: A general approach', *Econometrica* **77**, 1403–1455.

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. (2010), 'Causality effects in return volatility measures with random times', *Journal of Econometrics* . Forthcoming.

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

Shephard, N. & Sheppard, K. K. (2009), 'Realising the future: Forecasting with high frequency based volatility (HEAVY) models', *Journal of Applied Econometrics* . Forthcoming.

Silverman, B. W. (1986), *Density Estimation for Statistical and Data Analysis*, Chapman & Hall, London.

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. (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.

- Show the realised kernel is consistent and work out its limit theory for synchronized data. This is shown in Appendix A, where Propositions A1-A5 and Theorems A3-A4 are used to establish the multivariate result in Theorem 3 and the univariate result in Theorem 2 then follows as a corollary to Theorem 3.
- 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.

## Appendix A: Proofs for synchronised data

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

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

is positive semi-definite. This means that by taking the determinant of this matrix and rearranging we see that $K(Y^{(i)}, U^{(j)})^2 \leq K(Y^{(i)})K(U^{(j)})$, so that

$$K(X) = K(Y) + O\left(\sqrt{\max_i K\left(Y^{(i)}\right)}\sqrt{\max_j K\left(U^{(j)}\right)}\right) + K(U),$$

and the result follows. □

Next collect limit results about $K(Y)$ and $K(U)$. Due to Theorem 1 we can safely ignore the cross terms $K(U, Y)$ as long as $K(U)$ vanishes at the appropriate rate.

### A.1. Results concerning $K(U)$

The aim of this subsection to is prove the following Proposition.

**Theorem A.4.** *Under* **K** *and* **U** *then*

$$\frac{H^2}{n}K(U) \xrightarrow{p} -k''(0)\Omega, \quad \text{as } n, H, m \to \infty \text{ with } H^2/(mn) \to 0.$$

Before we prove Theorem A.4, we establish some intermediate results. The following definitions lead to a useful representation of $K(U)$. For $h = 0, 1, \ldots$ we define

$$V_h = \sum_{j=h+1}^{n-1} U_j U'_{j-h} + U_{j-h}U'_j, \quad \text{and} \qquad Z_h = \left(U_0 U'_h + U_h U'_0\right) + (U_n U'_{n-h} + U_{n-h}U'_n).$$

**Proposition A.1.** *The realised autocovariances of $U$ can be written as*

$$\Gamma_0(U) = V_0 - V_1 + \tfrac{1}{2}Z_0 - Z_1 \tag{A.1}$$

$$\Gamma_h(U) + \Gamma_h(U)' = -V_{h-1} + 2V_h - V_{h+1} + Z_h - Z_{h+1}, \tag{A.2}$$

*so with $k_h = k(\frac{h}{H})$ we have*

$$K(U) = (k_0 - k_1)\, V_0 - \sum_{h=1}^{n-1}(k_{h+1} - 2k_h + k_{h-1})\, V_h + \tfrac{1}{2}Z_0 - \sum_{h=1}^{n-1}(k_h - k_{h-1})\, Z_h. \tag{A.3}$$

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

$$\Gamma_0(U) = \sum_{j=1}^{n}(U_j - U_{j-1})(U_j - U_{j-1})' = U_0U_0' + U_nU_n' + \sum_{j=1}^{n-1}(U_jU_j' + U_jU_j')$$
$$- \sum_{j=2}^{n-1}(U_jU_{j-1}' + U_{j-1}U_j') - (U_nU_{n-1}' + U_{n-1}U_n' + U_0U_1' + U_1U_0'),$$

and (A.2) is proven similarly. □

We note that end-effects can only have an impact on $K(U)$ through $Z_h$, $h = 0, 1, \ldots,$ because $U_0$ and $U_n$ do not appear in the expressions for $V_h$, $h = 0, 1, \ldots$.

**Proposition A.2.** *Given* $\mathbf{U}$. *Then*

$$\frac{1}{n}V_h \xrightarrow{p} \begin{cases} 2\int_0^1 \left\{\Sigma_\zeta(u) + \gamma_0(u)\right\} \mathrm{d}u & \text{for } h = 0, \\ \int_0^1 \left\{\gamma_h(u) + \gamma_h(u)'\right\} \mathrm{d}u & \text{for } h > 0, \end{cases}$$

*and* $Z_h = O_p(m^{-1})$ *for all* $h = 0, 1, \ldots$ *and as* $m \to \infty$,

$$mZ_h \xrightarrow{p} \begin{cases} 2\left\{\Sigma_U(0) + \Sigma_U(1)\right\} & \text{for } h = 0, \\ \sum_{j=0}^{\infty}\{\gamma_{j+h}(0) + \gamma_{j+h}(0)' + \gamma_{j+h}(1) + \gamma_{j+h}(1)'\} & \text{for } h > 0. \end{cases}$$

Note that $\int_0^1 \left\{\Sigma_\zeta(u) + \gamma_0(u)\right\} \mathrm{d}u$ is the average local variance of $U$ as oppose to the average long-run variance $\Omega = \int_0^1 \left\{\Sigma_\zeta(u) + \sum_{h=-\infty}^{\infty} \gamma_h(u)\right\} \mathrm{d}u$.

**Proof of Proposition A.2.** The first result follows by the definition of $V_h$ and $\mathbf{U}$. Next, since $U_0 = m^{-1}\sum_{j=0}^{m-1} U(t_j)$ it follows that $Z_h$ is stochastic for any $m < \infty$, and

$$mU_0U_0' = m^{-1}\sum_{j=0}^{m-1}\sum_{i=0}^{m-1} U(t_j)U(t_i)' \xrightarrow{p} \Sigma_U(0),$$

and similar $mU_nU_n' \xrightarrow{p} \Sigma_U(1)$. So the result for $h = 0$ follows from $Z_0 = 2(U_0U_0' + U_nU_n')$. Next, for $h > 0$,

$$mU_0U_h' = \textstyle\sum_{j=0}^{m-1}U(t_j)U(t_{m-1+h})' \xrightarrow{p} \sum_{j=0}^{\infty}\gamma_{-j-h}(0) = \displaystyle\sum_{j=0}^{\infty}\gamma_{j+h}(0)',$$

and similarly we find $mU_nU_{n-h}' \xrightarrow{p} \sum_{j=0}^{m-1} \gamma_{j+h}(1)$. □

**Proof of Theorem A.4.** Since $k'(0) = 0$ and $k''(x)$ is continuous we have $k_0 - k_1 = -H^{-2}k''(\epsilon)/2$, for some $0 \leq \epsilon \leq H^{-1}$. Define $a_0 = -k''(\epsilon)$ and $a_h = H^2(-k_{|h|+1} + 2k_{|h|} - k_{|h|-1})$, and write $V$-terms of $K(U)$, see (A.3), as

$$(k_0 - k_1)V_0 - \sum_{h=1}^{n-1}(k_{h+1} - 2k_h + k_{h-1})V_h$$
$$= H^{-2}\sum_{h=-n+1}^{n-1} a_h \sum_j U_jU_{j-h}' = H^{-2}\sum_{|h|\leq\sqrt{H}} a_h \sum_j U_jU_{j-h}' + H^{-2}\sum_{|h|>\sqrt{H}} a_h \sum_j U_jU_{j-h}'.$$

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

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

so that the first term $\frac{n}{H^2}\sum_{|h|\leq\sqrt{H}} a_h \frac{1}{n}\sum_j U_j U'_{j-h} = -k''(0)\frac{n}{H^2}\Omega + o(\frac{n}{H^2})$. The second term vanishes because

$$\frac{n}{H^2}\left|\sum_{|h|>\sqrt{H}} a_h \frac{1}{n}\sum_j U_j U'_{j-h}\right| \leq \frac{n}{H^2}\sum_{|h|>\sqrt{H}}\left|H^2 a_h\right| \cdot \sup_{|h|>\sqrt{H}}\left|\frac{1}{n}\sum_j U_j U'_{j-h}\right|,$$

and $\sup_{|h|>\sqrt{H}}\left|\frac{1}{n}\sum_j U_j U'_{j-h}\right| = o_p(1)$.

