Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading

NEIL SHEPHARD
Nuffield College, New Road, Oxford OX1 1NF, UK,
Department of Economics, University of Oxford
neil.shephard@economics.ox.ac.uk

DACHENG XIU
5807 S. Woodlawn Ave,
Chicago, IL 60637, USA
Booth School of Business, University of Chicago
dacheng.xiu@chicagobooth.edu

First full draft: February 2012
This version: 22nd October 2012

Abstract
Estimating the covariance between assets using high frequency data is challenging due to market microstructure effects and asynchronous trading. In this paper we develop a multivariate realised quasi-likelihood (QML) approach, carrying out inference as if the observations arise from an asynchronously observed vector scaled Brownian model observed with error. Under stochastic volatility the resulting realised QML estimator is positive semi-definite, uses all available data, is consistent and asymptotically mixed normal. The quasi-likelihood is computed using a Kalman filter and optimised using a relatively simple EM algorithm which scales well with the number of assets. We derive the theoretical properties of the estimator and prove that it achieves the efficient rate of convergence. We show how to make it obtain the non-parametric efficiency bound for this problem. The estimator is also analysed using Monte Carlo methods and applied to equity data with varying levels of liquidity.

Keywords: EM algorithm; Kalman filter; market microstructure noise; non-synchronous data; portfolio optimisation; quadratic variation; quasi-likelihood; semimartingale; volatility.
JEL codes: C01; C14; C58; D53; D81

1 Introduction
1.1 Core message
The strength and stability of the dependence between asset returns is crucial in many areas of financial economics. Here we propose an innovative, theoretically sound, efficient and convenient method for estimating this dependence using high frequency financial data. We explore the properties of the methods theoretically, in simulation experiments and empirically.

Our realised quasi maximum likelihood (QML) estimator of the covariance matrix of asset prices is positive semidefinite and deals with both market microstructure effects such as bid/ask
bounce and crucially non-synchronous recording of data (the so-called Epps (1979) effect). Positive semidefiniteness allows us to define a coherent estimator of correlations and betas, objects of importance in financial economics. We derive the theoretical properties of our estimator and prove that it achieves the efficient rate of convergence. We show how to make it achieve the non-parametric efficiency bound for this problem and demonstrate theoretically the effect of asynchronous trading. The estimator is also analysed using Monte Carlo methods and applied on equity data in a high dimensional case.

Our results show our methods deliver particularly strong gains over existing methods for unbalanced data: that is where some assets trade slowly while others are more frequently available.

1.2 Quasi-likelihood context

Our approach can be thought to be the natural integration of three influential econometric estimators, completing a line of research and opening up many more areas of development and application.

The first is the realised variance estimator, which is the QML estimator of the quadratic variation of a univariate semimartingale and was econometrically formalised by Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002). There the quasi-likelihood is generated by assuming the log-price is Brownian motion. Multivariate versions of these estimators were developed and applied in Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2004). These estimators are called realised covariances and have been widely applied.

The second is the Hayashi and Yoshida (2005) estimator, which is the QML estimator for the corresponding multivariate problem where there is irregularly spaced non-synchronous data, though sampling intervals are not incorporated into their estimator. Again the underlying log-price is modelled as a rotated vector Brownian motion.

Neither of the above estimators deals with noise. Xiu (2010) studied the univariate QML estimator where the Brownian motion is observed with Gaussian noise. He called this the “realised QML estimator” and showed this was an effective estimator for semimartingales cloaked in non-Gaussian noise. Moreover, the QML estimator is asymptotically equivalent to the optimal realized kernel, by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), with a suboptimal bandwidth. Note also the related Zhou (1996), Zhou (1998), Andersen, Bollerslev, Diebold, and Ebens (2001) and Hansen, Large, and Lunde (2008).

Our paper moves beyond this work to produce a distinctive and empirically important result. It proposes and analyses in detail the multivariate realised QML estimator which deals with irregularly spaced non-synchronous noisy multivariate data. We develop methods to allow it to be easily
implemented and develop the corresponding asymptotic theory under realistic assumptions. We show this estimator has a number of optimal properties.

1.3 Alternative approaches

A number of authors have approached this sophisticated multivariate problem using a variety of techniques. Here we discuss them to place our work in a better context.

As we said above the first generation of multivariate estimators, realised covariances, were based upon moderately high frequency data. Introduced by Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2004), these realised covariances use synchronised data sampled sufficiently sparsely that they could roughly ignore the effect of noise and non-synchronous trading. Related is Hayashi and Yoshida (2005) who tried to overcome non-synchronous trading but did not deal with any aspects of noise (see also Voev and Lunde (2007)).

More recently there has been an attempt to use the finest grain of data where noise and non-synchronous trading become important issues. There are five existing methods which have been proposed. Two deliver positive semi-definite estimators, so allowing correlations and betas to be coherently computed. They are the multivariate realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) and the non-biased corrected pre-averaging estimator of Christensen, Kinnebrock, and Podolskij (2010). Both use a synchronisation device called refresh time sampling. Neither converges at the optimal rate.

Two other estimators have been suggested which rely on polarisation of quadratic variation. Each has the disadvantage that they are not guaranteed to be positive semi-definite, so ruling out their direct use for correlations and betas. The papers are Aït-Sahalia, Fan, and Xiu (2010) and Zhang (2011). The bias-corrected Christensen, Kinnebrock, and Podolskij (2010) is also not necessarily positive semi-definite. Further, none of them achieve the non-parametric efficiency bound.

In a paper written concurrently with this one, Park and Linton (2012a) develop a Fourier based estimator of covariances, which extends the multivariate work of Mancino and Sanfelici (2009) and Sanfelici and Mancino (2008).

and Horel (2009). Surveys include, for example, McAleer and Medeiros (2008), Aït-Sahalia and Mykland (2009), Park and Linton (2012b) and Aït-Sahalia and Xiu (2012).

1.4 More details on our paper

Here we use a QML estimator in the multivariate case where we model efficient prices as correlated Brownian motion observed at irregularly spaced and asynchronously recorded datapoints. Each observation is cloaked in noise. We provide an asymptotic theory which shows how this approach deals with general continuous semimartingales observed with noise irregularly sampled in time.

The above approach can be implemented computationally efficiently using Kalman filtering. The optimisation of the likelihood is most easily carried out using an EM algorithm, which is implemented using a smoothing algorithm. The resulting estimator of the integrated covariance is positive semidefinite. In practice it can be computed rapidly, even in significant dimensions.

1.5 Some particularly noteworthy papers

There are a group of papers which are closest to our approach.

Aït-Sahalia, Fan, and Xiu (2010) apply the univariate estimator of Xiu (2010) to the multivariate case using polarisation. That is they estimate the covariance between $x_1$ and $x_2$, by applying univariate methods to estimate $\text{Var}(x_1 + x_2)$ and $\text{Var}(x_1 - x_2)$ and then looked at a scaled difference of these two estimates. The implied covariance matrix is not guaranteed to be positive semi-definite.

During our work on this paper we were sent a copy of Corsi, Peluso, and Audrino (2012) in January 2012 which was carried out independently and concurrently with our work. This paper is distinct in a number of ways, most notably we have a fully developed econometric theory for the method under general conditions and our computations are somewhat different. However, the overarching theme is the same: dealing with the multivariate case using a missing value approach (see the related Elerian, Chib, and Shephard (2001), Roberts and Stramer (2001) and Papaspiliopoulos and Roberts (2012) for discretely observed non-linear diffusions) based on Brownian motion observed with error. A variant of this paper, also dated January 2012, by Peluso, Corsi, and Mira (2012), who carry out a related exercise to Corsi, Peluso, and Audrino (2012) but this time using Bayesian techniques. They assume the efficient price is Markovian and have no limiting theory for their quasi-likelihood based approach. Related to these papers is the earlier more informal univariate analysis of Owens and Steigerwald (2006) and the multivariate analysis of Cartea and Karyampas (2011).

In late April 2012 we also learnt of Liu and Tang (2012). They study a realised QML estimator of a multivariate exactly synchronised dataset. They propose using Refresh Time type devices to
achieve exact synchronicity. Under exact synchronicity their theoretical development is significant and independently generates the results in one of the theorems in this paper. We will spell out the precise theoretical overlap with our paper later. To be explicit they do not deal with asynchronous trading, which is the major contribution of our paper.

### 1.6 Structure of the paper

The structure of our paper is as follows. In Section 2 we define the model which generates the quasi-likelihood and more generally establish our notation. We also define our multivariate estimator. In Section 3 we derive the asymptotic theory of our estimator under some rather general conditions. In Section 4 we extend the core results in various important directions. In Section 5 we report on some Monte Carlo experiments we have conducted to assess the finite sample performance of our approach. In Section 6 we provide results from empirical studies, where the performance of the estimator is evaluated with a variety of equity prices. In Section 7 we draw our conclusions. The paper finishes with a lengthy appendix which contains the proofs of various theorems given in the paper, and a detailed online appendix with more empirical analysis.

### 2 Models

#### 2.1 Notation

We consider a $d$-dimensional log-price process $x = (x_1, ..., x_d)'$. These prices are observed irregularly and non-synchronous over the interval $[0, T]$, where $T$ is fixed and often thought of as a single day. These observations could be trades or quote updates. Throughout we will refer to them as trades.

We write the union of all times of trades as

$$t_i, \quad 1, 2, ..., n,$$

where we have ordered the times so that $0 \leq t_1 < ... < t_i < ... < t_n \leq T$. Note that the $t_i$ times must be distinct. Price updates can occur exactly simultaneously, a feature dealt with next.

Associated with each $t_i$ is an asset selection matrix $Z_i$. Let the number of assets which trade at time $t_i$ be $d_i$ and so $1 \leq d_i \leq d$. Then $Z_i$ is $d_i \times d$, full of zeros and ones where each row sums exactly to one. Unit elements in column $k$ of $Z_i$ shows the $k$-th asset traded at time $t_i$.

#### 2.2 Efficient price

$x$ is assumed to be driven by $y$, the efficient log-price, abstracting from market microstructure effects. The efficient price is modelled as a Brownian semimartingale defined on some filtered
probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t), P)\),
\[
y(t) = \int_0^t \mu(u)du + \int_0^t \sigma(u)dW(u),
\]
where \(\mu\) 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, and Renault (1996). Then the ex-post covariation is
\[
[y, y]_T = \int_0^T \Sigma(u)du,
\]
where
\[
[y, y]_T = \text{plim}_{n \to \infty} \sum_{j=1}^n \{y(\tau_j) - y(\tau_{j-1})\} \{y(\tau_j) - y(\tau_{j-1})\}',
\]
(e.g. Protter (2004, p. 66–77) and Jacod and Shiryaev (2003, p. 51)) for any sequence of deterministic synchronized partitions \(0 = \tau_0 < \tau_1 < ... < \tau_n = T\) with \(\sup_j \{\tau_{j+1} - \tau_j\} \to 0\) for \(n \to \infty\). This is the quadratic variation of \(y\). Our interest is in estimating \([y, y]_T\) using \(x\).

Throughout we will assume that \(y\) and the random times of trades \(\{t_i, Z_i\}\) are stochastically independent. This is a strong assumption and commonly used in the literature (but note the discussion in, for example, Engle and Russell (1998) and Li, Mykland, Renault, Zhang, and Zheng (2009)). This assumption means we can make our inference conditional on \(\{t_i, Z_i\}\) and so regard these times of trades as fixed.

Throughout we assume that we see a blurred version of \(y\), with our data being
\[
x_i = Z_i y(t_i) + Z_i \varepsilon_i, \quad i = 1, 2, ..., n,
\]
where \(\varepsilon_i\) is a vector of potential market microstructure effects. We will assume \(E(\varepsilon_i) = 0\) and write \(\text{Cov}(\varepsilon_i) = \Lambda\), where \(\Lambda\) is diagonal\(^1\). General time series discussions of missing data includes Harvey (1989, Ch. 6.4), Durbin and Koopman (2001, Ch. 2.7) and Ljung (1989).

### 2.3 A Gaussian quasi-likelihood

We proxy the Brownian semimartingale by Brownian motion, which is non-synchronously observed. This will be used to generate a quasi-likelihood. We model
\[
y(t) = \sigma W(t).
\]

\(^1\)Corsi, Peluso, and Audrino (2012) use a slightly different approach. They update at time points \(T_i/n\) whether there is new data or not. They used a linear Gaussian state space model \(x_i = Z_i y(T_i/n) + \varepsilon_i\), where a selection matrix \(Z_i\) is always \(d \times d\), but some rows are entirely made up of zeros if a price is not available at that particular time. If a price is entirely missing, the input for their Kalman innovations \(\varepsilon_i\) is set to zero.
Then writing $\Sigma = \sigma \sigma'$, we have that

$$y(t_i) - y(t_{i-1}) \sim N(0, \Sigma (t_i - t_{i-1})),$$

while all the non-overlapping innovations are independent.

Throughout we will write

$$u_i = y(t_i) - y(t_{i-1}), \quad \Delta_i^n = t_i - t_{i-1}.$$

Of course $\Delta_i^n > 0$ is a scalar.

At this point we assume that

$$\varepsilon_i \overset{iid}{\sim} N(0, \Lambda).$$

where $\Lambda$ is diagonal. Then we can think of the time series of observations $x_{1:n} = (x_1, \ldots, x_n)'$ as a Gaussian state space model. A discussion of the corresponding literature is available in, for example, Harvey (1989), West and Harrison (1989) and Durbin and Koopman (2001).

### 2.4 ML estimation via EM algorithm

Our goal is to develop positive semidefinite estimators of $\Sigma$, noting for us that $\Lambda$ is a nuisance. We would like our methods to work in quite high dimensions and so the EM approach to maximising the log-likelihood function is attractive. EM algorithms are discussed in, for example, Tanner (1996) and Durbin and Koopman (2001, Ch. 7.3.4).

We note that the complete log-likelihood is, writing and recalling, $e_i = x_i - Z_i y(t_i)$, $u_i = y(t_i) - y(t_{i-1})$, of the form, writing $y_{1:n} = (y_1, \ldots, y_n)'$, and assuming $y_1 \sim N(\hat{y}_1, P_1)$ which is independent of $(\Sigma, \Lambda)$,

$$\log f(x_{1:n}|y_{1:n}; \Lambda) + \log f(y_{1:n}; \Sigma) = c - \frac{1}{2} \sum_{i=1}^{n} \log |Z_i \Lambda Z_i'| - \frac{1}{2} \sum_{i=1}^{n} \epsilon_i' (Z_i \Lambda Z_i')^{-1} \epsilon_i$$

$$- \frac{1}{2} \sum_{i=2}^{n} \log |\Sigma| - \frac{1}{2} \sum_{i=2}^{n} \Delta_i^n u_i' \Sigma^{-1} u_i.$$

Then the EM algorithm works with the

$$E \{ [\log f(x_{1:n}|y_{1:n}; \Lambda) + \log f(y_{1:n}; \Sigma) ] | x_{1:n}; \Lambda, \Sigma \}$$

$$= c - \frac{1}{2} \sum_{i=1}^{n} \log |Z_i \Lambda Z_i'| - \frac{1}{2} \sum_{i=1}^{n} E \{ \epsilon_i' (Z_i \Lambda Z_i')^{-1} \epsilon_i | x_{1:n}; \Lambda, \Sigma \}$$

$$- \frac{1}{2} \sum_{i=2}^{n} \log |\Sigma| - \frac{1}{2} \sum_{i=2}^{n} \frac{1}{\Delta_i^n} E \{ u_i' \Sigma^{-1} u_i | x_{1:n}; \Lambda, \Sigma \}.$$

Writing $\hat{e}_{ij:n} = E(e_i | x_{1:n})$ and $D_{ij:n} = Mse(e_i | x_{1:n})$, then

$$E \{ \epsilon_i' (Z_i \Lambda Z_i')^{-1} \epsilon_i | x_{1:n} \} = tr \{ (Z_i \Lambda Z_i')^{-1} E(\epsilon_i' e_i' | x_{1:n}) \} = tr \{ (Z_i \Lambda Z_i')^{-1} \hat{e}_{ij:n} \hat{e}_{ij:n}' + D_{ij:n} \},$$

7
and, writing \( \hat{u}_{i|n} = E(u_i|x_{1:n}) \) and \( N_{i|n} = \text{Mse}(u_i|x_{1:n}) \), then
\[
E\{u'_i \Sigma^{-1} u_i | x_{1:n}\} = \text{tr} \{ \Sigma^{-1} E\{u_i u'_i | x_{1:n}\} \} = \text{tr} \left[ \Sigma^{-1} \left\{ \hat{u}_{i|n} \hat{v}_{i|n} + N_{i|n} \right\} \right].
\]

Then the EM update is
\[
\hat{\Sigma} = \frac{1}{n-1} \sum_{i=2}^{n} \frac{1}{\Delta_i} \left\{ \hat{u}_{i|n} \hat{v}_{i|n} + N_{i|n} \right\}, \quad \text{diag}(\hat{\Lambda}) = \left( \sum_{i=1}^{n} Z'_i Z_i \right)^{-1} \text{diag} \left( \sum_{i=1}^{n} Z'_i \{ \hat{\epsilon}_{i|n} \hat{\epsilon}_{i|n}' + D_{i|n} \} Z_i \right).
\]
Iterating these updates, the sequence of \( (\hat{\Sigma}, \hat{\Lambda}) \) converges to a maximum in the likelihood function.