For the $Z$-terms we have by Proposition A.2 that $Z_0 = O_p(m^{-1})$, and

$$\sum_{h=1}^{n-1}(k_h - k_{h-1})Z_h = \frac{1}{m}\frac{1}{H}\sum_{h=1}^{n-1}\left\{k'(h/H) + o(1)\right\} m Z_h = O_p(m^{-1}).$$

□

**Proof of Lemma 2.** When $k'(0) \neq 0$ we see that the first term of (A.3) is such that $\frac{H}{n}(k_0 - k_1)V_0 \xrightarrow{p} -k'(0)2\int_0^1\left\{\Sigma_\zeta(u) + \gamma_0(u)\right\}\mathrm{d}u$. From the proof of Theorem A.4 it follows that the other terms in (A.3) are of lower order. □

## A.2. Results concerning $K(Y)$

The aim of this subsection to is prove the following Theorem that concern $K(Y)$ in the univariate case. Then we extend the result to the multivariate case in the next subsection.

**Theorem A.5.** *Suppose* **K**, **SH**, **D**, *and* **U** *hold then as* $n, m, H \to \infty$ *with* $H/n = o(1)$ *and* $m^{-1} = o(\sqrt{H/n})$, *we have*

$$\sqrt{\frac{n}{H}}\left(K(Y) - \int_0^1 \sigma^2(u)\mathrm{d}u\right) \xrightarrow{Ls} \mathrm{MN}\left(0, 4k_\bullet^{0,0}\int_0^1 \sigma^4(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u\right). \tag{A.4}$$

Before we prove this Theorem for $K(Y)$ we introduce and analyze two related quantities,

$$\tilde{K}(Y) = \sum_{i=1}^{N}(\eta_{N,i}^{(1)} + \eta_{N,i}^{(2)}) \qquad \text{and} \qquad \hat{K}(Y) = \sum_{i=1}^{N}(\hat{\eta}_{N,i}^{(1)} + \hat{\eta}_{N,i}^{(2)})$$

where $y_{N,i} = Y(\tau_{N,i}) - Y(\tau_{N,i-1})$ and $\hat{y}_{N,i} = \sigma\left(\tau_{N,i-1}\right)\left(W_{\tau_{N,i}} - W_{\tau_{N,i-1}}\right)$ and

$$\eta_{N,i}^{(1)} = y_{N,i}^2, \quad \hat{\eta}_{N,i}^{(1)} = \hat{y}_{N,i}^2, \quad \eta_{N,i}^{(2)} = 2y_{N,i}\sum_{h=1}^{N-1}k_h y_{i-h}, \quad \hat{\eta}_{N,i}^{(2)} = 2\hat{y}_{N,i}\sum_{h=1}^{N-1}k_h \hat{y}_{N,i-h},$$

$\tilde{K}$ is similar to $K$, except that it is not subjected to the jittering, and $\hat{K}$ is similar to $\tilde{K}$, but is computed with auxiliary intraday returns. Note that we have (uniformly over $i$) the strong approximation (under **SH**)

$$y_{N,i} = \int_{\tau_{N,i-1}}^{\tau_{N,i}}\mu(u)\mathrm{d}u + \int_{\tau_{N,i-1}}^{\tau_{N,i}}\sigma(u)\mathrm{d}W(u) = \hat{y}_{N,i}\left\{1 + o_p(N^{-1/2})\right\}, \tag{A.5}$$

(Jacod (2008, (6.25)) and Phillips & Yu (2008, equation 66))). Let $\varepsilon_{N,i} = \sqrt{\Delta_{N,i}} \left( W_{\tau_{N,i}} - W_{\tau_{N,i-1}} \right)$ so that $\varepsilon_{N,i} \sim iid\, N(0,1)$ and note that $\hat{y}_{N,i} = N^{-1/2}\sigma\left(\tau_{N,i-1}\right) D_{N,i}^{1/2}\varepsilon_{N,i}$. We use $\hat{y}_{N,i}$ as our estimate of $y_{N,i}$ throughout, later showing it makes no impact on the result.

Note that $y_{N,i} - \hat{y}_{N,i} = \int_{\tau_{N,i-j-1}}^{\tau_{N,i-j}} \left\{ \sigma(u) - \sigma(\tau_{N,i-j-1}) \right\} \mathrm{d}W(u)$, so with $\mathrm{d}[\sigma]_t = \lambda_t \mathrm{d}t$ we find

$$\int_{\tau_{N,i-1}}^{\tau_{N,i}} \left\{ \sigma(u) - \sigma(\tau_{N,i-1}) \right\}^2 \mathrm{d}W(u) =$$
$$\lambda^2_{\tau_{N,i-1}} \int_{\tau_{N,i-1}}^{\tau_{N,i}} \left( u - \tau_{N,i-1} \right) \mathrm{d}u \left\{ 1 + o_p(1) \right\} = \frac{\lambda^2_{\tau_{N,i-1}}}{2} \frac{D_{N,i}^2}{N^2} \left\{ 1 + o_p(1) \right\}.$$

**Proposition A.3.** *Suppose* ***K***, ***SH****, and* ***D*** *hold. Then as* $n \to \infty$ *with* $H = o(n)$ *and* $m = O(\sqrt[3]{Hn})$, *then*

$$\sqrt{\frac{n}{H}} \left\{ K(Y) - \tilde{K}(Y) \right\} = o_p(1).$$

**Proof.** The difference between $K(Y)$ and $\tilde{K}(Y)$ is tied to the $m$ first and $m$ last observations. So the difference vanishes if $m$ does not grow at too fast a rate. We have

$$\sum_{i=1}^{m} y_{N,i}^2 = \sum_{i=1}^{m} \frac{D_{N,i}}{N} \sigma^2\left(\tau_{N,i-1}\right) \varepsilon_{N,i}^2 \left\{ 1 + o_p(N^{-1/2}) \right\}^2 = o_p\left(\frac{m^{3/2}}{N}\right),$$

since $\max_{i=1,\ldots,m} D_{N,i} = o_p(m^{1/2})$, $\sigma^2(t)$ is bounded, and $\sum_{i=1}^m \varepsilon_{N,i}^2 = O_p(m)$. So we need $\sqrt{\frac{n}{H}} \frac{m^{3/2}}{N} = O(1)$ which is implied by $m^3/(HN) \leq m^3/(Hn) = O(1)$. □

**Proposition A.4.** *Suppose* ***SH*** *and* ***D*** *hold then, so long as* $H = o(N)$,

$$\sqrt{\frac{N}{H}} \left\{ \tilde{K}(Y) - \hat{K}(Y) \right\} = o_p(1).$$

**Proof.** From, for example, Phillips & Yu (2008) it is known that $\sum_{i=1}^N \eta_{N,i}^{(1)} - \hat{\eta}_{N,i}^{(1)} = o_p(N^{-1/2})$. The only thing left to do is to prove that $\sum_{i=1}^N \eta_{N,i}^{(2)} - \hat{\eta}_{N,i}^{(2)} = o_p(\sqrt{H/N})$. First note that

$$\begin{aligned} \tfrac{1}{2}\sum_{i=1}^{N} \eta_{N,i}^{(2)} - \hat{\eta}_{N,i}^{(2)} &= \sum_{i=1}^{N} y_{N,i} \left( \sum_{h>0} k_h y_{N,i-h} \right) - \sum_{i=1}^{N} \hat{y}_{N,i} \left( \sum_{h>0} k_h \hat{y}_{N,i-h} \right) \\ &= \sum_{i=1}^{N} \left( y_{N,i} - \hat{y}_{N,i} \right) \left( \sum_{h>0} k_h y_{N,i-h} \right) + \sum_{i=1}^{N} \hat{y}_{N,i} \left( \sum_{h>0} k_h \left( y_{N,i-h} - \hat{y}_{N,i-h} \right) \right). \end{aligned} \quad \text{(A.6)}$$