### 2.5 Recalling the disturbance smoother

Computing \( \hat{\epsilon}_{i|n}, \hat{u}_{i|n}, D_{i|n} \) and \( N_{i|n} \) is routine and rapid, if rather tedious to write down. It is carried out computationally efficiently using the “disturbance smoother.” The following subsection is entirely computational and can be skipped on first reading without loss.

The smoother starts by running with the Kalman filter (e.g. Durbin and Koopman (2001, p. 67)), which is run forward in time \( i = 1, 2, \ldots, n \) through the data. In our case it takes on the form \( v_i = x_i - Z_i \hat{y}_i \), \( F_i = Z_i (P_i + \Lambda) Z'_i, K_i = P_i Z'_i F_i^{-1} \), \( L_i = I - K_i Z_i \) then \( \hat{y}_{i+1} = \hat{y}_i + K_i v_i \), \( P_{i+1} = P_i L_i + \Delta_i^{n+1} \Sigma \). Here \( \hat{y}_i = E(y_i|x_{1:i-1}) \) and \( F_i = \text{Cov}(x_i|x_{1:i-1}) \). These recursions need some initial conditions \( \hat{y}_1 \) and \( P_1 \). Throughout we will assume their choice does not depend upon \( \Sigma \) or \( \Lambda \). A typical selection for \( \hat{y}_1 \) is the opening auction price, whereas an alternative is to use a diffuse prior. Here \( v_i \) is \( d_i \times 1 \), \( F_i \) is \( d_i \times d_i \), \( K_i \) is \( d \times d_i \), \( \hat{y}_{i+1|1} \) is \( d \times 1 \) and \( P_{i+1} \) and \( L_i \) are \( d \times d \). Note that for large \( d \), the update for \( P_{i+1} \) is the most expensive, but it is highly sparse as \( L_i \) is sparse. Each iteration of the EM algorithm will lead to a non-negative change in the quasi log-likelihood
\[
\log f(x_{1:n}; \Lambda, \Sigma) = c - \frac{1}{2} \sum_{i=1}^{n} \log |F_i| - \frac{1}{2} \sum_{i=1}^{n} v'_i F_i^{-1} v_i.
\]

The disturbance smoother (e.g. Durbin and Koopman (2001, p. 76)) is run backwards \( i = n, n-1, \ldots, 1 \) through the data. It takes the form, writing \( H_i = Z_i \Lambda Z'_i \), a \( d_i \times d_i \) matrix, \( \hat{\epsilon}_{i|n} = H_i (F_i^{-1} v_i - K_i' r_i), D_{i|n} = H_i - H_i (F_i^{-1} + K'_i M_i K_i) H_i, \hat{u}_{i|n} = \Delta_i^o \Sigma r_{i-1}, N_{i|n} = \Delta_i^o \Sigma - (\Delta_i^o)^2 \Sigma M_{i-1} \Sigma \), where we recursively compute \( r_{i-1} = Z'_i F_i^{-1} v_i + L'_i r_i, M_{i-1} = Z'_i F_i^{-1} Z_i + L'_i M_i L_i, \) starting out with \( r_n = 0, M_n = 0 \). Here \( \hat{\epsilon}_{i|n} \) is \( d_i \times 1 \) and \( D_{i|n} \) is \( d_i \times d_i \). While \( \hat{u}_{i|n} \) and \( r_i \) are \( d \times 1 \), and \( N_{i|n} \) and \( M_i \) are \( d \times d \). Notice again the updates for \( M_i \) are highly sparse.

### 3 Econometric theory

In this section, we develop the asymptotic theory for the bivariate case, as the general multivariate case can be derived similarly. To get to the heart of the issues our analysis follows four steps.

First we look at the benchmark bivariate ML estimator case where the volatility matrix is fixed and there are equidistant observations. Secondly we show how those results change when the noise is non-Gaussian and we have stochastic volatility effects, but still have equidistant observations.
Thirdly and more realistically we discuss the impact of having unequally spaced data which is wrongly synchronised in the quasi-likelihood. Finally we discuss the impact of non-synchronised data on a fully non-synchronised quasi-likelihood.

3.1 Step one: benchmark bivariate MLE

3.1.1 The model

We start with the constant covariance matrix case with equidistant observations, which means that \( Z_t = I_2 \) and \( t_i = Ti/n \). This means that we have synchronised trading. We also assume the market microstructure effects are independent and initially normal.

Then we observe returns

\[
r_{j,i} = x_{j,i} - x_{j,i-1}, \quad i = 1, 2, ..., n, \quad j = 1, 2,
\]

with, we assume,

\[
x_i = y(i/n) + \varepsilon_i, \quad i = 0, 1, 2, ..., n, \quad y(i/n) = y((i - 1)/n) + \sqrt{T/nu_i},
\]

where

\[
\left( \varepsilon_i \atop u_i \right) \overset{i.i.d.}{\sim} N \left( \begin{array}{c} \Lambda_1 \\ 0 \\ \Sigma \end{array} \right), \quad \Lambda = \left( \begin{array}{cc} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{array} \right), \quad \Sigma = \left( \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{array} \right).
\]

In discrete time \( x \) would be called a bivariate “local level model” (e.g. Durbin and Koopman (2001, Ch. 2)).

Suppose the observed returns are collected as \( r = (r_{1,1}, r_{1,2}, ..., r_{1,n}, r_{2,1}, ..., r_{2,n})' \), then the likelihood can be rewritten as

\[
L = -n \log(2\pi) - \frac{1}{2} \log(\det \Omega) - \frac{1}{2} r'\Omega^{-1}r,
\]

where \( \Omega = \Delta\Sigma \otimes I_n + \Lambda \otimes J_n \). Here \( \otimes \) denotes the Kronecker product and \( J_n \) is a \( n \times n \) matrix

\[
J_n = \begin{pmatrix}
2 & -1 & 0 & \cdots & 0 \\
-1 & 2 & -1 & \ddots & \vdots \\
0 & -1 & 2 & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & -1 \\
0 & \cdots & 0 & -1 & 2
\end{pmatrix}.
\]

The likelihood (4) is tractable as we know the eigenvalues and eigenvectors of \( J_n \) and so \( \Omega \).

3.1.2 The asymptotic theory

Before we give the bivariate case we recall the univariate one

\[
n^{1/2} \left( \hat{\Sigma}_{11} - \Sigma_{11} \right) \overset{L}{\to} N \left( 0, 8\Lambda_{11}^{1/2}\Sigma_{11}^{3/2}T^{-1/2} \right),
\]

9
see, for example, Stein (1987), Gloter and Jacod (2001a), Gloter and Jacod (2001b), Aït-Sahalia, Mykland, and Zhang (2005) and Xiu (2010). This result establishes the optimal rate – $n^{1/4}$ – that can be obtained with noisy data. We now go onto the Gaussian bivariate case.

**Theorem 1 (Bivariate MLE)** Assume the model (2)-(3) is true. Then the ML estimators $\hat{\Sigma}$ and $\hat{\Lambda}$ satisfy the central limit theorem as $n \to \infty$

$$n^{1/4} \begin{pmatrix} \hat{\Sigma}_{11} - \Sigma_{11} \\ \hat{\Sigma}_{12} - \Sigma_{12} \\ \hat{\Sigma}_{22} - \Sigma_{22} \end{pmatrix} \xrightarrow{d} N(0, \Pi).$$

Here for a $3 \times 1$ vector $\Sigma_\theta = \text{vech}(\Sigma) = (\Sigma_{1,1}, \Sigma_{1,2}, \Sigma_{2,2})'$ the $\Pi$ matrix is such that $\Pi^{-1} = \frac{\partial \Psi_{\theta}}{\partial \Sigma_{\theta}}$ with

$$\frac{\partial \Psi_{\Sigma_{u,v}}}{\partial \Sigma_{i,j}} = -(1 + 1_{u \neq v}) \frac{1}{2} \left( \int_0^\infty \frac{\partial \omega_{i,j}(\Sigma, \Lambda, x)}{\partial \Sigma_{u,v}} \, dx \right), \quad i, j, u, v = 1, 2,$$

where, writing $\Lambda_{ii}^* = \Lambda_{ii}/T$,

$$\omega^{1,1}(\Sigma, \Lambda, x) = \frac{\Sigma_{22} + \Lambda_{22}^* \pi^2 x^2}{(\Sigma_{11} + \Lambda_{11}^* \pi^2 x^2)(\Sigma_{22} + \Lambda_{22}^* \pi^2 x^2) - \Sigma_{12}^2},$$

$$\omega^{2,2}(\Sigma, \Lambda, x) = \frac{\Sigma_{11} + \Lambda_{11}^* \pi^2 x^2}{(\Sigma_{11} + \Lambda_{11}^* \pi^2 x^2)(\Sigma_{22} + \Lambda_{22}^* \pi^2 x^2) - \Sigma_{12}^2},$$

$$\omega^{1,2}(\Sigma, \Lambda, x) = \frac{-\Sigma_{12}}{(\Sigma_{11} + \Lambda_{11}^* \pi^2 x^2)(\Sigma_{22} + \Lambda_{22}^* \pi^2 x^2) - \Sigma_{12}^2}.$$

**Proof.** Given in the Appendix.

Each of the integrals in $\Pi^{-1}$ has an analytic solution (e.g. Mathematica will solve the integrals), but the result is not informative and so we prefer to leave it in this compact form.

When $\Sigma_{12} = 0$, there is no “externality,” i.e. the asymptotic variances for $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{22}$ in the bivariate case reproduce the one-dimensional MLE case. As the correlation increases from 0 to 1, the multivariate MLE $\hat{\Sigma}_{11}$ becomes more efficient than the univariate one, because more information is collected via the correlation with the other series. This is illustrated in Figure 1.

### 3.2 Step two: bivariate QMLE with equidistant observations

#### 3.2.1 Assumptions

We now move to the cases which are more realistic. We deal with them one at a time: stochastic volatility, irregularly spaced synchronised data and finally and crucially asynchronised data.

**Assumption 1** The underlying latent $d$-dimensional log-price process satisfies (1), where the drift $\mu$ is predictable locally bounded, the $d \times d$ volatility process $\sigma$ is locally bounded Itô semimartingales and $W$ is a $d$-dimensional Brownian motion.
Figure 1: The figure plots the relative efficiency for bivariate MLE of $\Sigma_{11}$ over the univariate alternative, against the correlation. $\Sigma_{11} = 0.25^2$. As the number falls below zero the gains from bivariate MLE become greater. The blue line, the red dashed line, and the black dotted line correspond to the cases with $\Sigma_{22} = 0.3^2, 0.25^2$, and $0.2^2$, respectively.

**Assumption 2** The noise $\varepsilon_i$ is a vector random variable which is independent and identically distributed, and independent of $t_i$, $W$, $\sigma$ and has fourth moments$^2$.

**Assumption 3** The trades are synchronised at times $t_i = Ti/n$ and $Z_i = I_n$.

### 3.2.2 The asymptotic theory

Before we give the bivariate case we define $R_T = \left( \frac{1}{T} \int_0^T \sigma_t^2dt \right) / \left( \frac{1}{T} \int_0^T \sigma_t^4dt \right)^2 \geq 1$, by Jensen’s inequality and recall the univariate result

$$n^{1/4} \left( \hat{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_t^2dt \right) \xrightarrow{Ls} MN \left( 0, (5R_T + 3) \Lambda_{11}^{1/2} \left( \frac{1}{T} \int_0^T \sigma_t^4dt \right)^{3/2} T^{-1/2} \right),$$

which is a rewrite of the result due to Xiu (2010). Here the asymptotic variance of the estimator increases with $R_T$ keeping $\frac{1}{T} \int_0^T \sigma_t^2dt$ fixed.

We now extend this to the multivariate case. The asymptotic theory for the realised QML estimator $\hat{\Sigma}$ is given below.

---

$^2$The i.i.d. assumption can be replaced by more general noise process, which is independent conditionally on $Y$. This allows some dependence and endogeniety, but the noise is still uncorrelated with $Y$. This assumption is the focus of, for example, Jacod, Podolskij, and Vetter (2010). We choose not to adopt it as the idea of the proof remains the same except for some technicalities.
Theorem 2 (Bivariate QMLE) Under Assumptions 1-3, we have

\[
\frac{1}{n} \left( \begin{array}{c}
\hat{\Sigma}_{11} - \frac{1}{T} \int_0^T \Sigma_{11,t} dt \\
\hat{\Sigma}_{12} - \frac{1}{T} \int_0^T \Sigma_{12,t} dt \\
\hat{\Sigma}_{22} - \frac{1}{T} \int_0^T \Sigma_{22,t} dt
\end{array} \right) \xrightarrow{L} \mathcal{N}(0, \Pi_Q),
\]

where \( \Pi_Q \) is

\[
\Pi_Q = \frac{1}{4} \left( \frac{\partial \Psi_{\theta}}{\partial \Sigma_{\theta}} \right)^{-1} \left\{ \text{Avar}^{(2)} + \text{Avar}^{(3)} + \text{Avar}^{(4)} \right\} \left( \frac{\partial \Psi_{\theta}}{\partial \Sigma_{\theta}} \right)^{-1},
\]

\[
\text{Avar}^{(2)} = 2 \sum_{l,s,u,v=1}^2 \int_0^\infty \frac{\partial \omega^{u,v}(\Sigma, \Lambda, x)}{\partial \Sigma_{\theta}} \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma'_{\theta}} dx \left( \frac{1}{T} \int_0^T \Sigma_{sv,t} \Sigma_{ul,t} dt \right),
\]

\[
\text{Avar}^{(3)} = 4 \sum_{l,s,u,v=1}^2 \Lambda_{ll}^s \int_0^\infty \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma_{\theta}} \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma'_{\theta}} \pi^2 x^2 dx \left( \frac{1}{T} \int_0^T \Sigma_{sv,t} dt \right),
\]

\[
\text{Avar}^{(4)} = 2 \sum_{l,s=1}^2 \Lambda_{ll}^s \Lambda_{ss}^l \int_0^\infty \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma_{\theta}} \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma'_{\theta}} \pi^4 x^4 dx.
\]

Here all the derivatives are evaluated at \( \Sigma = \frac{1}{T} \int_0^T \Sigma_t dt \).

**Proof.** Given in the Appendix.

In independent and concurrent work Liu and Tang (2012) have established a similar result, although their derivation follows the steps in Xiu (2010), which is different from what we suggest here.

3.3 Step 3: bivariate QMLE with irregularly space synchronised observations

We now build a quasi-likelihood upon irregularly spaced but synchronised times \( \{t_1, t_2, \ldots, t_n\} \), so that \( Z_i = I_2 \). These times imply a collection of increments \( \Delta^n_i = t_i - t_{i-1}, 1 \leq i \leq n \). We then make the following assumption.

**Assumption 4** We assume that \( \Delta^n_i = \Delta (1 + \xi_i), i = 1, 2, \ldots, n, \Delta = \frac{T}{n}, \mathbb{E}(\xi_i) = 0, \) where \( \text{Var}(\xi_i) < \infty \) and \( \{\xi_i, 1 \leq i \leq n\} \) are i.i.d.. Also assume \( Y \) and \( \{\xi_i\} \) are independent.

Assumption 4 means that \( \xi_i = O_p(1) \) and so the individual gaps shrink at rate \( O_p(n^{-1}) \).

**Corollary 1** Assume Assumptions 1, 2 and 4 hold and additionally \( \Sigma \) is constant and \( \mu = 0 \). Then the asymptotic variance of the MLE is the same as that in Theorem 1.

**Proof.** Given in the Appendix.

This makes it clear that in the synchronised case irregularly spacing of the data has no impact on the asymptotic distribution of the multivariate realised QML estimator. The reason for this is that the effect is less important than the presence of noise, a result which appears in the univariate
work of A"ıt-Sahalia, Mykland, and Zhang (2005) and A"ıt-Sahalia and Mykland (2003). This result contrasts with the realised variance where Mykland and Zhang (2006) have shown that the quadratic variation of the sampling times impacts the asymptotic distribution in the absence of noise.

This synchronised case is important in practice. Some researchers have analysed covariances by applying a synchronisation scheme to the non-synchronous high frequency observations. This delivers an irregularly spaced sequence of synchronised times of trades, although some prices could be somewhat stale. Typically there is a very large drop in the sample size due to synchronisation. The most well known such scheme is the Refresh Time method analysed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) and subsequently employed by, for example, Christensen, Kinnebrock, and Podolskij (2010) and A"ıt-Sahalia, Fan, and Xiu (2010). See also the earlier more informal papers by Harris, McMcInish, Shoesmith, and Wood (1995) and Martens (2003). Alternatively, Zhang (2011) discusses the Previous Tick approach, which always discards more data than what is suggested by the Refresh Time.

3.4 Step 4: bivariate QMLE with non-synchronous observations

In this paper, synchronisation is not needed for our method. Instead we write

\[ x_{j,i} = y_j(t_{j,i}) + \varepsilon_{j,i}, \quad j = 1, 2, \quad i = 1, 2, \ldots, n_j, \]

where \( t_{j,i} \) is the \( i \)-th observation on the \( j \)-th asset. The corresponding returns are \( r_{j,i} = x_{j,i} - x_{j,i-1} \).

**Assumption 5** We define \( \Delta^n_{j,i} = t_{j,i} - t_{j,i-1}, \quad i = 1, 2, \ldots, n_j \) and \( \Delta^n_j = \text{diag}(\Delta^n_{j,i}) \). Assume that \( \Delta^n_{j,i} = \bar{\Delta}_j, \quad j = 1, 2, \quad \bar{\Delta}_2 = m\bar{\Delta}_1, \) where \( \bar{\Delta}_j = T/n_j \).

These assumptions mean that the data is asynchronous, unless \( m = 1 \), but always equally spaced in time. The latter assumption is made as the previous subsection has shown that irregularly spacing of the data does not impact the asymptotic analysis of the realised QML and so it is simply cumbersome to include that case here without any increase in understanding. This notation means that \( \bar{\Delta}_j \) is the average time gap between observations while \( n_j \) is the sample size for the \( j \)-th asset.

Collecting the observed returns \( r = (r_{1,1}, r_{1,2}, \ldots, r_{1,n_1}, r_{2,1}, \ldots, r_{2,n_2})' \), then our likelihood can be rewritten as

\[ L = -(n_1 + n_2) \log(2\pi) - \frac{1}{2} \log(\det \Omega) - \frac{1}{2} r' \Omega^{-1} r, \quad (7) \]

where

\[ \Omega = \begin{pmatrix} \Sigma_{11} \Delta^{n_1} + \Lambda_{11} J_{n_1} & \Sigma_{12} \Delta^{n_1,n_2} \\ \Sigma_{12} \Delta^{n_2,n_1} & \Sigma_{22} \Delta^{n_2} + \Lambda_{22} J_{n_2} \end{pmatrix}, \quad (8) \]

13
with

\[ \Delta^{n_1,n_2}_{i,j} = \begin{cases} 
    t_{2,j} - t_{1,i-1}, & \text{if } t_{2,j-1} \leq t_{1,i-1} < t_{2,j} \leq t_{1,i}; \\
    t_{1,i} - t_{2,j-1}, & \text{if } t_{1,i-1} \leq t_{2,j-1} < t_{1,i} \leq t_{2,j}; \\
    t_{1,i} - t_{2,i-1}, & \text{if } t_{2,j-1} \leq t_{1,i-1} < t_{1,i} \leq t_{2,j}; \\
    t_{1,j} - t_{2,j-1}, & \text{if } t_{2,i-1} \leq t_{1,i-1} < t_{1,j} \leq t_{2,i}; \\
    0, & \text{otherwise}. 
\end{cases} \tag{9} \]

As before \( J_n \) is given in (5). Note that writing out this likelihood function does not require Assumption 5.

**Theorem 3** Assume Assumptions 1, 2 and 5 hold and additionally \( \Sigma \) is constant and \( \mu = 0 \). Assume also that \( \Delta^{n_j}_{j} = \hat{\Delta}_j \), for \( j = 1, 2 \), and \( \hat{\Delta}_2 = m \hat{\Delta}_1 \), where \( n_1 > n_2 \), i.e. \( m \to \infty \). Then in this asynchronous case, the central limit theorem is given by:

\[
\left( \begin{array}{c}
  n_1^{1/4} (\hat{\Sigma}_{11} - \Sigma_{11}) \\
  n_2^{1/4} (\hat{\Sigma}_{12} - \Sigma_{12}) \\
  n_2^{1/4} (\hat{\Sigma}_{22} - \Sigma_{22})
\end{array} \right) \xrightarrow{L} N(0, \Pi_A), \quad \text{where} \quad \Pi_A = \left( \begin{array}{ccc}
  \frac{\partial \psi_{11}}{\partial \Sigma_{11}} & 0 & 0 \\
  0 & \frac{\partial \psi_{12}}{\partial \Sigma_{12}} & \frac{\partial \psi_{12}}{\partial \Sigma_{22}} \\
  0 & \frac{\partial \psi_{22}}{\partial \Sigma_{22}} & \frac{\partial \psi_{22}}{\partial \Sigma_{22}}
\end{array} \right)^{-1},
\]

such that

\[
\frac{\partial \psi_{u,v}}{\partial \Sigma_{i,j}} = -(1 + 1_{u\neq v}) \frac{1}{2} \left( \int_0^\infty \frac{\partial \omega^{i,j}(\Sigma, \Lambda, x)}{\partial \Sigma_{u,v}} \, dx \right), \quad i, j, u, v = 1, 2,
\]

where, writing \( \Lambda^{*}_{ii} = \Lambda_{ii}/T \),

\[
\begin{align*}
\omega^{1,1}(\Sigma, \Lambda, x) &= \frac{1}{(\Sigma_{11} + \Lambda^{*}_{11} \pi^2 x^2)}, \\
\omega^{1,2}(\Sigma, \Lambda, x) &= \frac{\Sigma_{11}}{\Sigma_{11} (\Sigma_{22} + \Lambda^{*}_{22} \pi^2 x^2) - \Sigma_{12}^2}, \\
\omega^{1,2}(\Sigma, \Lambda, x) &= \frac{-\Sigma_{12}}{\Sigma_{11} (\Sigma_{22} + \Lambda^{*}_{22} \pi^2 x^2) - \Sigma_{12}^2}.
\end{align*}
\]

**Proof.** Given in the Appendix.

This Theorem suggests that including the extremely illiquid assets into estimation should not greatly affect variance and covariance estimates of the liquid ones. This property is distinct from any approach in the literature for which the synchronization method matters and the estimates become considerably worse when some asset is illiquid.

Interestingly, the asymptotic covariance of \( \hat{\Sigma}_{22} \) and \( \hat{\Sigma}_{12} \) can be obtained by plugging \( \Lambda_{11} = 0 \) in Theorem 1, as if the liquid asset were not affected by the microstructure noise. This is due to a type of “sparse sampling” induced by the substantial difference in the number of observations.

### 3.4.1 Extension: general \( m \) case

Note that Theorems 1 and 3 represent two points at either end of an important continuum. Theorem 1 in effect deals with the \( m = 1 \) case and Theorem 3 deals with \( m \to \infty \). Of course results for finite values of \( m > 1 \) would be of great practical importance, but we still have not sufficient
control of the terms to state that result entirely confidently. However, our theoretical studies and some Monte Carlo results we do not report here suggest that the result is simply the asymptotic distribution given in Theorem 1 but

\[
\begin{pmatrix}
\frac{1}{n} \left( \hat{\Sigma}_{11} - \Sigma_{11} \right) \\
\frac{1}{n} \left( \hat{\Sigma}_{12} - \Sigma_{12} \right) \\
\frac{1}{n} \left( \hat{\Sigma}_{22} - \Sigma_{22} \right)
\end{pmatrix} \xrightarrow{L} N(0, \Pi_{A,m}),
\]

where \( \Pi_{A,m} \) takes on the form given in Theorem 1 except \( \Lambda^*_{22} = \Lambda_{22} / (mT) \) holding. Hence the role of \( m \) is simply to rescale the measurement error variances.

### 3.4.2 Extension: irregularly and asynchronously spaced data

It is clear from Corollary 1 that the previous results hold under the type of irregularly spaced data which obeys Assumption 6 which simply extends Assumption 4.

**Assumption 6** We assume that \( \Delta^*_i = t_{j,i} - t_{j,i-1} = \Delta_j \left( 1 + \xi_{j,i} \right) \), \( i = 1, 2, \ldots, n_j \), \( \Delta^*_i = \text{diag}(\Delta^*_i) \) where \( E(\xi_{j,i}) = 0 \), \( \text{Var}(\xi_{j,i}) < \infty \) and \( \{\xi_{j,i}, 1 \leq i \leq n_j\} \). Also assume \( Y \) and \( \{\xi_{j,i}\} \) are independent.

### 3.4.3 Extension: allowing stochastic volatility

Likewise the extension to more interesting dynamics, stated in Assumption 1, combined with Assumption 6 is now clear. Again

\[
\begin{pmatrix}
\frac{1}{n} \left( \hat{\Sigma}_{11} - \Sigma_{11} \right) \\
\frac{1}{n} \left( \hat{\Sigma}_{12} - \Sigma_{12} \right) \\
\frac{1}{n} \left( \hat{\Sigma}_{22} - \Sigma_{22} \right)
\end{pmatrix} \xrightarrow{L} MN(0, \Pi_{Q,m}),
\]

where \( \Pi_{Q,m} \) simply replaces \( \Pi_Q \) in Theorem 2 by adjusting \( \Lambda^*_{22} = \Lambda_{22} / (mT) \). The proof of such a result just combines the proofs of the previous theorems. It is tedious.

### 4 Additional developments

#### 4.1 Realised QML correlation and regression estimator

The theorems above have an immediate corollary for the estimator of the daily “realised QML correlation estimator”

\[
\hat{\rho}_{12} = \frac{\hat{\Sigma}_{12}}{\sqrt{\hat{\Sigma}_{11} \hat{\Sigma}_{22}}} \in [-1, 1], \quad \text{of} \quad \rho_{12} = \frac{1}{\sqrt{\left( \frac{1}{T} \int_0^T \Sigma_{11,t} \, dt \right) \left( \frac{1}{T} \int_0^T \Sigma_{22,t} \, dt \right)}} \in [-1, 1].
\]

Also of importance is the corresponding regression or “realised QML beta”

\[
\hat{\beta}_{1|2} = \frac{\hat{\Sigma}_{12}}{\hat{\Sigma}_{22}}, \quad \text{which estimates} \quad \beta_{1|2} = \frac{1}{\sqrt{\left( \frac{1}{T} \int_0^T \Sigma_{11,t} \, dt \right) \left( \frac{1}{T} \int_0^T \Sigma_{22,t} \, dt \right)}}.
\]
The corresponding limit theory follows by the application of the delta method: \( \text{Avar}(\hat{\beta}_{12}) = \nu_\rho V_{Q\rho}^\prime, \) and \( \text{Avar}(\hat{\beta}_{12}) = \nu_\beta V_{Q\beta}^\prime, \) where
\[
\nu_\rho = \left( -\frac{1}{2} \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}} \frac{1}{\sqrt{\Sigma_{11}\Sigma_{22}}} - \frac{1}{2} \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}} \right)' \quad \text{and} \quad \nu_\beta = \left( 0, \frac{1}{\Sigma_{22}} - \frac{\Sigma_{12}}{\Sigma_{22}} \right)'.
\]

These are noise and asynchronous trading robust versions of the realised quantities studied by Andersen, Bollerslev, Diebold, and Labys (2003) and Barndoff-Nielsen and Shephard (2004).

### 4.2 Miniature realised QML based estimation

So far we have carried out realised QML estimation using the data all at once over the interval 0 to \( T. \) It is possible to follow a different track which is to break up the time interval \([0, T]\) into non-stochastic blocks \( 0 = b_0 < b_1 < \ldots < b_B = T. \) Then we can compute a realised QML estimator within each block. We then sum the resulting estimator up to produce our estimator of the required covariance matrix. Such a blocking strategy was used by, in a different context, Mykland and Zhang (2009) and Mykland, Shephard, and Sheppard (2012) for example.

We call the \( i \)-th block estimator the “miniature realised QML” estimator and write it as \( \hat{\Sigma}_{i1}. \) For fixed block sizes the resulting estimator, as the sample goes to infinity, is
\[
n_i^{1/4} \left( \hat{\Sigma}_{11,i} - \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right) \xrightarrow{L^s} MN \left( 0, \frac{(5R_{ii} + 3)\Lambda_{11}^{1/2}}{(b_i - b_{i-1})^{1/2}} \left( \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^{3/2} \right),
\]
where \( n_i = n(b_i - b_{i-1})/T, \)
\[
R_{ii} = \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^4 dt \left( \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^2 \geq 1.
\]

For fixed non-overlapping blocks the joint limit theory for the group of miniature realised QML estimators is normal with uncorrelated errors across blocks.

**Corollary 2** Define the unblocked QML estimator of \( \Sigma_{11}, \) \( \hat{\Sigma}_{11} = \frac{1}{B} \sum_{i=1}^{B} (b_i - b_{i-1}) \hat{\Sigma}_{11,i} \) then for fixed \( b_i \) and \( B \) we have as \( n \to \infty \)
\[
n^{1/4} \left( \hat{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_t^2 dt \right) \xrightarrow{L^s} MN \left( 0, \frac{1}{T^{3/2}} \sum_{i=1}^{B} (b_i - b_{i-1}) (5R_{ii} + 3)\Lambda_{11}^{1/2} \left( \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^{3/2} \right).
\]

**Proof.** Immediate extension of Xiu (2010), noting each block is conditionally independent.

At first sight this does not look like much of an advance. The virtue though is that \( \int_0^1 \sigma_t^2 dt \) and \( \int_0^1 \sigma_t^4 dt \) are of bounded variation in \( t \) and so are both \( O_p(t) \) as \( t \downarrow 0. \) This means that \( R_{ii} \approx 1 + O_p(b_i - b_{i-1}) \) and \( (b_i - b_{i-1})^{-1} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt - \sigma_{b_{i-1}}^2 = O_p(b_i - b_{i-1}). \) Now \( R_{ii} \) is crucial for it drives the inefficiency of this quasi-likelihood approach to inference. Driving it down to one allows
the estimator to be optimal in the limit and is achieved by allowing \( \max_i (b_i - b_{i-1}) = o(1) \) as a function of \( n \). In practice we take the gaps to very slowly shrink with \( n \). Of course this shrinkage requires \( B \) to increase with \( n \) very slowly.

The result is very simple and achieves the non-parametric efficiency bound

\[
n^{1/4} \left( \hat{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_i^2 \, dt \right) \xrightarrow{L^2} MN \left( 0, 8 \Lambda_{11}^{1/2} \left( \frac{1}{T} \int_0^T \sigma_i^3 \, dt \right) T^{-1/2} \right).
\]

This approach is also efficient when the variance of the noise is time-varying, for \( \Lambda_{11} \) is estimated separately within each block. Reiss (2011) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) discuss other estimators which achieve this bound.

### 4.3 Multistep realised QML estimator

There may be robustness advantages in estimating the integrated variances using univariate QML methods \( \hat{\Sigma}_{11}, \hat{\Sigma}_{22} \). These two estimates can then be combined with the QML correlation estimator \( \hat{\rho}_{12} \), obtained by simply maximising the quasi-likelihood with respect to \( \rho_{12} \) keeping \( \Sigma_{11}, \Sigma_{22} \) fixed at the first stage \( \hat{\Sigma}_{11}, \hat{\Sigma}_{22} \). We call such an estimator the “multistep covariance estimator”.

A potential advantage of this approach is that model specification for one asset price will not impact the estimator of the integrated variance for the other asset. Of course volatility estimation is crucial in terms of risk scaling.

### 4.4 Sparse and subsampled realised QML

A virtue of the realised QML is that it is applied to all of the high frequency data. However, this estimator may have challenges if the noise has more complicated dynamics. Although we have proved results assuming the noise is i.i.d., it is clear from the techniques in the literature that the results will hold more generally if the noise is a martingale difference sequence (e.g. this covers some forms of price discreteness and diurnal volatility clustering in the noise). However, dependence which introduces autocorrelation in the noise could be troublesome. We might sometimes expect this feature if there are differential rates of price discovery in the different markets, e.g. an index fund leading price movements in the thinly traded Washington Post.

To overcome this kind of dependence in asset returns we define a “sparse realised QML” estimator, which corresponds to the sparse sampling realised variance. The approach we have explored is as follows.

We first list all the times of trades for asset \( i \), which are written as \( t_{j,i} \), which has a sample size of \( n_i \). Now think about collecting a subset of these times, taking every \( k \)-th time of trade. We write these times as \( t_{j,i}^* \), and the corresponding sample size as \( n_i^* \). We perform the same thinning operation for each asset. Then the union of the corresponding times will be written as \( t_i^* \). This
subset of the data can be analysed using the realised QML approach. Our asymptotic theory can be applied immediately to these \( i,j \) and \( n_i \), and the corresponding prices.

In practice it makes sense to amend this approach so that for each \( n_i > n_{\text{min}} \) where \( n_{\text{min}} \) is something like 20 or 50. This enforces that there is little thinning on infrequently traded assets.

Once we have defined a sparse realised QML, it is obvious that we could also simply subsample this approach, which means constructing \( k \) sets of subsampled datasets and for each computing the corresponding quasi-likelihood. We then average the \( k \) quasi-likelihoods and maximise them using the corresponding EM algorithm. We call this the “subsampled realised QML” estimator. This is simple to code and has the virtue that is employs all of the data in the sample while being less sensitive to the i.i.d. assumption.

5 Monte Carlo experiments

5.1 Monte Carlo design

Throughout we follow the design of Aït-Sahalia, Fan, and Xiu (2010), which is a bivariate model. Each day financial markets are open will be taken as lasting \( T = 1/252 \) units of time, so \( T = 1 \) would represent a financial year. Here we recall the structure of their model

\[
\text{dy}_{it} = \alpha_{it} dt + \sigma_{it} dW_{it}, \quad d\sigma_{it}^2 = \kappa_i (\bar{\sigma}_i^2 - \sigma_{it}^2) dt + s_i \sigma_{it} dB_{it} + \sigma_{it} J_{it}^Y dN_{it}
\]

where \( E(dW_{it} dB_{it}) = \delta_{ij} \rho_{it} dt \) and \( E(dW_{it} dW_{2t} | \rho^*) = \rho^* dt \). Here \( \kappa_i > 0 \).