The first term of (A.6) is a sum of martingale difference sequences. Its conditional variance is

$$V = \sum_{i=1}^{N} \frac{\lambda^2_{\tau_{N,i-1}}}{2} \frac{D_{N,i}^2}{N^2} \left( \sum_{h>0} k_h y_{N,i-h} \right)^2 \leq$$
$$\max_{i=1,\ldots,N} D_{N,i}^2 \left( \frac{1}{N} \sum_{i=1}^{N} \frac{\lambda^2_{\tau_{N,i-1}}}{2} \right) \frac{1}{N} \left( \sum_{h>0} k_h y_{N,i-h} \right)^2 = o_p(N) O_p(1) \frac{1}{N} O_p\left(\frac{H}{N}\right) = o_p(H/N),$$

where we have used that $\sum_{h>1} k_h y_{N,i-h} = O_p(\sqrt{H/N})$. The second term is

$$\sum_{i=1}^{N} \hat{y}_{N,i} \sum_{h>0} k_h \left(y_{N,i-h} - \hat{y}_{N,i-h}\right) = \sum_{i=1}^{N} \hat{y}_{N,i} \left(\sum_{h>0} k_h \int_{\tau_{N,i-h-1}}^{\tau_{N,i-h}} \left\{\sigma(u) - \sigma(\tau_{N,i-h-1})\right\} \mathrm{d}W(u)\right).$$

It has a zero conditional means and its conditional variance is

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right) D_{N,i} \sum_{h>0} k_h^2 \int_{\tau_{N,i-h-1}}^{\tau_{N,i-h}} \left\{\sigma(u) - \sigma(\tau_{N,i-h-1})\right\}^2 \mathrm{d}u \\
&= \frac{1}{N}\sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right) D_{N,i} \sum_{h>0} k_h^2 \lambda^2_{\tau_{N,i-h-1}} N^{-2} D^2_{N,i-h}/2 \left\{1 + o_p(1)\right\}, , \quad \text{where} \quad \mathrm{d}[\sigma]_t = \lambda_t \mathrm{d}t \\
&\le \frac{1}{N^3}\sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right) \max_{i=1,\ldots,N} D_{N,i} \frac{1}{2} \max_i \lambda_{N,i} \sum_{j=0}^{\lfloor N/q \rfloor} \max_{i=qj+1,\ldots,qj+q} D^2_{N,i} \sum_{h>qj}^{q(j+1)} k_h^2 \left\{1 + o_p(1)\right\} \\
&\le \frac{1}{N^3}\sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right) o_p(N^{1/2}) O_p(1) o_p(q) H o(\log \tfrac{N}{q}) \\
&= o_p(\frac{H}{N}), \qquad\qquad \text{take } q = N^{1/2}/\log(N).
\end{aligned}$$

Here we have used that $\max_{h\in\{c+1,\ldots,c+m\}} D_{N,i} = o_p(m^{1/2})$, that $\frac{1}{H}\sum_{h>qj}^{qj+q} k(\frac{h}{H})^2$ is at most of order $o(j^{-1})$, since $\frac{1}{H}\sum_{h=1}^{N} k(\frac{h}{H})^2 \simeq \sum_{j=0}^{\lfloor N/q \rfloor} \frac{1}{H}\sum_{h>q(j-1)}^{qj} k(\frac{h}{H})^2$ is convergent, and that $\sum_{j=1}^{m}\frac{1}{j} = O(\log m)$. □

**Proposition A.5.** *Suppose* ***SH*** *and* ***D*** *hold and $H = o(N)$, then as $N \to \infty$*

$$\sqrt{\frac{N}{H}}\left(\hat{K}(Y) - \int_0^1 \sigma^2(u)\mathrm{d}u\right) \xrightarrow{Ls} \mathrm{MN}\left(0, 4k_\bullet^{0,0}\int_0^1 \sigma^4(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u\right).$$

**Proof.** We have $\hat{K}(Y) = \sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right)\Delta_{N,i} + \hat{\eta}^{(1)}_{N,i} + \hat{\eta}^{(2)}_{N,i}$. Phillips & Yu (2008) imply that $\sqrt{\frac{N}{H}}\left(\sum_{i=1}^{N} \sigma^2\left(\tau_{N,i-1}\right)\Delta_{N,i} - \int_0^1 \sigma^2(u)\,\mathrm{d}u\right) = o_p(\frac{1}{\sqrt{H}})$. This means that

$$\sqrt{\frac{N}{H}}\left(\hat{K}(Y) - \int_0^1 \sigma^2(u)\,\mathrm{d}u\right) = \sqrt{\frac{N}{H}}\sum_{i=1}^{N}(\hat{\eta}^{(1)}_{N,i} + \hat{\eta}^{(2)}_{N,i}) + o_p(1),$$

which is the sum of the martingale differences: $\{\hat{\eta}^{(1)}_{N,i} + \hat{\eta}^{(2)}_{N,i}, \mathcal{F}_{\tau_{N,i}}\}$. So we just need to compute its contributions to the conditional variance.

The first term, $\hat{\eta}^{(1)}_{N,i}$, is the sampling error of the well known realized variance. In the present context, it was studied in Phillips & Yu (2008), and it follows that $\sqrt{\frac{N}{H}}\sum_{i=1}^{N}\hat{\eta}^{(1)}_{N,i} = O_p(H^{-1})$. This means that unless $H = O(1)$ this term will be asymptotically irrelevant for the realised kernel. Next

$$\frac{N}{H}(\hat{\eta}^{(2)}_{N,i})^2 = 4N\hat{y}^2_{N,i}\left(H^{-1}\sum_{h>1} k_h \hat{y}_{N,i-h}\right)^2,$$

where $\left(H^{-1}\sum_{h>1} k_h \hat{y}_{N,i-h}\right)^2 \xrightarrow{p} k_\bullet^{0,0}\sigma^2(\tau_{N,i})$. Since $N\hat{y}^2_{N,i} = D_{N,i}\sigma^2(\tau_{N,i})\varepsilon^2_{N,i}$ we have

$$\frac{N}{H}\sum_{i=1}^{N}\mathrm{Var}(\hat{\eta}^{(2)}_{N,i}|\mathcal{F}_{\tau_{N,i-1}}) = \frac{k_\bullet^{0,0}}{N}\sum_{i=1}^{N}\sigma^4\left(\tau_{N,i-1}\right)\mathrm{E}\left(D^2_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right) + o_p(1).$$

Now we follow Phillips & Yu (2008) and write

$$\mathrm{E}\left(D_{N,i}^2|\mathcal{F}_{\tau_{N,i-1}}\right) = D_{N,i}\frac{\mathrm{E}\left(D_{N,i}^2|\mathcal{F}_{\tau_{N,i-1}}\right)}{\mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)} - \left\{D_{N,i} - \mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)\right\}\frac{\mathrm{E}\left(D_{N,i}^2|\mathcal{F}_{\tau_{N,i-1}}\right)}{\mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)}.$$

Now

$$\frac{k_\bullet^{0,0}}{N}\sum_{i=1}^N \sigma^4\left(\tau_{N,i-1}\right)\left\{D_{N,i} - \mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)\right\}\frac{\mathrm{E}\left(D_{N,i}^2|\mathcal{F}_{\tau_{N,i-1}}\right)}{\mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)} = o_p(1),$$

as this is a temporal average of a martingale difference sequence. This means that

$$\begin{aligned}\frac{N}{H}\sum_{i=1}^N \mathrm{Var}(\hat{\eta}_{N,i}^{(2)}|\mathcal{F}_{\tau_{N,i-1}}) &= k_\bullet^{0,0}\sum_{i=1}^N \sigma^4\left(\tau_{N,i-1}\right)\Delta_{N,i}\frac{\mathrm{E}\left(D_{N,i}^2|\mathcal{F}_{\tau_{N,i-1}}\right)}{\mathrm{E}\left(D_{N,i}|\mathcal{F}_{\tau_{N,i-1}}\right)} + o_p(1),\\ &= k_\bullet^{0,0}\int_0^1 \sigma_u^4\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u + o_p(1),\end{aligned}$$

by Riemann integration. The results then follows by the martingale array CLT. □

**Proof of Theorem A.5.** Follows by combining the results of Propositions A.3, A.4, and A.5. □

### A.3. **Multivariate Results**

**Proof of Lemma 1.** The results (2) and (4) follow by combining Theorem 1 with Proposition A.4 and Theorem A.5. From the proof of Theorem 1 we have $K(X) = K(Y) + K(U) + O_p(\sqrt{K(U)})$, and (3) follows since $K(Y) \xrightarrow{p} [Y]$ and $K(U) \xrightarrow{p} \frac{-k''(0)}{c_0^2}\Omega$ when $H = c_0 n^{1/2}$. □

**Proof of Theorem 3.** We analyse the joint characteristic function of the realised kernel matrix

$$\mathrm{E}\exp\left[i\,\mathrm{tr}\left\{AK(X)\right\}\right] = \mathrm{E}\exp\left[i\sum_{j=1}^d \lambda_j \mathrm{tr}\left\{K(X)a_j a_j'\right\}\right].$$

where $A = \sum_{j=1}^d \lambda_j a_j a_j'$ is symmetric matrix of constants[8]. Hence it is sufficient for us to study the joint law of $a_j' K(X) a_j$, for any fixed $a_j$, $j = 1, \ldots, d$. This is a convenient form as $a_j' K(X) a_j = K(a_j' X)$, the univariate kernel applied to the process $a_j' X$. This is very convenient as $a_j' X$ is simple a univariate process in our class.

The univariate results imply that the only thing left to study is the joint distribution of $K(a_j' Y)$. Now under the conditions of the Theorem with $n, H \to \infty$ and $H \propto n^\eta$ for $\eta \in (0,1)$ and $m/n \to 0$, we will establish that

$$\sqrt{\frac{n}{H}}\left(K(Y) - \int_0^t \Sigma(u)\mathrm{d}u\right) \xrightarrow{Ls} \mathrm{MN}\left(0, 4k_\bullet^{0,0}\int_0^1 \Psi(u)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u\right). \tag{A.7}$$

where $\Psi(u) = \Sigma(u)\otimes\Sigma(u)$. This will then complete the theorem. The univariate proof implies we that can replace $K(a_j' Y)$ by $\hat{K}(a_j' Y) = \sum_{i=1}^n x_{n,j,i}^2 + 2\sum_{i=1}^n x_{n,j,i}\sum_{h=1}^{n-1} k_h x_{n,j,i-h}$, where $x_{n,j,i} = a_j'\sigma\left(\tau_{n,i-1}\right)\Delta_{n,i}^{1/2}\varepsilon_{n,i}$

[8] It is well known that the distribution is characterised by the matrix characteristic function $\mathrm{E}\exp\left[i\,\mathrm{tr}\left\{A'K(X)\right\}\right]$. Without loss of generality we can assume $A$ is symmetric as $K(X)$ is symmetric and $tr(A'K(X)) = \sum_{i,j} a_{ij}K(X)_{ji} = \sum_i a_{ii}K(X)_{ii} + \sum_{i<j}(a_{ij} + a_{ji})K(X)_{ij} = \sum_{i,j}(a_{ij} + a_{ji})/2K(X)_{ij} = \frac{1}{2}tr\left\{(A + A')'K(X)\right\}$.

and $\varepsilon_{n,i} = W_{\tau_{n,i}} - W_{\tau_{n,i-1}}$. But this raises no new principles and so we can see that using the same method as before

$$\sqrt{\frac{n}{H}}\left[\left\{\begin{array}{c} K(a_j' Y) \\ K(a_k' Y) \end{array}\right\} - \left(\begin{array}{c} a_j' \\ a_k' \end{array}\right)\left\{\int_0^t \Sigma(u)\mathrm{d}u\right\}\left(\begin{array}{cc} a_j & a_k \end{array}\right)\right]$$
$$\xrightarrow{Ls} \mathrm{MN}\left(0, 4k_\bullet^{0,0}\int_0^1 \left(\begin{array}{cc} \left(a_j'\Sigma(u)a_j\right)^2 & \left(a_j'\Sigma(u)a_k\right)^2 \\ \left(a_j'\Sigma(u)a_k\right)^2 & \left(a_j'\Sigma(u)a_j\right)^2 \end{array}\right)\frac{\varkappa_2(u)}{\varkappa_1(u)}\mathrm{d}u\right).$$

Unwrapping the results delivers (A.7) as required. □

**Proof of Theorem 2.** Follows as a corollary to Theorem 3. □

### A.4. **Optimal choice of bandwidth**

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

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

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

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

and the asymptotic variance is

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

## Appendix B: Errors induced by stale prices

The stale prices induce a particular form of noise with an endogenous component. The price indexed by time $\tau_j$ 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

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

The endogenous component that is induced by refresh time is $Y^{(i)}(\tau_j) - Y(t_j^{(i)})$, But this is exactly the sort of dependence that Assumption **U** can accommodate through correlation between $\tilde{W}$ and $W$, and the (random) coefficients $\psi_h(t)$, $h = 0, 1, \ldots$.

## List of Figures

[tbh]

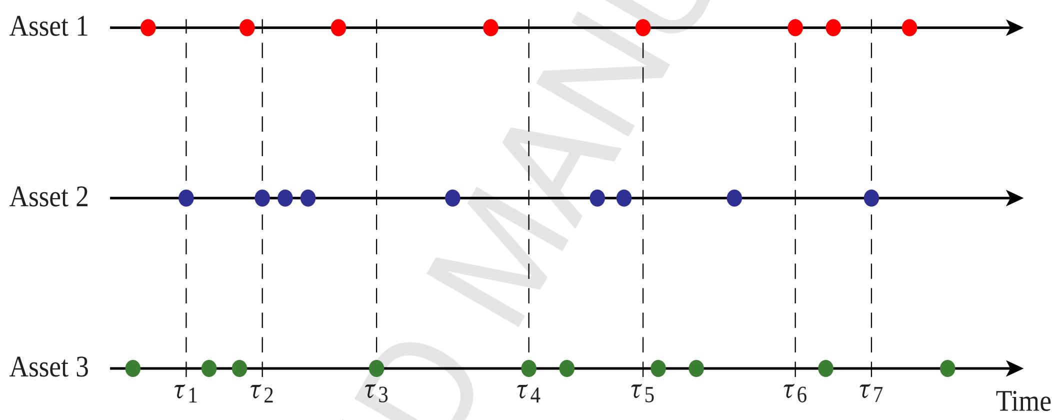


Figure 1: *This figure illustrates Refresh Time in a situation with three assets. The dots represent the times* $\{t_j^{(i)}\}$. *The vertical lines represent the sampling times generated from the three assets with refresh time sampling. Note, in this example,* $N = 7$, $n^{(1)} = 8$, $n^{(2)} = 9$ and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