Ignoring the impact of jumps, the variance process \( \sigma_{it}^2 \) has a marginal distribution given by \( \Gamma(2 \kappa_i \bar{\sigma}_i^2 / s_i^2, 2 \bar{\sigma}_i^2 / 2 \kappa_i) \). Throughout when jumps happen the log-jumps log \( J_{it}^Y \) are iid \( N(\theta_i, \mu_i) \), while \( N_{it} \) is a Poisson process with intensity \( \lambda_i \). Likewise \( \varepsilon_{it} \) are iid \( N(0, \sigma_i^2) \).

We now depart slightly from their setup. For each day we draw independently \( \sigma_{it}^2 \sim \Gamma(2 \kappa_i \bar{\sigma}_i^2 / s_i^2, 2 \bar{\sigma}_i^2 / 2 \kappa_i) \) over \( i = 1, 2 \), which means each replication will be independent. For each separate day we simulate independently \( \rho^* \sim \rho_0 \text{Beta}(\rho_1^*, \rho_2^*) \), where \( \rho_0 = \sqrt{(1 - \rho_1^*)(1 - \rho_2^*)} \), guarantees the positive-definiteness of the covariance matrix of \((W_1, W_2, B_1, B_2)\). This means \( \text{E}(\rho^*) = \rho_0 \rho_1^* / (\rho_1^* + \rho_2^*) \) and \( \text{sd}(\rho^*) = \rho_0 \sqrt{\rho_1^* \rho_2^*/\{ (\rho_1^* + \rho_2^*) \sqrt{\rho_1^* + \rho_2^* + 1} \}} \). The values of \( a_i, \alpha_i, \rho_i, \kappa_i, \theta_i, \mu_i, \lambda_i, s_i, \bar{\sigma}_i^2, \rho_1^* \) and \( \rho_2^* \) are given in Table 1. To check our limit theory calculations, Figure 2 plots the histograms of the

<table>
<thead>
<tr>
<th>( i )</th>
<th>( a_i )</th>
<th>( \alpha_i )</th>
<th>( \rho_1^* )</th>
<th>( \kappa_i )</th>
<th>( \theta_i )</th>
<th>( \mu_i )</th>
<th>( \lambda_i )</th>
<th>( s_i )</th>
<th>( \bar{\sigma}_i^2 )</th>
<th>( \rho_1^* )</th>
<th>( \rho_0 )</th>
</tr>
</thead>
</table>
| 1 | 0.005 | 0.05 | -0.6 | 3 | -5 | 0.8 | 12 | 0.16 | 0.8 | \( \rho_1^* = 2 \) | \( \rho_0 = 0.529 \)
| 2 | 0.001 | 0.01 | -0.75 | 2 | -6 | 1.2 | 36 | 0.09 | 0.5 | \( \rho_2^* = 1 \) | \( \text{E}(\rho^*) = 0.176 \)

Table 1: Parameter values which index the Monte Carlo design. Simulates from a bivariate model.

standardized pivotal statistics (standardising using the infeasible true random asymptotic variance
in each case) with 1,000 Monte Carlo repetitions sampled regularly in time at frequency of every 10 seconds, that is \( n = 2,340 \). This corresponds to an 6.5 hour trading day, which is the case for the NYSE and NASDAQ (we note the LSE and Xetra are open for 8.5 hours a day). The histograms show the limiting result provides a reasonable guide to the finite sample behaviour in these cases.

![Figure 2: The figure plots the histograms of the standardized pivotal statistics, which verify the asymptotic theory developed in Theorem 2. The standardisation is carried out using the infeasible true random asymptotic variance for each replication.](image)

In our main Monte Carlo we take \( n \in \{117,1170,11700\} \) and all results are based on 1,000 stochastically independent replications. Having fixed the overall sample size \( n \) we randomly and uniform scatter these points over the time interval \( t \in [0,T] \), recalling \( T = 1/252 \). For asset 1 we will scatter exactly \( nF \) points and for asset 2 there will be exactly \( n(1-F) \) points. This kind of stratified scatter corresponds to a sample from a Poisson bridge process with intensity \( nF/T \) and \( n(1-F)/T \), respectively. We call \( F \) the mixture rate and take \( F \in \{0.1,0.5,0.9\} \).

We will report on the accuracy on the daily estimation of the random \( \Sigma_{11} = \frac{1}{T} \int_0^T \Sigma_{11,t} dt \), \( \Sigma_{22} = \frac{1}{T} \int_0^T \Sigma_{22,t} dt \), \( \Sigma_{12} = \frac{1}{T} \int_0^T \Sigma_{12,t} dt \), \( \rho_{1,2} = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}} \), \( \beta_{1|2} = \Sigma_{12}/\Sigma_{22} \), \( \beta_{2|1} = \Sigma_{12}/\Sigma_{11} \).

### 5.2 Our suite of estimators

We will compute six estimators of \( \Sigma_{11}, \Sigma_{22}, \Sigma_{12}, \beta_{1|2}, \beta_{2|1} \) and \( \rho_{1,2} \). The six are: (i) realised QML, (ii) multistep realised QML estimator, (iii) blocked realised QML\(^3\), (iv) realised QML but using the reduced data synchronised by Refresh Time, (v) the realised kernel of Barndorff-Nielsen,

\(^3\)The number of blocks was taken to be \( \sqrt{n}/3 \). Sometimes this yielded very unequally sized blocks in which case we decreased the number of blocks so that there were at least two observations for each asset.
Hansen, Lunde, and Shephard (2011) which uses Refresh Time, (vi) Aït-Sahalia, Fan, and Xiu (2010) which uses polarisation and Refresh time. We will use the notation $\theta_{QML}$, $\theta_{Step}$, $\theta_{Bloc}$, $\theta_{RT}$, $\theta_{Kern}$, $\theta_{Pol}$, respectively, where $\theta$ is some particular parameter. We write this generically as $\theta_{L}$, with $L \in \{QML, Step, Bloc, RT, Kern, Pol\}$, and the corresponding estimator as $\hat{\theta}_{L}$.

All the estimators but (vi) deliver positive semi-definite estimators. Only (i) and (ii) use all the data, the others are based on Refresh Time. (i)-(iv) and (vi) converge at the optimal rate. (iii) should be the most efficient, followed by (i), then (ii), then (iii), then (vi) and finally (v).

5.3 Results

Throughout we report in Table 2 simulation based estimates of the $0.9$ quantiles of $|n^{1/4}(\hat{\theta}_{L} - \theta_{L})|$ for various values of $n$, $L$ and $\varphi$. The table also shows $n^{1/4}$, which allows us to see the actual speed by which the quantiles for $|\hat{\theta}_{L} - \theta_{L}|$ contract.

The results indicate that all six estimators perform roughly similarly for $\Sigma_{11}$ and $\Sigma_{22}$ when $F = 0.5$, with a small degree of underperformance for RT, Kern and Pol. When the data was more unbalanced, with $F = 0.1$ or $0.9$, then RT, Kern and Pol were considerably worse while realised QML being the best by a small margin. Bloc was a little disappointing when estimating the volatility of the less liquid asset. QML almost always outperformed Step but not by a great deal. The quantiles for Kern seem to mildly increase with $n$, which is what we would expect due to their slower rate of convergence.

For the measure of dependence $\Sigma_{12}$ there are signs that the realised QML type estimators “QML”, “Step” and “Bloc” perform better than “RT”. All seem to perform more strongly than the existing “Kern” and “Pol” estimators. The differences are less important in the case where $F = 0.5$.

When we move onto $\rho_{1,2}$ the differences become more significant, although we recall that realised QML and Step are identical in this case. When $F = 0.9$ then Pol estimator struggles with the quantiles being around twice that of QML and Bloc. A doubling of the quantile is massive, for these estimators are converging at rate $n^{1/4}$ so halving a quantile needs the sample size to increase by $2^{4} = 16$ fold. The results for Kern sit between Pol and QML, while RT is disappointing. This latter result shows it is the effect of refresh time sampling which is hitting these estimators. QML is able to coordinate the data more effectively. Similar results hold for $F = 0.1$. Overall the new methods seem to deliver an order of magnitude improvement in the accuracy of the estimator. In the balanced sampling case of $F = 0.5$ the differences are more moderate but similar.

Before we progress to the regression case it is helpful to calibrate how accurately we have estimated $\rho_{1,2}$ in the realised QML case. When $F = 0.1$ the quantile is 2.61 with $n = 117$, so
<table>
<thead>
<tr>
<th>$n$</th>
<th>$n^{1/4}$</th>
<th>$f = 0.9$</th>
<th>$f = 0.5$</th>
<th>$f = 0.1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QML</td>
<td>Step</td>
<td>Bloc</td>
<td>RT</td>
</tr>
<tr>
<td>117</td>
<td>3.29</td>
<td>0.40</td>
<td>0.43</td>
<td>0.42</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>0.38</td>
<td>0.39</td>
<td>0.43</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>0.36</td>
<td>0.35</td>
<td>0.43</td>
</tr>
<tr>
<td>117</td>
<td>3.29</td>
<td>0.23</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>0.19</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>0.18</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>117</td>
<td>3.29</td>
<td>0.29</td>
<td>0.35</td>
<td>0.31</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>0.23</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>0.21</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>117</td>
<td>3.29</td>
<td>2.35</td>
<td>2.35</td>
<td>2.34</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>1.60</td>
<td>1.60</td>
<td>1.83</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>1.48</td>
<td>1.48</td>
<td>1.92</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>3.18</td>
<td>3.24</td>
<td>3.99</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>2.82</td>
<td>2.86</td>
<td>4.23</td>
</tr>
<tr>
<td>117</td>
<td>3.29</td>
<td>2.97</td>
<td>2.55</td>
<td>2.57</td>
</tr>
<tr>
<td>1,170</td>
<td>5.84</td>
<td>2.05</td>
<td>2.02</td>
<td>2.13</td>
</tr>
<tr>
<td>11,700</td>
<td>10.3</td>
<td>1.68</td>
<td>1.67</td>
<td>2.67</td>
</tr>
</tbody>
</table>

Table 2: Monte Carlo results for the volatility, covariance, correlation and beta estimation. Throughout we report the 0.9 quantiles of $|n^{1/4}(\hat{\theta}_L - \theta_L)|$ over the 1,000 independent replications. $f$ denotes the percentage of the data corresponding to trades in asset 1. “QML” is our multivariate QMLE. “Step” is our multistep QMLE. “Bloc” is our blocked multivariate QMLE. “RT” is our multivariate QML using the Refresh Time. “Kern” is the existing multivariate realised kernel. “Pol” is the existing polarisation and Refresh Time estimator. The numbers in bold indicate the minimum of quantiles in comparison.
the corresponding quantile for \( \hat{\theta}_{QML} - \theta_{QML} \) is 0.794. When \( n = 1,170 \) it is 0.388. When \( n = 11,700 \) it is 0.191. In the balanced case \( F = 0.5 \) the corresponding results are 0.556, 0.277 and 0.137. Hence balanced data helps, but not by very much as long as \( n \) is moderately large and the realised QML method is used. Balancing is much more important for RT, Kern and Pol. We think this makes the realised QML approach distinctly promising. A final point is worth noting. Even though \( n = 11,700 \) the quantiles in the balanced case of 0.137 are not close to zero. Hence although we can non-parametrically estimate the correlation between assets, the estimation in practice is not without important error. This is important econometrically when we come to using these objects for forecasting or decision making.

The regression cases deliver the same type of results to the correlation, with again the QML, Step, and Bloc performing around an order of magnitude better than RT, Kern and Pol.

6 Empirical implementation

6.1 Our database

We use data from the cleaned trade database developed by Lunde, Shephard, and Sheppard (2012). It is taken from the TAQ database accessed through the Wharton Research Data Services (WRDS) system. They followed the step-by-step cleaning procedure used in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). Their cleaning rules include using data from a single exchange, selected as the one which generates the most trades on each day. It should also be noted that no quote information is used in the data cleaning process. The exchanges open at 9.30 and close at 16.00 local time.

An important feature of this TAQ data is that times are recorded to a second, so we take the median of multiple trades which occur in the same second. This can be thought of as a form of miniature preaveraging. As prices are recorded in seconds the maximum sample size is \( 60 \times 60 \times 6.5 = 23,400 \).

The data range from 1st January 2006 until 31st December 2009. We have selected 13 stocks from the S&P 500 with the aim of having 2 infrequently traded and 11 highly traded assets. This will allow us to assess the estimator in different types of data environments.

The assets we study are the Spyder (SPY), an S&P 500 ETF, along with some of the most liquid stocks in the Dow Jones 30 index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). We supplement these 11 series with two relatively infrequently traded stocks: Washington Post (WPO) and Berkshire Hathaway Inc. New Com (BRK-B). These 13 series are “unbalanced” in terms of
individual daily sample sizes, while the restricted 11 series are reasonably “balanced”.

6.2 Summaries

6.2.1 Sample sizes

Figure 3 shows the sample sizes of each asset on each day through time. What we plot is the median sample size of the 13 series together with the following quantile ranges: 0 to 25%, 25% to 75%, 75% to maximum. These ranges are indicated by shading. This is backed up by a line indicating the median. In addition we show the Refresh Time sample sizes when we use the 11 assets (denoted RefT1) and the corresponding result for all the 13 assets (RefT2). This type of “trading intensity graph” is highly informative and has the property that it scales with the number of assets. It first appeared in Lunde, Shephard, and Sheppard (2012) (who used it to look at 100s of assets).

![Trading Intensity Graph](image)

Figure 3: The trading intensity graph for our 13 assets. The figure plots the min, max, 25%, 50%, and 75% quantiles of the number of observations for the cleaned dataset. The number of refresh sampling for the 11 asset (RefT1) and 13 asset (RefT2) databases are also plotted.

The trading intensity graph shows the median intensity for the 13 assets is around 5,000 a day, slightly increasing through time. The maximum daily sample sizes are around 15,000, a tad below the feasible maximum of 23,400. The Refresh time for the 11 assets delivers a sample size of around 1,000 a day. However, for the 13 asset case the Refresh time dives down to around 60 a day. This is, of course, driven by the presence of the slow trading WPO and BRK-B.

This trading intensity graph demonstrates that the Refresh time approach is limited, for in large unbalanced systems it will lead to a significant reduction in the amount of data available to us. This could damage the effectiveness of the realised kernel or preaveraging in properly estimating
6.3 Volatilities

6.3.1 Summary statistics

We start our analysis of the realised quantities by looking at the univariate volatilities. We will compare realised QML with realised kernels and realised volatilities. The comparison will be made using estimators which use the one dimensional datasets and those based on the 13 dimensional series. The question is whether the use of the high dimensional series will disrupt the behaviour of the volatility estimators, due to the use of Refresh time\(^4\).

In our web appendix we give a detailed analysis of each of the 13 sets of series and their associated volatility estimators. Here we will focus on a single representative series, Exxon Mobile Corporation common stock (XOM), and some cross sectional summaries. It is important not to overreact to the specific features of a single series, we will make remarks only on characteristics which work out in the cross section.

Figure 4 shows 6 graphs. The first graph has the daily open to close returns, scaled by \(\sqrt{252}\) to place the data on an annual scale. Here 2 roughly represents 200\% annualised volatility. The middle and right hand top graphs show the time series of the daily estimates of the QML volatility in the series, drawn on the log10 scale. The first uses the univariate data, the second comes out of the 13 dimensional covariance fit. There is not a great deal of difference. For those unfamiliar with these type of non-parametric graphs, there is absolutely no time series smoothing here, the open to close volatility each day is estimated only with data on that day. The results are stunning, the volatility changes by an order of magnitude during this period. As the web appendix shows, this is entirely common across the assets.

For the moment we can now turn to Table 3 which shows summaries of the volatility estimators. They work in the following ways. Means are the square root of the average daily variance measure. For Open to Close returns this is simple the daily standard deviation. For the realised quantities this is the square root of the time series mean of the daily square volatility. For the autocorrelations, we report here results for lag 1 and lag 100. The results are always the autocorrelations of the

---

\(^4\)To be explicit, when we compute the realised kernel we take the data and approximately synchronise it using Refresh Time. This synchronised dataset is then used in all the realised kernel calculations. In the case where the analysis is carried out using the univariate databases, Refresh Time has no impact. In the 13 dimensional case it dramatically reduces the sample size due to the inclusion of the slow trading markets.

When we sparsely sample, we first sparsely sample and then compute the Refresh Time coordination. This has less impact than one might expect at first sight, as sparsely sampling has little impact on the slow trading stocks and these largely determine the Refresh Times. Hence the realised kernel will not be very impacted by sparse sampling in the multivariate case.

There is an argument that with the realised kernel we should only report results for sparsity being one, as that is the way it was introduced. For completeness though we have recomputed it for all the different levels of sparsity.
Figure 4: The log daily returns scaled by $\sqrt{252}$ to produce annual numbers. Vol for XOM computed using QML on the univariate series and in the 13 dimensional case. Multiply numbers by 100 to get yearly % changes. All acf calculations are based on volatilities, not variances, except for the squared returns. Vol sig are volatility signature plots, which plot as a function of sparsity the square root of the temporal average of the realised variance, realised QML, etc. Middle picture is the results for QML, right hand side is the results for the realised kernel (RK).
volatilities, not the squared volatilities. In terms of Open to Close returns, this means we report the autocorrelations of the absolute returns.

The Table reports results for 1 and 13 dimensions. It also gives results for different levels of sparse sampling, ranging from 1 trade (which means all the available 1 second return data is used) to 300 trades (which means taking every 300 trade). As the level of sparsity increases we might expect any bias in these estimators to fall but for them to become more variable.