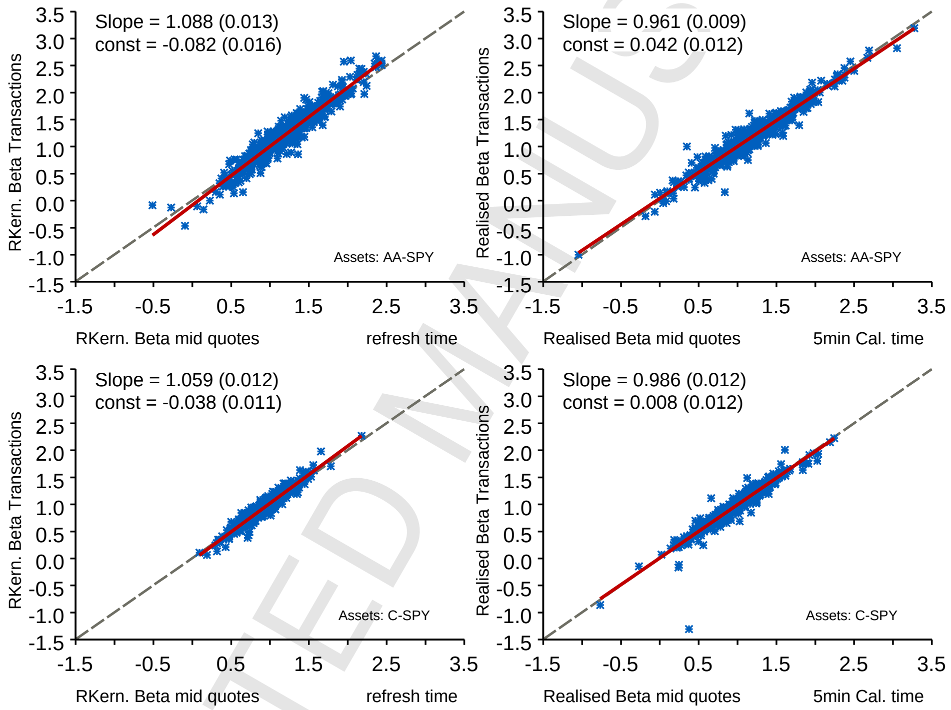


Figure 2: Scatter plots for daily realised kernel betas for the AA and SPY asset combination.

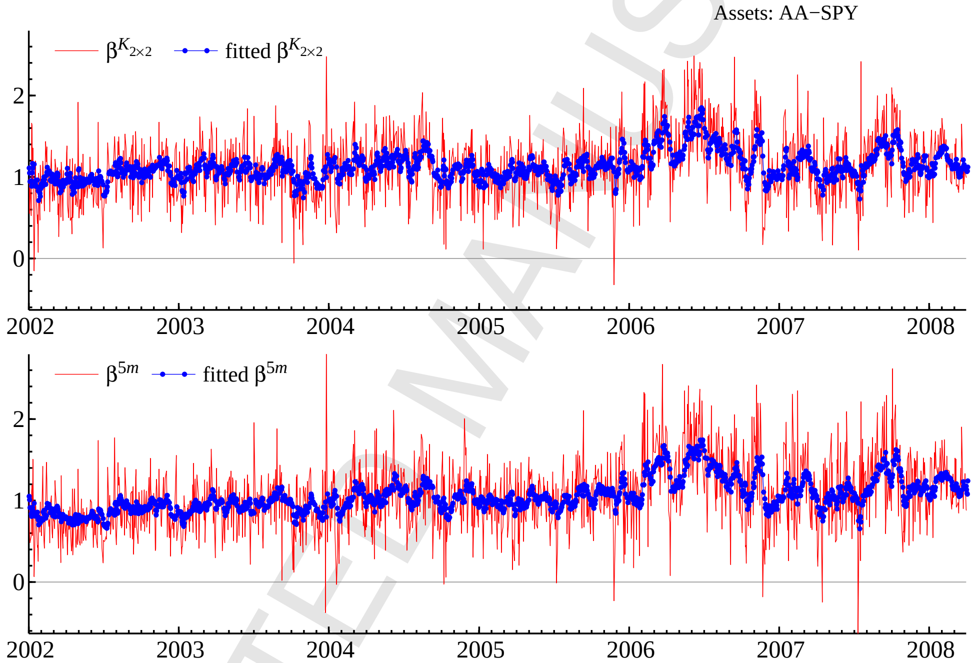


Figure 3: ARMA(1,1) model for transaction based realised kernel betas for the AA and SPY combination.

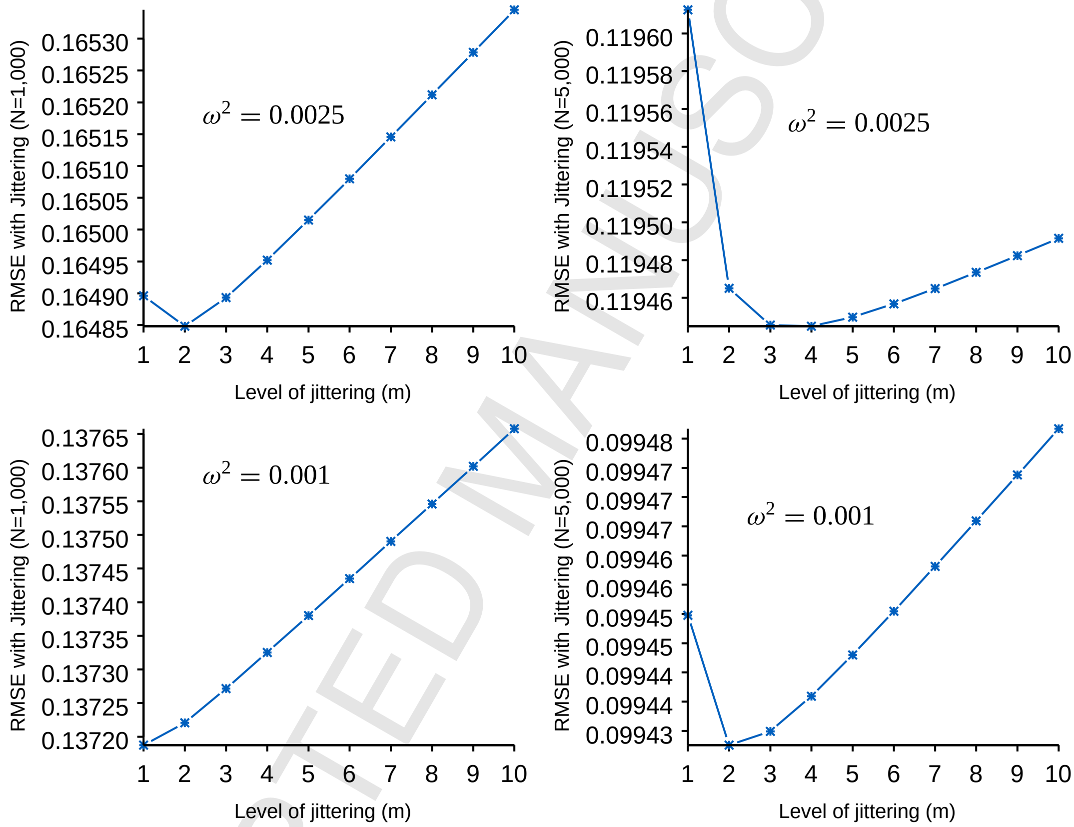


Figure 4: Sensitivity to the the choice of $m$. The Figure shows the RMSE as a function of $m$ for the sample sizes $N = 1,000$ and $N = 5,000$, and $\omega^2 = 0.001$ and $\omega^2 = 0.0025$.

## List of Tables

Table 1: Properties of some realised kernels.

| | Kernel function, $k(x)$ | | $\lvert k''(0)\rvert$ | $k_{\bullet}^{0,0}$ | $c^*$ | $\left\lvert k''(0)(k_{\bullet}^{0,0})^2\right\rvert^{1/5}$ |
|---|---|---|---|---|---|---|
| Parzen | $k(x)=\begin{cases}1-6x^2+6x^3 & 0\le x\le \frac{1}{2}\\ 2(1-x)^3 & \frac{1}{2}\le x\le 1\\ 0 & x>1\end{cases}$ | | 12 | 0.269 | 3.51 | 0.97 |
| Quadratic Spectral | $k(x)=\frac{3}{x^2}\left(\frac{\sin x}{x}-\cos x\right)$ | $x\ge 0$ | 1/5 | $3\pi/5$ | 0.46 | 0.93 |
| Fejér | $k(x)=\left(\frac{\sin x}{x}\right)^2$ | $x\ge 0$ | 2/3 | $\pi/3$ | 0.84 | 0.94 |
| Tukey-Hanning$_\infty$ | $k(x)=\sin^2\left(\frac{\pi}{2}e^{-x}\right)$ | $x\ge 0$ | $\pi^2/2$ | 0.52 | 2.16 | 1.06 |
| BNHLS (2008) | $k(x)=(1+x)\,e^{-x}$ | $x\ge 0$ | 1 | 5/4 | 0.96 | 1.09 |

$|k''(0)(k_{\bullet}^{0,0})^2|^{1/5}$ *measures the relative asymptotic efficiency of* $k\in\mathcal{K}$.

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 results are based on 1000 repetitions.*

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

| $2 \times 2$ *case* | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | | 0.589 | 0.599 | 0.595 | 0.586 | 0.565 | 0.593 | 0.582 | 0.602 | 0.574 |
| AIG | 0.635 | | 0.607 | 0.596 | 0.616 | 0.608 | 0.584 | 0.605 | 0.598 | 0.601 |
| AXP | 0.639 | 0.651 | | 0.613 | 0.617 | 0.595 | 0.596 | 0.594 | 0.619 | 0.584 |
| BA | 0.634 | 0.642 | 0.652 | | 0.596 | 0.577 | 0.595 | 0.591 | 0.609 | 0.582 |
| BAC | 0.636 | 0.656 | 0.662 | 0.649 | | 0.631 | 0.579 | 0.612 | 0.597 | 0.613 |
| C | 0.635 | 0.657 | 0.660 | 0.647 | 0.680 | | 0.554 | 0.610 | 0.575 | 0.619 |
| CAT | 0.630 | 0.631 | 0.641 | 0.636 | 0.633 | 0.630 | | 0.575 | 0.597 | 0.569 |
| CVX | 0.641 | 0.657 | 0.659 | 0.651 | 0.671 | 0.675 | 0.634 | | 0.588 | 0.618 |
| DD | 0.639 | 0.646 | 0.656 | 0.649 | 0.653 | 0.651 | 0.639 | 0.652 | | 0.575 |
| SPY | 0.609 | 0.642 | 0.630 | 0.622 | 0.667 | 0.685 | 0.602 | 0.670 | 0.625 | |

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

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| Trades | 3,442 | 4,228 | 3,461 | 3,529 | 4,544 | 5,480 | 3,330 | 4,845 | 3,307 | 5,412 |
| Quotes | 8,460 | 9,270 | 8,626 | 8,553 | 10,091 | 10,809 | 8,026 | 10,254 | 8,521 | 15,973 |

*Summary statistics for the refresh sampling scheme. In the upper panel we present averages over the daily data of the data maintained 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.*

Table 4: Average high frequency realised covariance and open to close covariance

*Average of Hayashi-Yoshida covariances (all times)*

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | 4.331 | **0.415** | **0.484** | **0.383** | **0.372** | **0.446** | **0.424** | **0.360** | **0.465** | **0.349** |
| AIG | **0.225** | 2.904 | **0.612** | **0.387** | **0.520** | **0.597** | **0.415** | **0.332** | **0.415** | **0.361** |
| AXP | **0.251** | **0.315** | 3.383 | **0.447** | **0.580** | **0.661** | **0.476** | **0.382** | **0.477** | **0.430** |
| BA | **0.197** | **0.206** | **0.230** | 3.274 | **0.344** | **0.400** | **0.393** | **0.319** | **0.389** | **0.336** |
| BAC | **0.198** | **0.275** | **0.299** | **0.177** | 2.466 | **0.596** | **0.374** | **0.315** | **0.375** | **0.345** |
| C | **0.237** | **0.305** | **0.319** | **0.199** | **0.303** | 4.014 | **0.431** | **0.354** | **0.430** | **0.382** |
| CAT | **0.227** | **0.224** | **0.254** | **0.210** | **0.197** | **0.220** | 2.274 | **0.332** | **0.413** | **0.341** |
| CVX | **0.194** | **0.196** | **0.219** | **0.176** | **0.180** | **0.190** | **0.187** | 2.192 | **0.340** | **0.305** |
| DD | **0.240** | **0.225** | **0.251** | **0.206** | **0.196** | **0.224** | **0.222** | **0.186** | 2.643 | **0.342** |
| SPY | **0.149** | **0.152** | **0.179** | **0.143** | **0.142** | **0.147** | **0.151** | **0.141** | **0.142** | 1.068 |

*Open-to-close covariance*

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | | 1.100 | 1.228 | 1.091 | 1.019 | 1.267 | 1.377 | 1.027 | 1.219 | 1.036 |
| AIG | 1.092 | | 1.776 | 0.978 | 1.808 | 2.076 | 1.062 | 0.675 | 1.076 | 1.120 |
| AXP | 1.220 | 1.763 | | 1.045 | 1.661 | 2.055 | 1.191 | 0.853 | 1.101 | 1.180 |
| BA | 1.075 | 0.976 | 1.049 | | 0.882 | 1.071 | 0.990 | 0.630 | 0.850 | 0.820 |
| BAC | 1.015 | 1.804 | 1.653 | 0.887 | | 2.036 | 0.967 | 0.619 | 0.927 | 0.988 |
| C | 1.265 | 2.088 | 2.053 | 1.077 | 2.041 | | 1.230 | 0.819 | 1.134 | 1.261 |
| CAT | 1.365 | 1.045 | 1.178 | 0.981 | 0.965 | 1.233 | | 0.768 | 1.005 | 0.933 |
| CVX | 1.026 | 0.671 | 0.862 | 0.627 | 0.624 | 0.824 | 0.771 | | 0.658 | 0.705 |
| DD | 1.211 | 1.070 | 1.092 | 0.850 | 0.923 | 1.134 | 0.996 | 0.656 | | 0.829 |
| SPY | 1.029 | 1.120 | 1.177 | 0.816 | 0.985 | 1.268 | 0.927 | 0.708 | 0.829 | |

*The upper panel presents average estimates for* $\mathrm{Cov}_s^{HY}$ *and the lower panel displays these for* $\mathrm{Cov}_s^{OtoC}$. *In both panels the upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. The diagonal elements are computed with transaction prices. In the upper panel numbers outside the diagonal are boldfaced if the bias is significant at the 1 percent level.*

Table 5: Averages for alternative integrated covariance estimators

*Average of Parzen covariances (30×30)*

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | 3.338 | 0.945 | 1.033 | **0.808** | 0.865 | 1.118 | **1.055** | 0.818 | 1.047 | **0.862** |
| AIG | 0.933 | 2.725 | **1.300** | **0.722** | 1.254 | **1.556** | 0.856 | 0.533 | **0.820** | **0.878** |
| AXP | 1.003 | **1.273** | 2.580 | **0.780** | 1.326 | **1.656** | **0.974** | **0.578** | **0.900** | **0.930** |
| BA | **0.792** | **0.730** | **0.773** | 2.194 | **0.654** | **0.831** | **0.765** | **0.488** | 0.702 | **0.670** |
| BAC | 0.850 | 1.226 | 1.305 | **0.658** | 2.057 | 1.681 | 0.805 | 0.485 | 0.768 | **0.811** |
| C | 1.102 | **1.528** | **1.627** | **0.840** | 1.649 | 2.942 | 1.011 | 0.638 | 0.986 | **1.057** |
| CAT | **1.028** | 0.841 | **0.939** | **0.760** | 0.792 | 1.001 | 2.225 | **0.603** | 0.882 | **0.776** |
| CVX | 0.822 | 0.534 | **0.570** | **0.500** | 0.489 | 0.640 | **0.607** | 1.655 | 0.576 | 0.589 |
| DD | 1.049 | **0.813** | **0.890** | 0.706 | 0.759 | 0.984 | 0.874 | 0.589 | 1.810 | 0.736 |
| SPY | **0.862** | **0.876** | **0.918** | **0.678** | **0.808** | **1.058** | **0.775** | 0.602 | 0.744 | 0.745 |