<table>
<thead>
<tr>
<th>Mkt</th>
<th>Dimen</th>
<th>Spar</th>
<th>Vol QML acf 1 acf 100</th>
<th>Mean acf 1 acf 100</th>
<th>RV acf 1 acf 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>XOM</td>
<td>OtC</td>
<td>1</td>
<td>0.30 0.89 0.13</td>
<td>0.29 0.84 0.10</td>
<td>0.29 0.88 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>2</td>
<td>0.29 0.84 0.10</td>
<td>0.29 0.83 0.09</td>
<td>0.29 0.87 0.11</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>3</td>
<td>0.28 0.83 0.10</td>
<td>0.29 0.82 0.09</td>
<td>0.29 0.86 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>5</td>
<td>0.28 0.83 0.10</td>
<td>0.29 0.81 0.09</td>
<td>0.29 0.85 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>10</td>
<td>0.29 0.82 0.11</td>
<td>0.29 0.80 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>15</td>
<td>0.28 0.82 0.10</td>
<td>0.29 0.80 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>20</td>
<td>0.28 0.81 0.10</td>
<td>0.28 0.80 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>30</td>
<td>0.28 0.79 0.09</td>
<td>0.28 0.81 0.09</td>
<td>0.29 0.82 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>60</td>
<td>0.28 0.80 0.10</td>
<td>0.28 0.82 0.08</td>
<td>0.29 0.79 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>120</td>
<td>0.28 0.81 0.10</td>
<td>0.27 0.81 0.09</td>
<td>0.29 0.79 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>180</td>
<td>0.28 0.78 0.09</td>
<td>0.27 0.80 0.09</td>
<td>0.29 0.79 0.08</td>
</tr>
<tr>
<td>XOM</td>
<td>1</td>
<td>300</td>
<td>0.28 0.76 0.09</td>
<td>0.27 0.78 0.08</td>
<td>0.28 0.81 0.08</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>1</td>
<td>0.27 0.86 0.12</td>
<td>0.26 0.78 0.11</td>
<td>0.29 0.88 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>2</td>
<td>0.28 0.84 0.10</td>
<td>0.25 0.78 0.12</td>
<td>0.29 0.87 0.11</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>3</td>
<td>0.28 0.84 0.10</td>
<td>0.25 0.78 0.12</td>
<td>0.29 0.86 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>5</td>
<td>0.28 0.83 0.10</td>
<td>0.25 0.76 0.11</td>
<td>0.29 0.85 0.10</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>10</td>
<td>0.29 0.82 0.11</td>
<td>0.24 0.74 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>15</td>
<td>0.29 0.82 0.11</td>
<td>0.24 0.72 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>20</td>
<td>0.29 0.81 0.10</td>
<td>0.24 0.72 0.09</td>
<td>0.29 0.83 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>30</td>
<td>0.29 0.80 0.10</td>
<td>0.24 0.71 0.09</td>
<td>0.29 0.82 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>60</td>
<td>0.29 0.82 0.10</td>
<td>0.24 0.74 0.09</td>
<td>0.29 0.79 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>120</td>
<td>0.29 0.78 0.10</td>
<td>0.24 0.72 0.08</td>
<td>0.29 0.79 0.09</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>180</td>
<td>0.30 0.78 0.09</td>
<td>0.24 0.71 0.08</td>
<td>0.29 0.79 0.08</td>
</tr>
<tr>
<td>XOM</td>
<td>13</td>
<td>300</td>
<td>0.29 0.80 0.10</td>
<td>0.24 0.72 0.09</td>
<td>0.28 0.81 0.08</td>
</tr>
</tbody>
</table>

Table 3: OtC summaries. Unconditional volatilities of annualised daily log returns. Multiply by 100 to produce percentages. The acf of the OtC is the acf of the absolute value of the daily returns.

We first focus on the one dimensional case. The Table indicates at sparsity of 1 the QML estimator is slightly above the OtC average level of volatility, while this average level is roughly constant as sparsity varies above 1. The autocorrelation for these realised quantities is very much higher than for the absolute returns, both at lag 1 and 100. Roughly similar results hold for the realised kernel. Here the realised volatility does quite well, with roughly the same average value and a high degree of autocorrelation. This is the case in roughly half the series, the other half show quite pronounced upward bias in the estimator for low levels of sparsity. This is the famous upward bias of realised volatility due to market microstructure noise which has prompted
Table 4: Forecasting exercises. GARCHX models. LogL denotes increase in the log-likelihood compared to the GARCH model. $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma x_{t-1}$, where $x_t$ is a realised quantity.

such a large amount of theoretical work (e.g. Zhou (1996), Zhang, Mykland, and Aït-Sahalia (2005), Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)).

When we move to the 13 dimensional case the only substantial impact is on the realised kernel when the autocorrelations considerably fall. This is consistent with a deterioration in their quality caused by the use of Refresh Time. The QML estimator is hardly effected and of course realised volatility is not effected at all.

If we return to Figure 4 now some of these points are reiterated. The bottom left shows the autocorrelation function of the different estimators and show the deterioration in the realised kernel with dimension. It also shows the acf of the absolute and squared returns, indicating how much more noisy these estimators are. The middle graph shows the volatility signature plot of the realised quantities as a function of the level of sparsity. Recall the volatility signature plot is the square root of the temporal average of the realised estimator of the variance. It should be roughly flat as a function of sparsity if the estimator is robust to market microstructure effects. This is
true here. The seemingly small upward bias in some of the estimator is not material and will be matched by some other series which go the other way — QML and realised kernel are basically unbiased in this cross section. The results do not change when we look at the right hand picture, which shows the same thing for the 13 dimensional case. Realised kernels do not really exhibit bias in this case, Refresh Time impacts their variability.

6.3.2 Volatility forecasting

Another way of seeing the relative importance of these realised measures is through a prediction exercise. Here we use GARCHX models, supplementing the usual GARCH models of returns with $X$ variables which are logged realised variance type quantities. In particular we fit $\sigma_t^2 = \text{Var}(y_t | \mathcal{F}_{t-1})$ where

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma x_{t-1}.$$ 

Here $y_t$ is the $t$-th open to close return. These kind of extended GARCH models are now common in the literature, examples include Engle and Gallo (2006), Brownlees and Gallo (2010), Shephard and Sheppard (2010), Noureldin, Shephard, and Sheppard (2012), Hansen, Huang, and Shek (2011) and Hansen, Lunde, and Voev (2010).

The model is fitted using a Gaussian quasi-likelihood with

$$-\frac{1}{2} \sum_{i=1}^{n} \left( \log \sigma_i^2 + \frac{y_i^2}{\sigma_i^2} \right),$$

taking $\sigma_1^2 = \frac{1}{11} \sum_{i=1}^{11} y_i^2$. Here we will report only the estimated $\alpha, \beta, \gamma$ and the change the log likelihood in comparison with the simpler GARCH model. Clearly those changes are going to be non-negative by construction. If the presence of the realised quantity moves the likelihood up a great deal we will think this is evidence for its statistical usefulness. The results for all 13 assets are in our Web Appendix.

We first just focus on the XOM case. Table 4 shows the results in the univariate and 13 dimensional cases. The results show across the board important improvements when using the realised quantities and $\alpha$ is basically forced to zero. This is the common feature of these models in the literature, once the realised quantities are there there is no need for the square return (Shephard and Sheppard (2010)). Further $\beta$ falls dramatically and meaningfully. It means that the average lookback of the forecast on past data has reduced considerably, making it also more robust to structural breaks.

For some series the realised volatility adds little when the sparsity is 1 (due to the impact of the market microstructure), but this is not the case for XOM. There is some evidence that the QML
Figure 5: logL improvement in GARCHX compared to GARCH, where X is the realised quantity. High values are thus good. X-axis is the level of sparsity. Top left is the 1 dimensional realised QML result. Top right is the 1 dimensional realised kernel result. Bottom left is the 13 dimensional realised QML. Bottom right is the 13 dimensional realised kernel.

estimator does a little better when sparsity is a tiny amount above 1. All the estimators trail off as the sparsity gets large.

When we move to the 13 dimensional case the QML results hardly change and obviously the RVol case does not change at all. The realised kernel results are reasonably consistently damaged in this case, although the damage is not exceptional.

We can now average the log-likelihood changes using the cross-section of thickly traded stocks. The results are given in Figures 5, 6 and 7.

Figure 5 shows the change in the likelihood by including the realised quantities for the different assets, drawn against the level of sparsity. The thick line is the cross-sectional median. QML1 is the QML estimator based on the univariate series, QML13 estimator uses all 13 assets. RK denotes the corresponding results for the realised kernel. There are three noticeable features. The QML1 and QML13 results are very close. QML1 results get better as we move sparsity to 2 or 3 and then tail off. RK just tails off. RK13 looks a little worse than RK1.

Figure 6 reports the likelihood for QML minus the likelihood for RK. On the left hand side we deal with the univariate case. On the right the 13 dimensional case is the focus. So negative numbers prefer RK. This shows in the univariate case at sparsity of 1 RK is better, but this...
preference is removed by the time we reach sparsity of 3. After that they are basically the same. When we look at the 13 dimensional case, except for sparsity of 1, QML is better. This is consistent across many different levels of sparsity.

Finally, Figure 7 shows the difference in likelihood of the 13 dimensional model minus the likelihood for the 1 dimensional model. Thus a positive number means some damage is done to the predictions by using the 13 dimensional data rather than the univariate data. On the left hand side we report QML, on the right is RK. There are two things to see. First QML shows little change on average and the scatter in changes has little width. This means QML is hardly effected by the increase in dimension of the problem. The results for RK are very different, there are is a substantial reduction in the average fit, shown by the dark line, at all levels of sparsity. But further, the scatter in changes is quite large. Hence RK is indeed sensitive to the dimension, as expected.

6.4 Dependence

6.4.1 Summary statistics

We now turn to looking at covariation amongst the assets. We will focus on QML, realised kernel and realised covariance estimators, which are computed each day. In all cases they will be based
Figure 7: logL from the GARCHX model using 1 dimensional realised measure minus the logL from the GARCHX model using the 13 dimensional realised measure. Positive results thus suggest a fall in the fit using as the dimensional of the realised quantity increases. X-axis is the level of sparsity. Left hand side has the realised QML results. Right hand side has the realised kernel results.
on the 13 unbalanced database, which means we compute each day a 13 by 13 dimensional positive semidefinite estimator of the covariance matrix.

To look inside the covariance matrix we will focus on pairs of assets. To be concrete our focus will be on Bank of America (BAC) and SPY, the other 77 pairs are discussed in our web appendix. Again we will only flag up issues which hold up in the cross section.

The top left of Figure 8 shows the time series evolution of the conditional volatilities for these series based upon the past QMLs. Again these are plotted on the log10 scale. Of course it shows the typically lower level in the SPY series. What is key is that the wedge between these two series dramatically opens from 2008 onwards which means the ratio of the standard deviation of BAC and SPY has increased a great deal. If the correlation between the two series is stable this would deliver a massive increase in the “beta” of BAC. This is what actually happened, as can be seen in the middle top graph, which we will discuss in a moment.

The top right hand graph shows the time series of the daily correlations computed using the QML method. This shows a moderate increase in correlation during the crisis from around 0.55
up to around 0.7, with some weakening of the correlation from 2009 onwards after TARP.

If we return to the top middle graph we can see how enormously the beta changes through time. It was relatively stable until close to the end of 2007, but then it rapidly exploded reaching around 5 during some periods of the crisis. Nearly all of this move is a volatility induced change, although the correlation shift also has some impact. This boosting of the beta is also seen in our sample, but to a lesser extend, for J P Morgan and GE. The other stocks have more stable betas. Of course like Bank of America, J P Morgan is another financial company, while GE at the time has a significant exposure finance business.

Bottom left shows the autocorrelations of the QML, realised kernel, realised covariance and cross products of the open to close returns. These time series are quite heavy tailed and so heavily influenced by a handful of datapoints. This means it is particularly important to be careful in thinking through what these pictures means. There are a number of interesting features here. First the raw open to close returns have a small amount of autocorrelation in them, just as square daily returns are only moderately autocorrelated. The realised kernel is quite a bit better, but the realised covariance and QML show stronger autocorrelation. This indicates they are less noisy estimators than the realised kernel and the open to close return based measure, although they have biases.

Table 5: BAC and SPY pair summaries for the covariances. OtC denotes the open to close (annualised) daily log returns and the figure which follows it is the sample covariance of the returns. The acf of the OtC is the acf of the cross product of the returns, here reported at lags 1 and 100. QML, RK and RV are, respectively, the realised QML, the realised realised kernel and the realised covariance. All estimate the daily covariance of the series and are computed using the 13 dimensional database. Spar denotes sparsity. Mean is the temporal average of the time series.

<table>
<thead>
<tr>
<th>Spar</th>
<th>OtC</th>
<th>Mean</th>
<th>acf 1</th>
<th>acf 100</th>
<th>QML</th>
<th>Mean</th>
<th>acf 1</th>
<th>acf 100</th>
<th>RK</th>
<th>Mean</th>
<th>acf 1</th>
<th>acf 100</th>
<th>RV</th>
<th>Mean</th>
<th>acf 1</th>
<th>acf 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.09</td>
<td>0.20</td>
<td>0.07</td>
<td></td>
<td>0.03</td>
<td>0.79</td>
<td>0.21</td>
<td></td>
<td>0.06</td>
<td>0.55</td>
<td>0.17</td>
<td></td>
<td>0.01</td>
<td>0.68</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.80</td>
<td>0.19</td>
<td></td>
<td>0.04</td>
<td>0.80</td>
<td>0.18</td>
<td></td>
<td>0.06</td>
<td>0.53</td>
<td>0.16</td>
<td></td>
<td>0.02</td>
<td>0.69</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
<td>0.80</td>
<td>0.18</td>
<td></td>
<td>0.06</td>
<td>0.45</td>
<td>0.14</td>
<td></td>
<td>0.06</td>
<td>0.51</td>
<td>0.16</td>
<td></td>
<td>0.04</td>
<td>0.72</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.79</td>
<td>0.18</td>
<td></td>
<td>0.06</td>
<td>0.51</td>
<td>0.16</td>
<td></td>
<td>0.05</td>
<td>0.74</td>
<td>0.15</td>
<td></td>
<td>0.05</td>
<td>0.74</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.06</td>
<td>0.76</td>
<td>0.18</td>
<td></td>
<td>0.06</td>
<td>0.51</td>
<td>0.18</td>
<td></td>
<td>0.06</td>
<td>0.76</td>
<td>0.17</td>
<td></td>
<td>0.06</td>
<td>0.76</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.06</td>
<td>0.76</td>
<td>0.19</td>
<td></td>
<td>0.06</td>
<td>0.54</td>
<td>0.18</td>
<td></td>
<td>0.05</td>
<td>0.75</td>
<td>0.16</td>
<td></td>
<td>0.06</td>
<td>0.74</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.07</td>
<td>0.74</td>
<td>0.20</td>
<td></td>
<td>0.05</td>
<td>0.54</td>
<td>0.19</td>
<td></td>
<td>0.06</td>
<td>0.76</td>
<td>0.17</td>
<td></td>
<td>0.06</td>
<td>0.74</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.07</td>
<td>0.70</td>
<td>0.20</td>
<td></td>
<td>0.05</td>
<td>0.51</td>
<td>0.19</td>
<td></td>
<td>0.06</td>
<td>0.76</td>
<td>0.17</td>
<td></td>
<td>0.06</td>
<td>0.74</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.07</td>
<td>0.61</td>
<td>0.16</td>
<td></td>
<td>0.05</td>
<td>0.55</td>
<td>0.20</td>
<td></td>
<td>0.07</td>
<td>0.67</td>
<td>0.18</td>
<td></td>
<td>0.07</td>
<td>0.67</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.07</td>
<td>0.52</td>
<td>0.16</td>
<td></td>
<td>0.05</td>
<td>0.56</td>
<td>0.21</td>
<td></td>
<td>0.07</td>
<td>0.61</td>
<td>0.16</td>
<td></td>
<td>0.07</td>
<td>0.61</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>0.08</td>
<td>0.66</td>
<td>0.22</td>
<td></td>
<td>0.05</td>
<td>0.54</td>
<td>0.20</td>
<td></td>
<td>0.07</td>
<td>0.55</td>
<td>0.14</td>
<td></td>
<td>0.07</td>
<td>0.55</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.08</td>
<td>0.56</td>
<td>0.21</td>
<td></td>
<td>0.05</td>
<td>0.44</td>
<td>0.17</td>
<td></td>
<td>0.07</td>
<td>0.63</td>
<td>0.18</td>
<td></td>
<td>0.07</td>
<td>0.63</td>
<td>0.18</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: BAC and SPY pair summaries for the covariances. OtC denotes the open to close (annualised) daily log returns and the figure which follows it is the sample covariance of the returns. The acf of the OtC is the acf of the cross product of the returns, here reported at lags 1 and 100. QML, RK and RV are, respectively, the realised QML, the realised realised kernel and the realised covariance. All estimate the daily covariance of the series and are computed using the 13 dimensional database. Spar denotes sparsity. Mean is the temporal average of the time series.
sample. Then the Table shows the average level of the daily covariance estimator, using the QML, realised kernel and realised covariance. This is printed out for different levels of sparsity. The focus here is thus on the different biases in the estimators. Estimating covariances is hard. The famous Epps effect can be seen through RV, which underestimates the covariance by an order of magnitude when using every trade. It takes sparsity of 10 to reproduce half of the correlation.