*Average of Parzen covariances (10×10)*

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | 3.400 | 0.953 | 1.024 | **0.808** | 0.863 | 1.121 | **1.040** | 0.825 | 1.046 | **0.868** |
| AIG | 0.931 | 2.766 | **1.298** | **0.734** | 1.248 | **1.553** | 0.860 | 0.544 | **0.829** | **0.885** |
| AXP | 0.995 | **1.271** | 2.593 | **0.778** | 1.315 | **1.642** | **0.957** | **0.581** | **0.900** | **0.926** |
| BA | **0.798** | **0.739** | **0.783** | 2.222 | **0.656** | **0.845** | **0.772** | **0.501** | 0.712 | **0.679** |
| BAC | 0.844 | 1.204 | **1.290** | **0.660** | 2.080 | 1.669 | 0.799 | 0.493 | 0.765 | **0.811** |
| C | 1.101 | **1.518** | **1.616** | **0.854** | 1.625 | 2.985 | 1.013 | 0.653 | 0.992 | 1.062 |
| CAT | **1.019** | 0.843 | **0.931** | **0.767** | 0.785 | 1.003 | 2.234 | **0.613** | 0.879 | **0.779** |
| CVX | 0.824 | 0.544 | **0.583** | 0.512 | 0.501 | 0.656 | **0.619** | 1.682 | 0.592 | 0.603 |
| DD | 1.045 | **0.819** | **0.892** | 0.719 | 0.755 | 0.991 | 0.876 | 0.603 | 1.850 | 0.745 |
| SPY | **0.865** | **0.876** | **0.917** | **0.688** | **0.806** | **1.061** | **0.779** | 0.616 | 0.752 | 0.755 |

*Average of Parzen covariances (2×2)*

| | AA | AIG | AXP | BA | BAC | C | CAT | CVX | DD | SPY |
|---|---|---|---|---|---|---|---|---|---|---|
| AA | 3.531 | 0.931 | 0.986 | **0.812** | 0.837 | 1.098 | **1.004** | 0.824 | 1.034 | **0.867** |
| AIG | 0.912 | 2.803 | **1.273** | **0.775** | 1.180 | **1.490** | 0.863 | 0.583 | **0.847** | **0.891** |
| AXP | **0.956** | **1.238** | 2.707 | **0.832** | **1.268** | **1.599** | **0.945** | **0.633** | 0.919 | **0.941** |
| BA | **0.793** | 0.773 | **0.825** | 2.371 | **0.679** | 0.892 | **0.800** | 0.560 | 0.753 | 0.731 |
| BAC | 0.820 | 1.156 | **1.237** | **0.683** | 2.096 | 1.565 | 0.786 | 0.541 | 0.763 | 0.811 |
| C | 1.078 | **1.469** | **1.568** | 0.893 | 1.550 | 3.108 | 1.011 | 0.719 | 0.995 | 1.066 |
| CAT | **0.975** | 0.851 | **0.914** | **0.794** | 0.775 | 1.005 | 2.299 | 0.642 | 0.883 | **0.798** |
| CVX | 0.810 | 0.595 | **0.642** | 0.577 | 0.553 | 0.723 | 0.648 | 1.738 | 0.637 | 0.662 |
| DD | 1.018 | **0.842** | **0.904** | 0.759 | 0.760 | 0.999 | 0.873 | 0.644 | 1.961 | 0.774 |
| SPY | **0.853** | **0.875** | **0.910** | 0.726 | **0.795** | 1.054 | **0.791** | 0.664 | 0.768 | 0.783 |

*The upper panel presents average estimates for* $\mathrm{Cov}_s^{K_{30\times30}}$*, the middle panel for* $\mathrm{Cov}_s^{K_{10\times10}}$*, and the lower panel gives results for* $\mathrm{Cov}_s^{K_{2\times2}}$*. In both panels the upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. The diagonal elements are computed with transaction prices. Outside the diagonals numbers are boldfaced if the bias is significant at the 1 percent level.*