The same impression can be gleaned from looking at the lower part of Figure 8, which is a covariance signature plot — showing the temporal average of the daily covariance estimators as a function of the level of sparsity. We can see that for low levels of sparsity the realised kernel is the least biased and the realised covariance the most. For higher levels of sparsity the realised kernel gets a little worse and both the QML and realised covariance improves and overtakes the realised kernel. Throughout QML is better than the realised covariance by a considerable margin.

The Table also shows the autocorrelation of the individual time series, recording results at lags 1 and 100. These results are striking, with more dependence for the QML series than the realised kernel. This matches results we saw for the volatilities in the previous subsection.

6.4.2 Dependence forecasting

We now move on to forecasting. The focus will be on the conditional covariance matrix

$$\Sigma_t = \text{Cov}(y_t|\mathcal{F}_{t-1}^{y,x}),$$

where $y_t$ is the $d$-dimensional open to close daily return vector. We will write $\Sigma_t = D_t R_t D_t$ where $D_t$ is a diagonal matrix with conditional standard deviations on the diagonal where the conditional variances are $\sigma^2_{i,t} = \text{Var}(y_{i,t}|\mathcal{F}_{t-1}^{y,x})$ where

$$\sigma^2_{i,t} = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i \sigma^2_{i,t-1} + \gamma_i x_{i,t-1}, \quad \omega_i, \alpha_i, \beta_i, \gamma_i \geq 0.$$  

This is the same conditional volatility model as we used in the previous subsection. We use these volatilities to construct

$$e_{i,t} = \frac{y_{i,t}}{\sigma_{i,t}},$$

the $i$-th devolatilised series.

Here $R_t$ is a conditional correlation matrix with $i,j$-th element

$$\text{Cor}(y_{i,t}, y_{j,t}|\mathcal{F}_{t-1}^{y,x}) = \text{Cor}(e_{i,t}, e_{j,t}|\mathcal{F}_{t-1}^{y,x}).$$

It is the focus of this subsection.

\footnote{Recall the realised kernel uses Refresh Time which is a kind of sparse sampling. Hence the fact that it has less bias at small levels of sparse sampling is not a surprise.}
In this exercise we will assume the following dynamic evolution

$$R_t = \omega \Pi + \alpha C_{t-1} + \beta R_{t-1} + \gamma X_{t-1}, \quad \omega, \alpha, \beta, \gamma \geq 0,$$

where $\omega + \alpha + \beta + \gamma = 1$, $X$ is a realised type correlation matrix and $\Pi$ is a matrix of parameters which form a correlation matrix. Here the $ij$-th element of $C_t$ is a moving block correlation of the devolatilised series

$$C_{i,j,t} = \sum_{s=1}^{M} e_{i,t-s} e_{j,t-s} \sqrt{\left( \sum_{s=1}^{M} e_{i,t-s}^2 \right) \left( \sum_{s=1}^{M} e_{j,t-s}^2 \right)} \in [-1,1].$$

In the case where $M = d$ and there is no $X$ variables, then (12) is the Tse and Tsui (2002) model. However, throughout we take $M = 66$, representing around 3 months of past data. In our experiments $X$ will represent the realised QML, realised kernel and realised covariance matrices. We will take $R_t = C_M$ for all $t \leq M$.

The key feature of (12) is that it is made up of the weighted sum of four correlation matrices, where the weights sum to one. Hence $R_t$ is always a correlation matrix. If $X$ is biased, for example, then $\Pi$ can partially compensate by not being the unconditional correlations of the innovations (11). We will see this happen in practice.

In order to tune the model and to assess the fit, we will work with the joint log-likelihood function

$$\log L = -\frac{1}{2} \log |\Sigma_t| - \frac{1}{2} y_t^\prime \Sigma_t^{-1} y_t = \log L_M + \log L_C$$

where

$$\log L_M = -\frac{1}{2} \log |D_t|^2 - \frac{1}{2} \left( y_t^\prime D_t^{-1} \right) \left( y_t D_t^{-1} \right) = \sum_{j=1}^{d} \log L_{M_i}$$

$$\log L_{M_i} = -\frac{1}{2} \sum_{t=M+1}^{n} \left( \log \sigma_{it}^2 + \frac{y_{it}^2}{\sigma_{it}^2} \right),$$

$$\log L_C = -\frac{1}{2} \sum_{t=M+1}^{n} \left( \log |R_t| + \frac{1}{2} \epsilon_t^\prime R_t^{-1} \epsilon_t - \epsilon_t^\prime \epsilon_t \right).$$

An alternative would be to use the DCC model which would have the form of

$$Q_t = \omega \Pi + \alpha e_{t-1} e_{t-1}^\prime + \beta Q_{t-1} + \gamma X_{t-1},$$

where

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}.$$

Unfortunately the impact of the non-linear transform for $R_t$ could be rather gruesome on the realised correlation matrix $X_{t-1}$ as the rescaling by the diagonal elements of $Q_t$ destroys all of its attractive properties. DCC models are discussed in Engle (2009).
We define here $e_t = D_{t-1} y_t$, the vector of “devolatilised returns”. The log $L_C$ term is setup to be a copula type likelihood.

We estimate the model using a two-step procedure, which can be formalised using the method of moments (e.g. Newey and McFadden (1994)). First we estimate the univariate models, and fix the volatility dynamic parameters at those estimated values. We then estimate the dependence model by optimising log $L_C$. A review of the literature on multivariate models is given by Silvennoinen and Teräsvirta (2009) and Engle (2009).

$$
\begin{array}{ccccccccc}
\text{Spar} & \rho & \alpha & \beta & \gamma & \log L & \rho & \alpha & \beta & \gamma & \log L & \rho & \alpha & \beta & \gamma & \log L \\
1 & 0.69 & 0.31 & 0.13 & 0.25 & 5.5 & 0.94 & 0.06 & 0.73 & 0.14 & 6.8 & 0.99 & 0.43 & 0.00 & 0.17 & 2.0 \\
2 & 1.00 & 0.12 & 0.41 & 0.25 & 8.3 & 0.99 & 0.00 & 0.69 & 0.17 & 8.3 & 0.99 & 0.31 & 0.00 & 0.23 & 3.7 \\
3 & 1.00 & 0.08 & 0.53 & 0.23 & 9.0 & 0.80 & 0.09 & 0.71 & 0.11 & 3.2 & 0.99 & 0.13 & 0.34 & 0.18 & 4.8 \\
5 & 0.99 & 0.05 & 0.61 & 0.21 & 9.6 & 0.78 & 0.09 & 0.70 & 0.10 & 2.8 & 0.99 & 0.07 & 0.44 & 0.18 & 6.0 \\
10 & 0.99 & 0.06 & 0.64 & 0.20 & 7.5 & 0.81 & 0.05 & 0.77 & 0.10 & 4.4 & 0.99 & 0.07 & 0.50 & 0.19 & 5.9 \\
15 & 0.99 & 0.07 & 0.72 & 0.15 & 5.1 & 0.80 & 0.07 & 0.71 & 0.12 & 4.8 & 0.99 & 0.07 & 0.55 & 0.18 & 5.9 \\
20 & 0.99 & 0.07 & 0.74 & 0.14 & 4.1 & 0.80 & 0.07 & 0.74 & 0.10 & 2.9 & 0.99 & 0.08 & 0.55 & 0.19 & 5.5 \\
30 & 0.99 & 0.06 & 0.78 & 0.11 & 4.0 & 0.84 & 0.05 & 0.74 & 0.12 & 5.1 & 0.99 & 0.10 & 0.57 & 0.18 & 4.9 \\
60 & 0.96 & 0.07 & 0.79 & 0.11 & 3.4 & 0.87 & 0.04 & 0.76 & 0.12 & 5.9 & 0.99 & 0.10 & 0.64 & 0.16 & 4.7 \\
120 & 0.95 & 0.08 & 0.76 & 0.11 & 3.7 & 0.82 & 0.08 & 0.71 & 0.09 & 3.2 & 0.99 & 0.14 & 0.56 & 0.20 & 4.9 \\
180 & 0.89 & 0.09 & 0.71 & 0.14 & 4.3 & 0.82 & 0.08 & 0.72 & 0.08 & 2.5 & 0.99 & 0.13 & 0.57 & 0.21 & 4.7 \\
300 & 0.85 & 0.09 & 0.70 & 0.12 & 2.9 & 0.82 & 0.08 & 0.72 & 0.08 & 1.8 & 0.99 & 0.10 & 0.69 & 0.15 & 4.6 \\
\end{array}
$$

Table 6: Forecasting exercises for cross-sectional dependence between BAC (Bank of America) and SPY (S&P 500 exchange traded fund). The estimated model is (12). Here $\rho$ is the non-unit element of $\Pi$. Note that $\omega = 1 - \alpha - \beta - \gamma$ and so is not reported here. LogL is the improvement in the log likelihood compared to the base model with no realised quantities, that is $\gamma = 0$.

The results from this forecasting exercise are given in Table 6. These are based upon the innovations from the univariate volatility models for BAC and SPY conditioning on lagged realised QML statistics. The results for the dependence model when we do not condition on any additional realised quantities, that is $\gamma = 0$, are given above the line in the Table. As $\beta = 0$ it means $\omega = 1 - \alpha$ and so is roughly 0.41. Here $\rho$ is the non-unit element of $\Pi$. Hence the estimated model for the conditional correlation is $0.41 \times 0.69 + 0.59 C_{1,2,t-1}$ where $C_{1,2,t-1}$ is the block correlation amongst the BAC and SPY innovations. This can be thought of as simply a shrunk block correlation.

When we condition on lagged realised quantities the log likelihood will typically rise. The improvement is recorded as logL in the Table. The results are reported for the QML, realised kernel and realised covariance. Obviously the results vary with the level of sparsity. For low levels of sparsity, RK does best. It drives $\alpha$ down to near zero, reminding us of the results we saw in the univariate cases discussed in the previous subsection. However, the improvement in the log likelihood is relatively modest, certainly less than we were used to from the univariate cases.
Figure 9: Improvements in logL for pairs involving SPY by including the realised quantities. On the x-axis is the level of sparsity. The heavy line denotes the median at each level of sparsity. All realised quantities are computed using all 13 series. Bottom right is the log likelihood for the model including realised QML minus the corresponding figure for realised kernel. This is plotted against sparsity. Also drawn is the corresponding result for the realised QML against realised covariance.

For low levels of sparsity QML is downward biased and so $\rho$ is estimated to be high to compensate. In the QML case we need larger sparsity to successfully drive down $\alpha$, but that estimator is certainly low with sparsity being 5 or more. This kind of levels of sparsity delivers a better fitting model than the results for RK, but the difference is not particularly large.

6.4.3 Cross section

Here we just focus on the cross section involving SPY based pairs. Of course there are 12 of these. The top left of Figure 9 shows the log-likelihood improvement in $\log L_C$ by including the realised QML information, i.e. allowing $\gamma$ to be greater than zero. The improvement is shown for each level of sparsity and is plotted separately for each of the 12 pairs. The median improvement is shown by the dark line. Almost throughout the improvement is modest, for a sole series the improvement is quite large. The realised QML performs better as the level of sparsity increases, but once again it tails off at the very end with very large sparsity for in those cases the sample sizes tend to be moderate and so the realised estimator is noisy.

Bottom left shows the corresponding results for the realised covariance. The results here are
poor for low levels of sparsity, adding basically nothing to the forecasting model. This is the influence of the Epps effect again. For higher levels of sparsity the realised covariance performs much better and approaches the realised QML estimator in terms of added value.

The top right shows the results for the realised kernel. This does best for very low levels of sparsity, making an important improvement in forecasting performance. However, as the level of sparsity increases the improvement due to the realised kernel falls away.

The bottom right graph shows the median log-likelihood improvement of QML minus the median log-likelihood improvement for the realised kernel. Positive numbers give a preference for QML. For low levels of sparsity, as we would expect, the realised kernel outperforms. However, for moderate degrees of sparsity the QML estimator has better performance.

Overall we can see that for the dependence modelling the inclusion of the realised information does add value, but the effects are not enormous. Realised QML again needs a moderate degree of sparsity to be competitive. For this level of sparsity realised QML slightly outperforms the realised kernel. Unlike the univariate case, the open to close information is not tested out of the model. But its importance does reduce a great deal by including the realised quantities.

7 Conclusion

This paper proposes and systematically studies a new method for estimating the dependence amongst financial asset price processes. The realised QML estimator is robust to certain types of market microstructure noise and can deal with non-synchronised time stamps. It is also guaranteed to be positive semi-definite and converges at the optimal asymptotic rate. This combination of properties is unique in the literature and so it is worthwhile exploring this estimator in some detail.

In this paper we develop the details of the quasi-likelihood and show how to numerically optimise it in a simple way even in large dimensions. We also develop some of the theory needed to understand the properties of the estimator and the corresponding results for realised QML estimators or betas and correlations. Particularly important is our theory for asynchronous data. Our Monte Carlo experiments are extensive, comparing the estimator to various alternatives. The realised QML performs well in these comparisons, in particular in unbalanced cases.

Our initial empirical results are somewhat encouraging, although much work remains. The volatilities seem to be robust to the presence of slowly trading stocks in the dataset. The improvement in the fit of the model by including these realised quantities is large. The results for measures for dependence are mixed, with the improvements from the realised quantities being modest. The realised QML is underestimating long-run dependence in these empirical experiments unless the level of sparsity is quite high. This underestimation also does not appear in our Monte
Carlo experiments. There are various explanations for this, but it would seem clear to need a more sophisticated model of market microstructure effects.

At the moment our recommendation for empirical work is for researchers to use realised QML or realised kernel estimators inside their conditional volatility models. When modelling dependence amongst the devolatilised returns the decision to use extra realised information is more balanced — it does increase the performance of the model but not by a great deal.

8 Acknowledgements

We thank Siem Jan Koopman for some early comments on some aspects of filtering with massive missing data, and Markus Bibinger for discussions on quadratic variation of time. We are particularly grateful to Fulvio Corsi, Stefano Peluso and Francesco Audrino for sharing with us a copy of their related work on this topic. The same applies to Cheng Liu and Cheng Yong Tang. We also thank Kevin Sheppard for allowing us to use the cleaned high frequency data he developed for Lunde, Shephard, and Sheppard (2012), as well as advice on all things multivariate. Last but not least, we thank seminar participants at CEMFI, especially Stéphane Bonhomme, Enrique Sentana, and David Veredas for helpful comments. This research was supported in part by the FMC Faculty Scholar Fund at the University of Chicago Booth School of Business.

References


**Appendices**
A Mathematical proofs

A.1 Proof of Theorem 1

There exists an orthogonal matrix \( U = (u_{ij}) \) given below, such that

\[
\begin{pmatrix}
U \\
U
\end{pmatrix}
\begin{pmatrix}
\Omega_{11} & \Omega_{12} \\
\Omega_{12} & \Omega_{22}
\end{pmatrix}
\begin{pmatrix}
U' \\
U'
\end{pmatrix}
= \begin{pmatrix}
\text{diag}(\mu_{1j}) & \Omega_{12} \\
\Omega_{12} & \text{diag}(\mu_{2j})
\end{pmatrix} =: V
\]

where

\[
\Omega_{12} = \Sigma_{12} \Delta \otimes I,
\]

\[
\Omega_{ii} = \Sigma_{ii} \Delta \otimes I + \Lambda_{ii} \otimes J, \quad i = 1, 2,
\]

\[
u_{ij} = \sqrt{\frac{2}{n+1}} \sin \left( \frac{i \cdot j}{n+1} \pi \right), \quad i, j = 1, \ldots, n,
\]

\[
\mu_{ij} = \Sigma_{ii} \Delta + 2\Lambda_{ii} \left( 1 - \cos \left( \frac{j}{n+1} \pi \right) \right), \quad i = 1, 2, \text{ and } j = 1, \ldots, n.
\]

Since \( U = U'^{-1} \), we have

\[
\Omega^{-1} = \begin{pmatrix}
U' \\
U'
\end{pmatrix} V^{-1} \begin{pmatrix}
U \\
U
\end{pmatrix}
\]

where

\[
V^{-1} = \begin{pmatrix}
\frac{\Sigma_{11} \Delta}{\mu_{11} \mu_{21} - \Sigma_{12} \Delta^2} & -\frac{\Sigma_{12} \Delta}{\mu_{11} \mu_{21} - \Sigma_{12} \Delta^2} \\
-\frac{\Sigma_{12} \Delta}{\mu_{11} \mu_{21} - \Sigma_{12} \Delta^2} & \frac{\mu_{11} \mu_{21} - \Sigma_{12} \Delta^2}{\mu_{11} \mu_{21} - \Sigma_{12} \Delta^2}
\end{pmatrix}
\]

One important observation is that \( U \) does not depend on parameters, hence taking derivatives of \( \Omega \) becomes very convenient with the help of the decomposition.

Note that

\[
\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \theta} = -\frac{1}{2 \sqrt{n}} \left( tr \left( \Omega^{-1} \frac{\partial \Omega}{\partial \theta} \right) - tr \left( \Omega^{-1} \frac{\partial \Omega}{\partial \theta} \Omega^{-1} r r' \right) \right)
\]

where for \( \theta = \Sigma_{11} \) and \( \Sigma_{12}, \)

\[
tr \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{11}} \right) = tr \left( V^{-1} \frac{\partial V}{\partial \Sigma_{11}} \right) = \sum_{i=1}^{n} \frac{\mu_{2i} \Delta}{\mu_{1i} \mu_{2i} - \Sigma_{12}^2 \Delta^2}
\]

\[
tr \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{12}} \right) = tr \left( V^{-1} \frac{\partial V}{\partial \Sigma_{12}} \right) = \sum_{i=1}^{n} \frac{-2 \Sigma_{12} \Delta^2}{\mu_{1i} \mu_{2i} - \Sigma_{12}^2 \Delta^2}
\]

and

\[
tr \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{11}} \Omega^{-1} r r' \right)
\]
Therefore, by direct calculations and using symmetry, we have

\[
E\left(-\frac{1}{\sqrt{n}} \frac{\partial^2 L}{\partial \Sigma_{11}^2}\right) = -\frac{1}{2 \sqrt{n}} \sum_{i=1}^{n} \left( \frac{\mu_i^2 \Delta^2}{(\mu_i \mu_{2i} - \Sigma_{12}^2 \Delta^2)^2} \right) \cdot \int_{0}^{\infty} \frac{(\Sigma_{12}^2 T + \Lambda_{12} \pi^2 x^2)^2 T^2}{((\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{12} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2)^2} dx := I_{11}^\Sigma,
\]

\[
E\left(-\frac{1}{\sqrt{n}} \frac{\partial^2 L}{\partial \Sigma_{12} \partial \Sigma_{22}} \right) = -\frac{1}{2 \sqrt{n}} \sum_{i=1}^{n} \left( \frac{\Sigma_{12}^2 \Delta^2}{(\mu_i \mu_{2i} - \Sigma_{12}^2 \Delta^2)^2} \right) \cdot \int_{0}^{\infty} \frac{\Sigma_{12}^2 T^4}{((\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{12} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2)^2} dx := I_{12}^\Sigma,
\]

\[
E\left(-\frac{1}{\sqrt{n}} \frac{\partial^2 L}{\partial \Sigma_{12}^2} \right) = -\frac{1}{2 \sqrt{n}} \sum_{i=1}^{n} \left( \frac{2 \Sigma_{12}^2 \Delta^2 + \Lambda_{12} \mu_{2i}^2}{(\mu_i \mu_{2i} - \Sigma_{12}^2 \Delta^2)^2} \right) \cdot \int_{0}^{\infty} \frac{\Sigma_{12}^2 T^4}{((\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{12} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2)^2} dx := I_{12}^\Sigma.
\]

By symmetry, we can obtain $I_{23}^\Sigma$, $I_{33}^\Sigma$, and $I_{22}^\Sigma$ by simply switching the indices 1 and 2 in $I_{12}^\Sigma$ and $I_{11}^\Sigma$. The asymptotic variance for $(\hat{\Sigma}_{11}, \hat{\Sigma}_{12}, \hat{\Sigma}_{22})$ is given by

\[
\Pi = \left( \begin{array}{ccc}
I_{11}^\Sigma & I_{12}^\Sigma & I_{12}^\Sigma \\
I_{12}^\Sigma & I_{22}^\Sigma & I_{22}^\Sigma \\
I_{12}^\Sigma & I_{22}^\Sigma & I_{22}^\Sigma 
\end{array} \right)^{-1}.
\]

Similarly derivations on 1/n-scaled likelihood can show that the asymptotic variance for $(\hat{\Lambda}_{11}, \hat{\Lambda}_{22})$ is given by

\[
\left( \begin{array}{c}
I_{11}^\Lambda \\
I_{22}^\Lambda 
\end{array} \right)^{-1} = \left( \begin{array}{cc}
2\Lambda_{11}^2 & 2\Lambda_{22}^2 
\end{array} \right).
\]

Notice that the above integrals have explicit forms, which can be obtained easily by Mathematica. However, the explicit formulae are tedious and hence omitted here.

### A.2 Proof of Theorem 2

The proof is made of the following steps: first, we show that the differences of the score vectors, scaled by appropriate rates, and their target “conditional expectations” converge uniformly to 0, and
satisfy the identification condition. (This step is easily achieved from the following calculations).
Second, we derive the stable CLTs for the differences, and this where the higher order moments of volatility process come into play. Third, we solve the equations that the target equal to 0, and find that the difference between the pseudo true parameter values and the parameters of interest are asymptotically negligible. Last, we use the sandwich theorem and consistency to establish the CLT for the QMLE.

To clarify our notation, we use subscript 0 to mark quantities that are made of true values. The true values for the Brownian covariances are obviously written in integral forms. The pseudo true parameters are marked with a superscript such as \( \bar{\Sigma} \) and \( \bar{\Lambda} \), and the QML estimators are marked as \( \hat{\Sigma} \) and \( \hat{\Lambda} \). The other \( \Sigma, \Lambda \) etc without any special marks represent any parameter values within the parameter space, which is assumed to be a compact set.

The drift term can be ignored without loss of generality, as a simple change of measure argument makes it sufficient to investigate the case without drift.

Recall that in (4), we have

\[ L = -n \log(2\pi) - \frac{1}{2} \log(\det \Omega) - \frac{1}{2} r' \Omega^{-1} r \]

Now we consider the following function:

\[ \bar{L} = -n \log(2\pi) - \frac{1}{2} \log(\det \Omega) - \frac{1}{2} \text{tr}(\Omega^{-1} \Omega_0) \]

where the subscript 0 denotes the true value, 

\[ \Omega_0 = \begin{pmatrix} \Omega^{11}_0 & \Omega^{12}_0 \\ \Omega^{21}_0 & \Omega^{22}_0 \end{pmatrix} \]

with \( \Omega^{11}_{0,i,i} = \int_{t_{i-1}}^{t_i} \Sigma_{it,t} dt + 2\Lambda_{0,ii} \\
\Omega^{12}_{0,i,i+1} = \Lambda_{0,ii}, \text{ and } \Omega^{12}_{0,i,i} = \Omega^{21}_{0,i,i} = \int_{t_{i-1}}^{t_i} \Sigma_{12,t} dt \).

Therefore, the difference between \( L \) and \( \bar{L} \) is given by:

\[ L - \bar{L} = \frac{1}{2} \text{tr} \left( \Omega^{-1}(rr' - \Omega_0) \right) = \frac{1}{2} \text{tr} \left( \begin{pmatrix} U' & U' \end{pmatrix} V^{-1} \left( \begin{array}{cc} U & U \end{array} \right) (rr' - \Omega_0) \right) \]

\[ = \frac{1}{2} \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \omega_{il}^{ls} \left( (\Delta_{i}y_l)(\Delta_{i}y_s) - \int_{t_{i-1}}^{t_i} \Sigma_{ls,t} dt \right) + \frac{1}{2} \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \sum_{j<i}^{n} \omega_{ij}^{ls} \Delta_{i}^{n} \Delta_{j}^{n} \] \( \sin \left( k_i n + 1 \pi \right) \sin \left( k_j n + 1 \pi \right) \)

Consider \( \omega_{ij}^{11} \) first.

\[ \omega_{ij}^{11} = \frac{2}{n+1} \sum_{k=1}^{n} \frac{\mu_{2k}}{\mu_{1k} \mu_{2k} - \Sigma_{12}^2} \sin \left( \frac{k_i}{n+1} \pi \right) \sin \left( \frac{k_j}{n+1} \pi \right) \]

45
We define the score vectors and their targets as

\[
\omega_{1k} = \frac{1}{n+1} \sum_{k=1}^{n} \frac{\mu_{2k}}{\mu_{1k} \mu_{2k} - \Sigma_{12}^2} \left( \cos \left( \frac{k(i-j)}{n+1} \pi \right) - \cos \left( \frac{k(i+j)}{n+1} \pi \right) \right).
\]

Similarly, we can derive

\[
\omega_{2k} = \frac{1}{2(n+1)} \sum_{k=1}^{n} \frac{\mu_{2k}}{\mu_{1k} \mu_{2k} - \Sigma_{12}^2} \left( \sin \left( \frac{(k+1) i}{n+1} \pi \right) - \sin \left( \frac{(k-1) i}{n+1} \pi \right) \right)
\]

hence, for any \( n^{1/2+\delta} \leq i \leq \frac{n+1}{2} \), we have

\[
\left| \frac{1}{n+1} \sum_{k=1}^{n} \frac{\mu_{2k}}{\mu_{1k} \mu_{2k} - \Sigma_{12}^2} \left( \cos \left( \frac{k l}{n+1} \pi \right) \right) \right| \leq C \frac{1}{n} \sum_{k=1}^{n} \frac{1}{n+1} \left( \frac{1}{n} + \frac{\sqrt{n}}{n} \right)^2 \sim o(\sqrt{n})
\]

hence, for any \( n^{1/2+\delta} \leq i \leq n - n^{1/2+\delta} \),

\[
\omega_{1i} = \left( \int_0^\infty \frac{\Sigma_{11} T + \Lambda_{11} \pi^2 x^2}{(\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2} \cdot \sqrt{n}(1+o(1)).
\]

Similarly, we can derive

\[
\omega_{2i} = \left( \int_0^\infty \frac{-\Sigma_{12} T}{(\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2} \cdot \sqrt{n}(1+o(1)).
\]

To simply our notation, let

\[
\omega_{1}(\Sigma, \Lambda, x) = \frac{\Sigma_{22} T + \Lambda_{22} \pi^2 x^2}{(\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2}
\]

\[
\omega_{2}(\Sigma, \Lambda, x) = \frac{\Sigma_{11} T + \Lambda_{11} \pi^2 x^2}{(\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2}
\]

\[
\omega_{12}(\Sigma, \Lambda, x) = \frac{-\Sigma_{12} T}{(\Sigma_{11} T + \Lambda_{11} \pi^2 x^2)(\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2}
\]

We define the score vectors and their targets as

\[
\Psi_\Sigma = - \frac{1}{\sqrt{n}} \frac{\partial L}{\partial \Sigma}, \quad \Psi_\Lambda = - \frac{1}{\sqrt{n}} \frac{\partial L}{\partial \Lambda}, \quad \Psi_\Lambda = - \frac{1}{\sqrt{n}} \frac{\partial L}{\partial \Lambda}.
\]

46
where
\[
\frac{\partial}{\partial \Sigma} = \begin{pmatrix}
\frac{\partial}{\partial \Sigma_{11}} \\
\frac{\partial}{\partial \Sigma_{12}} \\
\frac{\partial}{\partial \Sigma_{22}}
\end{pmatrix}, \quad \text{and} \quad \frac{\partial}{\partial \Lambda} = \begin{pmatrix}
\frac{\partial}{\partial \Lambda_{11}} \\
\frac{\partial}{\partial \Lambda_{12}} \\
\frac{\partial}{\partial \Lambda_{22}}
\end{pmatrix}.
\]

Then we have
\[
\Psi_{\Sigma} - \bar{\Psi}_{\Sigma} = \frac{1}{2\sqrt{n}} \left( M_1^{(\Sigma)} + 2M_2^{(\Sigma)} + 2M_3^{(\Sigma)} + M_4^{(\Sigma)} \right),
\]

where
\[
M_1^{(\Sigma)} = \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \left( (\Delta_i y_l)(\Delta_i y_s) - \int_{\tau_{i-1}}^{\tau_{i}} \xi_{ls,t} dt \right)
\]
\[
M_2^{(\Sigma)} = \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \sum_{j<i}^{n} \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \Delta_i y_l \Delta_j y_s
\]
\[
M_3^{(\Sigma)} = \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \Delta_i \varepsilon_l \Delta_j y_s
\]
\[
M_4^{(\Sigma)} = \sum_{l=1}^{2} \sum_{s=1}^{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \left( \Delta_i \varepsilon_l \Delta_j y_s - E(\Delta_i \varepsilon_l \Delta_j y_s) \right).
\]

Following the same argument in Xiu (2010) and Theorem 7.1 in Jacod (2012), we can show
\[
n^{-\frac{1}{4}} \left( M_1^{(\Sigma)} + 2M_2^{(\Sigma)} \right) \xrightarrow{\mathcal{L}_x} MN(0, \text{Avar}(2)(\Sigma)),
\]

where
\[
\text{Avar}(2)(\Sigma) = \lim_{n \to \infty} 4n^{-\frac{1}{4}} \sum_{l,s,u,v=1}^{2} \sum_{i=1}^{n} \sum_{j<i}^{n} \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \frac{\partial \omega_{vu}^{i}}{\partial \Sigma'} \Sigma_{sv,t} \Sigma_{ul,t} \Delta^2
\]
\[
= 2T \sum_{l,s,u,v=1}^{2} \int_{0}^{\infty} \frac{\partial \omega_{ls}^{i}(\Sigma, \Lambda, x)}{\partial \Sigma} \frac{\partial \omega_{vu}^{i}(\Sigma, \Lambda, x)}{\partial \Sigma'} dx \int_{0}^{T} \Sigma_{sv,t} \Sigma_{ul,t} dt. \tag{A.1}
\]

All the elements of the covariance matrix have closed-forms. Also, we have
\[
n^{-\frac{1}{4}} 2M_3^{(\Sigma)} \xrightarrow{\mathcal{L}_x} MN(0, \text{Avar}(3)(\Sigma)),
\]

where
\[
\text{Avar}(3)(\Sigma)
\]
\[
= \lim_{n \to \infty} 4 \sum_{j=1}^{n} n^{-\frac{1}{4}} \sum_{l,s,u,v=1}^{2} \Lambda_{0,lt} \left( \frac{\partial \omega_{ls}^{i}}{\partial \Sigma} \left( 2 \frac{\partial \omega_{lv}^{j}}{\partial \Sigma'} - \frac{\partial \omega_{lv}^{j-1}}{\partial \Sigma'} - \frac{\partial \omega_{lv}^{j+1}}{\partial \Sigma'} \right) \right) \Sigma_{sv,t} \Delta
\]
\[
= 4 \sum_{l,s,u,v=1}^{2} \Lambda_{0,lt} \int_{0}^{\infty} \frac{\partial \omega_{ls}^{i}(\Sigma, \Lambda, x)}{\partial \Sigma} \frac{\partial \omega_{vu}^{i}(\Sigma, \Lambda, x)}{\partial \Sigma'} \pi^2 x^2 dx \int_{0}^{T} \Sigma_{sv,t} dt. \tag{A.3}
\]

47
Finally, we have
\[ n^{-\frac{1}{2}} M_4 (\Sigma) \xrightarrow{\mathcal{L}} N(0, \text{Avar}^{(4)}(\Sigma)), \]
where
\[
\text{Avar}^{(4)}(\Sigma) = \lim_{n \to \infty} n^{-\frac{1}{2}} \sum_{i,j,k,l=1}^{n} \left( \frac{\partial \omega_{ij}^{1}}{\partial \Sigma} \frac{\partial \omega_{kl}^{1}}{\partial \Sigma} K_{i,j,k,l}^{1} + 4 \frac{\partial \omega_{ij}^{2}}{\partial \Sigma} \frac{\partial \omega_{kl}^{2}}{\partial \Sigma'} K_{i,j,k,l}^{2} \right)
\]
\[ = \lim_{n \to \infty} n^{-\frac{1}{2}} \left( V_1 \left( \frac{\partial \omega_{11}^{1}}{\partial \Sigma} \frac{\partial \omega_{11}^{1}}{\partial \Sigma} \right) + V_2 \left( \frac{\partial \omega_{11}^{1}}{\partial \Sigma} \frac{\partial \omega_{11}^{1}}{\partial \Sigma} \right) + 2 V_2 \left( \frac{\partial \omega_{12}^{1}}{\partial \Sigma}, \frac{\partial \omega_{12}^{1}}{\partial \Sigma} \right) + V_1 \left( \frac{\partial \omega_{22}^{2}}{\partial \Sigma}, \frac{\partial \omega_{22}^{2}}{\partial \Sigma} \right) + V_2 \left( \frac{\partial \omega_{22}^{2}}{\partial \Sigma}, \frac{\partial \omega_{22}^{2}}{\partial \Sigma} \right) \right),
\]
and
\[
V_1 \left( \frac{\partial \omega_{ii}^{ll}}{\partial \Sigma}, \frac{\partial \omega_{ii}^{ll}}{\partial \Sigma} \right) = \left( \sum_{i=1}^{n-1} - 8 \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma'} + 2 \frac{\partial \omega_{i,i}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma'} + 4 \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma'} \right)
\[ + 2 \sum_{i=1}^{n} \left( \frac{\partial \omega_{i,i}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i,i}^{ll}}{\partial \Sigma'} \right) \text{cum}_4[\varepsilon_i] \sim O(1),
\]
\[
V_2 \left( \frac{\partial \omega_{ii}^{12}}{\partial \Sigma}, \frac{\partial \omega_{ii}^{12}}{\partial \Sigma} \right) = 2 (\Lambda_{0,ii}^0)^2 \sum_{i,j=1}^{n} \left( \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma} \left( \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} \right) \right)
\[ + \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \left( \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} \right) \right)
\sim 2 (\Lambda_{0,ii}^0)^2 \left( \int_0^\infty \frac{\partial \omega_{i,i}^{12}}{\partial \Sigma} \frac{\partial \omega_{i,i}^{12}}{\partial \Sigma'} \pi^4 x^4 dx \right) n^{\frac{3}{2}},
\]
\[
V_2 \left( \frac{\partial \omega_{ii}^{12}}{\partial \Sigma}, \frac{\partial \omega_{ii}^{12}}{\partial \Sigma} \right) = 2 (\Lambda_{0,ii}^0) (\Lambda_{0,ss}^0) \sum_{i,j=1}^{n} \left( \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma} \left( \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma'} \right) \right)
\[ + \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \left( \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{i,i+1}^{12}}{\partial \Sigma'} \right) \right)
\sim 2 (\Lambda_{0,ii}^0) (\Lambda_{0,ss}^0) \left( \int_0^\infty \frac{\partial \omega_{i,i}^{12}}{\partial \Sigma} \frac{\partial \omega_{i,i}^{12}}{\partial \Sigma'} \pi^4 x^4 dx \right) n^{\frac{1}{2}}.
\]
Here, \( K_{i,j,kl}^{11}, K_{i,j,kl}^{22}, K_{i,j,kl}^{11} \) and \( K_{i,j,kl}^{22} \) are the corresponding cumulants for \( \Delta_\varepsilon^0 \) and \( \Delta_\varepsilon^2 \), and \( \text{cum}_4[\varepsilon_1] \) and \( \text{cum}_4[\varepsilon_2] \) are the fourth cumulants of \( \varepsilon_1 \) and \( \varepsilon_2 \).