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* | | | | | | | | | | |
| $Cov^{K_{30\times30}}$ | 0.8844 [ 0.089] | 1.607 | -0.229 | 1.000 | 0.67 | 0.58 | 0.52 | 0.45 | 0.44 | 0.35 |
| $Cov^{K_{10\times10}}$ | 0.8862 [ 0.089] | 1.596 | -0.227 | 0.992 | 0.69 | 0.61 | 0.54 | 0.47 | 0.45 | 0.36 |
| $Cov^{K_{2\times2}}$ | 0.8900 [ 0.088] | 1.518 | -0.223 | 0.960 | 0.75 | 0.66 | 0.59 | 0.53 | 0.51 | 0.42 |
| $Cov^{HY}$ | 0.2113 [ 0.022] | 0.362 | -0.902 | 0.767 | 0.80 | 0.73 | 0.68 | 0.64 | 0.62 | 0.53 |
| $Cov^{1/4m}$ | 0.4534 [ 0.050] | 0.805 | -0.660 | 0.660 | 0.84 | 0.74 | 0.68 | 0.64 | 0.61 | 0.52 |
| $Cov^{5m}$ | 0.8505 [ 0.085] | 1.511 | -0.262 | 0.942 | 0.71 | 0.62 | 0.54 | 0.48 | 0.46 | 0.37 |
| $Cov^{30m}$ | 0.9049 [ 0.091] | 1.838 | -0.208 | 0.866 | 0.49 | 0.46 | 0.37 | 0.34 | 0.35 | 0.25 |
| $Cov^{3h}$ | 0.9566 [ 0.105] | 2.659 | -0.156 | 0.640 | 0.22 | 0.25 | 0.20 | 0.17 | 0.16 | 0.15 |
| $Cov^{OtoC}$ | 1.1116 [ 0.150] | 4.255 | | 0.508 | 0.12 | 0.15 | 0.17 | 0.14 | 0.08 | 0.15 |
| *Summary stats for correlations* | | | | | | | | | | |
| $Corr^{K_{30\times30}}$ | 0.3862 [ 0.008] | 0.203 | | 1.000 | 0.28 | 0.26 | 0.23 | 0.23 | 0.22 | 0.19 |
| $Corr^{K_{10\times10}}$ | 0.3825 [ 0.008] | 0.188 | | 0.975 | 0.32 | 0.29 | 0.26 | 0.26 | 0.25 | 0.22 |
| $Corr^{K_{2\times2}}$ | 0.3698 [ 0.007] | 0.155 | | 0.824 | 0.44 | 0.40 | 0.37 | 0.35 | 0.34 | 0.30 |
| $Corr^{1/4m}$ | 0.1836 [ 0.007] | 0.113 | | 0.277 | 0.78 | 0.75 | 0.72 | 0.71 | 0.70 | 0.66 |
| $Corr^{5m}$ | 0.3619 [ 0.007] | 0.169 | | 0.756 | 0.34 | 0.31 | 0.28 | 0.26 | 0.24 | 0.22 |
| $Corr^{30m}$ | 0.4030 [ 0.010] | 0.298 | | 0.684 | 0.14 | 0.13 | 0.11 | 0.12 | 0.11 | 0.10 |
| $Corr^{3h}$ | 0.3869 [ 0.016] | 0.594 | | 0.347 | 0.03 | 0.03 | 0.03 | 0.04 | 0.02 | 0.02 |
| Average unconditional Open-to-Close correlation = 0.5185 | | | | | | | | | | |
| | | | | *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* | | | | | | | | | | |
| $Cov^{K_{30\times30}}$ | 0.8917 [ 0.089] | 1.656 | -0.221 | 1.000 | 0.62 | 0.55 | 0.48 | 0.43 | 0.42 | 0.32 |
| $Cov^{K_{10\times10}}$ | 0.8940 [ 0.090] | 1.636 | -0.219 | 0.992 | 0.66 | 0.58 | 0.51 | 0.45 | 0.44 | 0.34 |
| $Cov^{K_{2\times2}}$ | 0.9000 [ 0.089] | 1.546 | -0.213 | 0.941 | 0.74 | 0.66 | 0.59 | 0.53 | 0.51 | 0.41 |
| $Cov^{HY}$ | 0.4144 [ 0.038] | 0.627 | -0.699 | 0.788 | 0.82 | 0.74 | 0.69 | 0.65 | 0.62 | 0.53 |
| $Cov^{1/4m}$ | 0.4470 [ 0.048] | 0.776 | -0.666 | 0.669 | 0.83 | 0.74 | 0.68 | 0.64 | 0.61 | 0.52 |
| $Cov^{5m}$ | 0.8530 [ 0.084] | 1.481 | -0.260 | 0.922 | 0.72 | 0.61 | 0.55 | 0.50 | 0.47 | 0.39 |
| $Cov^{30m}$ | 0.9056 [ 0.091] | 1.833 | -0.207 | 0.897 | 0.50 | 0.46 | 0.37 | 0.34 | 0.35 | 0.25 |
| $Cov^{3h}$ | 0.9574 [ 0.105] | 2.661 | -0.156 | 0.672 | 0.22 | 0.25 | 0.21 | 0.17 | 0.16 | 0.16 |
| $Cov^{OtoC}$ | 1.1143 [ 0.150] | 4.234 | | 0.534 | 0.12 | 0.15 | 0.18 | 0.14 | 0.08 | 0.15 |
| *Summary stats for correlations* | | | | | | | | | | |
| $Corr^{K_{30\times30}}$ | 0.3904 [ 0.009] | 0.221 | | 1.000 | 0.25 | 0.23 | 0.21 | 0.20 | 0.20 | 0.18 |
| $Corr^{K_{10\times10}}$ | 0.3870 [ 0.008] | 0.200 | | 0.968 | 0.30 | 0.27 | 0.24 | 0.24 | 0.23 | 0.20 |
| $Corr^{K_{2\times2}}$ | 0.3763 [ 0.008] | 0.165 | | 0.818 | 0.41 | 0.37 | 0.34 | 0.32 | 0.31 | 0.27 |
| $Corr^{1/4m}$ | 0.1815 [ 0.006] | 0.103 | | 0.282 | 0.75 | 0.71 | 0.69 | 0.67 | 0.66 | 0.62 |
| $Corr^{5m}$ | 0.3650 [ 0.007] | 0.168 | | 0.724 | 0.35 | 0.30 | 0.28 | 0.26 | 0.25 | 0.22 |
| $Corr^{30m}$ | 0.4027 [ 0.010] | 0.299 | | 0.734 | 0.14 | 0.13 | 0.11 | 0.12 | 0.11 | 0.10 |
| $Corr^{3h}$ | 0.3873 [ 0.016] | 0.593 | | 0.382 | 0.03 | 0.04 | 0.03 | 0.04 | 0.02 | 0.02 |
| Average unconditional Open-to-Close correlation = 0.5169 | | | | | | | | | | |

*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.*

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

*Panel A: 30×30 case*

| $H_{s-1}$ | $r_{s-1}(r_{s-1})'$ | $\mathrm{K}_{t-1}$ | $\mathrm{RV}_{s-1}^{5m}$ | $\log L$ |
|---|---|---|---|---|
| 0.943 (0.006) | 0.005 (0.000) | 0.040 (0.004) | – | -27029.9 |
| 0.742 (0.014) | 0.013 (0.001) | – | 0.115 (0.006) | -27077.7 |
| 0.984 (0.001) | 0.008 (0.000) | – | – | -28477.5 |
| 0.777 (0.012) | – | 0.076 (0.004) | 0.061 (0.005) | -26948.3 |
| 0.784 (0.013) | 0.009 (0.001) | 0.067 (0.004) | 0.059 (0.005) | -26904.3 |

*Panel B: 10×10 case*

| $H_{s-1}$ | $r_{s-1}(r_{s-1})'$ | $\mathrm{K}_{t-1}$ | $\mathrm{RV}_{s-1}^{5m}$ | $\log L$ |
|---|---|---|---|---|
| 0.768 (0.017) | 0.015 (0.003) | 0.151 (0.011) | – | -7920.7 |
| 0.687 (0.025) | 0.022 (0.003) | – | 0.160 (0.013) | -7935.9 |
| 0.965 (0.003) | 0.023 (0.001) | – | – | -8307.5 |
| 0.705 (0.023) | – | 0.126 (0.014) | 0.067 (0.014) | -7923.3 |
| 0.716 (0.023) | 0.017 (0.003) | 0.106 (0.014) | 0.065 (0.013) | -7903.0 |

*Panel C: 2×2 cases*

*AIG-CAT*

| $H_{s-1}$ | $r_{s-1}(r_{s-1})'$ | $\mathrm{K}_{t-1}$ | $\mathrm{RV}_{s-1}^{5m}$ | $\log L$ |
|---|---|---|---|---|
| 0.837 (0.028) | 0.038 (0.008) | 0.126 (0.030) | – | -2584.9 |
| 0.863 (0.023) | 0.044 (0.007) | – | 0.098 (0.025) | -2591.2 |
| 0.951 (0.006) | 0.045 (0.005) | – | – | -2629.7 |
| 0.764 (0.036) | – | 0.236 (0.063) | 0 (–) | -2592.4 |
| 0.837 (0.028) | 0.038 (0.008) | 0.126 (0.050) | 0 (–) | -2584.9 |

*BA-SPY*

| $H_{s-1}$ | $r_{s-1}(r_{s-1})'$ | $\mathrm{K}_{t-1}$ | $\mathrm{RV}_{s-1}^{5m}$ | $\log L$ |
|---|---|---|---|---|
| 0.844 (0.043) | 0.031 (0.008) | 0.094 (0.032) | – | -1516.9 |
| 0.843 (0.041) | 0.032 (0.008) | – | 0.091 (0.030) | -1517.8 |
| 0.958 (0.006) | 0.036 (0.005) | – | – | -1544.4 |
| 0.717 (0.074) | – | 0.125 (0.071) | 0.083 (0.065) | -1521.1 |
| 0.837 (0.045) | 0.031 (0.009) | 0.068 (0.047) | 0.031 (0.045) | -1516.6 |

*Estimation results for scalar BEKK models for close-to-close $d = 30, 10, 2$ dimensional return vectors.*

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

| | $\omega^2 = 0.001$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | normalised bias$^2$ | | | normalised variance | | | normalised mse | | |
| $n$ | $K_B^F(X)$ | $K_P^F(X)$ | $K(X)$ | $K_B^F(X)$ | $K_P^F(X)$ | $K(X)$ | $K_B^F(X)$ | $K_P^F(X)$ | $K(X)$ |
| $U \in \mathcal{WN}$ | | | | | | | | | |
| 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 |

*Estimation results for scalar BEKK models for close-to-close Relative efficiency of the realised kernel $K(X)$ and the flat-top realised kernel, $K^F(X)$. Results for five different types of noise are presented. In the MA(1) and AR(1) designs, the variance of $\epsilon$ was scaled such that $Var(U) = \omega^2$. The squared bias, variance, and MSE have been 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). The results are based on 50000 repetitions.*