Therefore, we have
\[
\text{Avar}^{(4)}(\Sigma) = 2 \sum_{i,s=1}^{2} \Lambda_{0,ii}^0 \Lambda_{0,ss}^0 \int_0^\infty \frac{\partial \omega_{ii,s}^s(\Sigma, \Lambda, x)}{\partial \Sigma'} \frac{\partial \omega_{ii,s}^s(\Sigma, \Lambda, x)}{\partial \Sigma'} \pi^4 x^4 dx. \quad (A.4)
\]
48
Similarly, we can obtain
\[ n^{-\frac{1}{2}} \left( \Psi_{11} - \bar{\Psi}_{11} \right) \xrightarrow{\mathcal{L}} N \left( 0, \frac{1}{4} \left( \text{Avar}^{(2)}(\Sigma) + \text{Avar}^{(3)}(\Sigma) + \text{Avar}^{(4)}(\Sigma) \right) \right). \]

Similarly, we can obtain
\[ n^{-\frac{1}{2}} \left( \Psi_{22} - \bar{\Psi}_{22} \right) \xrightarrow{\mathcal{L}} N \left( 0, \frac{1}{4} \left( \frac{2(\Lambda_{0,11})^2 + \text{cum}_{4}[\varepsilon_1]}{\Lambda_{11}^2} \right) \right). \quad (A.5) \]

Further, we need to solve \( \bar{\Psi}_\Sigma = 0 \) and \( \bar{\Psi}_\Lambda = 0 \) for the pseudo-true parameters \( \theta^* \), and show that the distance between \( \theta^* \) and the values of interest are negligible asymptotically. In fact, for any \( \Sigma_{uv} \in \{ \Sigma_{11}, \Sigma_{12}, \Sigma_{22} \} \), we have
\[ \bar{\Psi}_{\Sigma_{uv}} = \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \right) + \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \right) \right\} \]
\[ = \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \right) + \text{tr} \left( \Omega^{-1} \left( \Omega + J \otimes (\Lambda_0 - \Lambda) + \Gamma \right) \right) \right\} \]
\[ = \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} J \otimes (\Lambda_0 - \Lambda) \right) + \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \Gamma \right) \right\} \]
\[ = \frac{1}{2\sqrt{n}} \left\{ 2 \sum_{l=1}^{n} \left( \frac{\partial \omega_{ll}^T}{\partial \Sigma_{uv}} - \frac{\partial \omega_{ll}^T}{\partial \Sigma_{uv}} \right) (\Lambda_0, - \Lambda) \Gamma \right\} \]
\[ + \frac{1}{2\sqrt{n}} \left\{ \left( \int_{0}^{\infty} \partial \omega_{ll}^T(\Sigma, \Lambda, x) dx \right) \left( \int_{0}^{\infty} \Sigma_{\Delta, t} dt - \Sigma_{\Delta} \right) (1 + o(1)) \right\} \]
\[ + \frac{1}{2\sqrt{n}} \left\{ \left( \int_{0}^{\infty} \partial \omega_{ll}^T(\Sigma, \Lambda, x) dx \right) \left( \int_{0}^{\infty} \Sigma_{\Delta, t} dt - \Sigma_{\Delta} \right) (1 + o(1)) \right\} \]

where \( \Lambda_0 \) denotes the true covariance matrix of noise, \( \Gamma \) is block diagonal matrix, with \( \Gamma_{ll} = \int_{t_{l-1}}^{t_l} \Sigma_{\Delta, t} dt - \Sigma_{\Delta} \), and \( J \) is an \( n \times n \) tridiagonal matrix with matrix diagonal elements equal to 2 and off-diagonal elements equal to \(-1\).

Similarly, for \( \Lambda_{uu} \in \{ \Lambda_{11}, \Lambda_{22} \} \), we have
\[ \bar{\Psi}_{\Lambda_{uu}} = \frac{1}{2n} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \right) + \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \right) \right\} \]
\[ = \frac{1}{2n} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \right) + \text{tr} \left( \Omega^{-1} \left( \Omega + J \otimes (\Lambda_0 - \Lambda) + \Gamma \right) \right) \right\} \]
\[ = \frac{1}{2n} \left\{ \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} J \otimes (\Lambda_0 - \Lambda) \right) + \text{tr} \left( \Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \Gamma \right) \right\} \]
\[ = \frac{1}{2n} \left\{ 2 \sum_{l=1}^{n} \left( \frac{\partial \omega_{ll}^T}{\partial \Lambda_{uu}} - \frac{\partial \omega_{ll}^T}{\partial \Lambda_{uu}} \right) (\Lambda_0, - \Lambda) \Gamma \right\} \]
\[ + \frac{1}{2n} \left\{ \left( \int_{0}^{\infty} \partial \omega_{ll}^T(\Sigma, \Lambda, x) dx \right) \left( \int_{0}^{\infty} \Sigma_{\Delta, t} dt - \Sigma_{\Delta} \right) (1 + o(1)) \right\} \]
\[ + \frac{1}{2n} \left\{ \left( \int_{0}^{\infty} \partial \omega_{ll}^T(\Sigma, \Lambda, x) dx \right) \left( \int_{0}^{\infty} \Sigma_{\Delta, t} dt - \Sigma_{\Delta} \right) (1 + o(1)) \right\} \]
Therefore, solving for $\Sigma$ and $\Lambda$, we obtain:

$$
\hat{\Lambda}_{ll} = \Lambda_{0, ll} + \frac{\overline{\Lambda}_l^2}{\sqrt{n}} \left\{ \sum_{i=1}^{2} \left( \int_{0}^{\infty} \frac{\partial \omega^{iJ}(\Sigma_{ij}, \Lambda)}{\partial \Lambda_{uu}} dx \right) \left( \int_{0}^{T} \Sigma_{iJ,t} dt - \Sigma_{iJ,T} \right) \right\} (1 + o_p(1)), \text{ for } l = 1, 2,
$$

$$
\Sigma_{ls} = \frac{1}{T} \int_{0}^{T} \Sigma_{ls,t} dt + O_p(n^{-\frac{1}{2}}) = \Sigma_{0,ls} + O_p(n^{-\frac{1}{2}}), \text{ for } l, s = 1, 2.
$$

Further, applying Theorem 2 in Xiu (2010), we have

$$
\hat{\Sigma}_{ls} - \overline{\Sigma}_{ls} = o_p(1), \text{ and } \hat{\Lambda}_{ll} - \overline{\Lambda}_{ll} = o_p(1),
$$

hence consistency is established.

To find the central limit theorem, we do the usual “sandwich” calculations. Denote,

$$
\frac{\partial \overline{\Psi}_\Sigma}{\partial \Sigma} = \left( \begin{array}{ccc}
\frac{\partial \overline{\Psi}_{\Sigma_{11}}}{\partial \Sigma_{11}} & \frac{\partial \overline{\Psi}_{\Sigma_{11}}}{\partial \Sigma_{12}} & \frac{\partial \overline{\Psi}_{\Sigma_{21}}}{\partial \Sigma_{11}} \\
\frac{\partial \overline{\Psi}_{\Sigma_{12}}}{\partial \Sigma_{12}} & \frac{\partial \overline{\Psi}_{\Sigma_{12}}}{\partial \Sigma_{22}} & \frac{\partial \overline{\Psi}_{\Sigma_{22}}}{\partial \Sigma_{12}} \\
\frac{\partial \overline{\Psi}_{\Sigma_{21}}}{\partial \Sigma_{21}} & \frac{\partial \overline{\Psi}_{\Sigma_{21}}}{\partial \Sigma_{22}} & \frac{\partial \overline{\Psi}_{\Sigma_{22}}}{\partial \Sigma_{22}} 
\end{array} \right) \xrightarrow{P} \frac{\partial \Psi_{\Sigma_0}}{\partial \Sigma},
$$

where

$$
\frac{\partial \Psi_{\Sigma_{0, uv}}}{\partial \Sigma_{ij}} = -\frac{T}{2} \left( \int_{0}^{\infty} \frac{\partial \omega^{ij}(\Sigma_{0}, \Lambda_{0}, x)}{\partial \Sigma_{uv}} dx \right) - \frac{T}{2} \left( \int_{0}^{\infty} \frac{\partial \omega^{ij}(\Sigma_{0}, \Lambda_{0}, x)}{\partial \Sigma_{uv}} dx \right) \{u \neq v\}
$$

and $\Sigma_0$ denotes the true parameter value. So, the central limit theorem is:

$$
n^{\frac{1}{2}}(\hat{\Sigma} - \Sigma_0) = n^{\frac{1}{2}} \left( \begin{array}{ccc}
\hat{\Sigma}_{11} - \frac{1}{T} \int_{0}^{T} \Sigma_{11,t} dt \\
\hat{\Sigma}_{12} - \frac{1}{T} \int_{0}^{T} \Sigma_{12,t} dt \\
\hat{\Sigma}_{22} - \frac{1}{T} \int_{0}^{T} \Sigma_{22,t} dt 
\end{array} \right) \xrightarrow{L} MN(0, V_Q),
$$

where

$$
P_Q = \frac{1}{4} \left( \frac{\partial \Psi_{\Sigma_0}}{\partial \Sigma} \right)^{-1} \left( \text{Avar}^{(2)}(\Sigma_0) + \text{Avar}^{(3)}(\Sigma_0) + \text{Avar}^{(4)}(\Sigma_0) \right) \left( \frac{\partial \Psi_{\Sigma_0}}{\partial \Sigma} \right)^{-1}.
$$

Note that

$$
\frac{\partial \Psi_{\Lambda_0}}{\partial \Lambda} = \left( \begin{array}{cc}
-\frac{1}{2N_{0,11}} & -\frac{1}{2N_{0,22}} \\
-\frac{1}{2N_{0,21}} & -\frac{1}{2N_{0,12}} 
\end{array} \right),
$$

hence the CLT for $\hat{\Lambda}$ follows immediately from (A.5). This concludes the proof of Theorem 2.

### A.3 Proof of Corollary 1

Denote $\Delta_i = \Delta(1 + \xi_i)$ and $\xi_i$ is i.i.d $O_p(1)$. Note that

$$
\Omega = \Omega + \Delta \Sigma \otimes \Theta,
$$
where $\Xi = \text{diag}(\xi_1, \ldots, \xi_i, \ldots, \xi_n)$, and $\bar{\Omega}$ is the covariance matrix in the equidistant case with $\Delta$ replaced by $\bar{\Delta}$. It turns out that

$$\Omega^{-1} = (\Omega(I + \bar{\Delta}^{-1}\Sigma \otimes \Xi))^{-1} = (I + \bar{\Delta}^{-1}\Sigma \otimes \Xi)^{-1}\Omega^{-1}$$

$$= \Omega^{-1} + \sum_{k=1}^{\infty} (-1)^k \bar{\Delta}^k (\Omega^{-1}\Sigma \otimes \Xi)^k\Omega^{-1}.$$ 

For any $\theta_1, \theta_2 \in \{\Sigma_{11}, \Sigma_{12}, \Sigma_{22}\}$, we have

$$\frac{\partial \Omega}{\partial \theta_1} = \frac{\partial \bar{\Omega}}{\partial \theta_1} + \bar{\Delta} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi$$

hence,

$$E\left( \frac{\partial \log(\det \Omega)}{\partial \theta_1} \right) = E\left( \text{tr}\left( \Omega^{-1} \frac{\partial \Omega}{\partial \theta_1} \right) \right)$$

$$= E\left( \text{tr}\left( \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} \right) + \bar{\Delta} E\left( \text{tr}\left( \Omega^{-1}\Sigma \otimes \Xi^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} + \Omega^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi \right) \right) \right)$$

$$+ \Delta^2 E\left( \text{tr}\left( (\Omega^{-1}\Sigma \otimes \Xi)^2 \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \Omega^{-1}\Sigma \otimes \Xi \Omega^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi \right) \right) + o(\Delta^2).$$

Because $E(\Xi) = 0$,

$$E\left( \text{tr}\left( \Omega^{-1}\Sigma \otimes \Xi^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} + \Omega^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi \right) \right) = 0.$$

Also,

$$E\left( \text{tr}\left( (\Omega^{-1}\Sigma \otimes \Xi)^2 \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \Omega^{-1}\Sigma \otimes \Xi \Omega^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi \right) \right)$$

$$= \text{tr}\left( \Omega^{-1}(\Sigma \otimes I) D(\Sigma \otimes I) \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \Omega^{-1}(\Sigma \otimes I) D(\frac{\partial \Sigma}{\partial \theta_1} \otimes I) \right) \text{var}(\xi),$$

where

$$D = \left( \begin{array}{cc} \text{diag}(\bar{\Omega}_{11}^{-1}) & \text{diag}(\bar{\Omega}_{12}^{-1}) \\ \text{diag}(\bar{\Omega}_{21}^{-1}) & \text{diag}(\bar{\Omega}_{22}^{-1}) \end{array} \right)$$

and $\bar{\Omega}_{ij}^{-1}$ is the $(i, j)$ block of the $\bar{\Omega}^{-1}$. Therefore,

$$E\left( - \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} \right) = - \frac{1}{2} E\left( \frac{\partial^2 \log(\det \Omega)}{\partial \theta_1 \partial \theta_2} \right) = \frac{1}{2} \text{tr}\left( \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_2} \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} + \phi_{\theta_1, \theta_2}(\Sigma, \bar{\Omega}, \bar{\Delta}) \text{var}(\xi) + o(\Delta^2) \right)$$

where

$$\phi_{\theta_1, \theta_2}(\Sigma, \bar{\Omega}, \bar{\Delta}) = - \frac{1}{2} \frac{\partial}{\partial \theta_2} \text{tr}\left( \Omega^{-1}(\Sigma \otimes I) D(\Sigma \otimes I) \Omega^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \Omega^{-1}(\Sigma \otimes I) D(\frac{\partial \Sigma}{\partial \theta_1} \otimes I) \right) \Delta^2.$$ 

In fact, we can show that

$$\phi_{\theta_1, \theta_2}(\Sigma, \bar{\Omega}, \bar{\Delta}) = o(\Delta^{3/2}).$$

Hence, the new fisher information converges to the previous one given in the proof of Theorem 1, as $\bar{\Delta} \to 0$, which concludes the proof.
A.4  Proof of Theorem 3

Since $n_1 \gg n_2$, we have:

$$\Delta^{n_1,n_2} = (\Delta^{n_2,n_1})' = \begin{pmatrix} \Delta^{n_1}_{1:m} & 0 & 0 & \ldots & 0 \\ 0 & \Delta^{n_1}_{m+1:2m} & 0 & \ddots & \vdots \\ 0 & 0 & \Delta^{n_1}_{2m+1:3m} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 0 & \Delta^{n_1}_{n_1-m+1:n_1} \end{pmatrix}_{n_1 \times n_2},$$

where $\Delta^{n_1}_{km+1:(k+1)m} = (\Delta^{n_1}_{km+1}, \Delta^{n_1}_{km+2}, \ldots, \Delta^{n_1}_{(k+1)m})'$ is a $m$-dimensional vector.

Using similar orthogonal matrices $U^{n_1}$, and $U^{n_2}$, such that

$$\begin{pmatrix} U^{n_1} \\ U^{n_2} \end{pmatrix} \Omega \begin{pmatrix} (U^{n_1})' \\ (U^{n_2})' \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{pmatrix} =: V,$$

where

$$u_{ij}^{nk} = \sqrt{2} \sin\left(\frac{i \cdot j}{n_k + 1}\pi\right), \quad i, j = 1, \ldots, n_k, k = 1, 2,$$

$$V_{ii} = \text{diag}(\mu_{ij}) = \text{diag}(\Sigma_{ii} \Delta_i + 2\Lambda_{ii} \left(1 - \cos\left(\frac{j}{n_i + 1}\pi\right)\right)), \quad i = 1, 2, \text{and } j = 1, \ldots, n_i,$$

$$V_{12} = (v_{ij}^{12}) = \Sigma_{12} U^{n_1} \Delta^{n_1,n_2}(U^{n_2})'.$$

So for $i = 1, 2, \ldots, n_1$, $j = 1, 2, \ldots, n_2$, we have

$$v_{ij}^{12} = \Sigma_{12} \Delta_1 \sum_{k=1}^{n_1} u_{i,k}^{n_1} u_{j,k(n-1)/m+1}^{n_2} = \Sigma_{12} \Delta_1 \sum_{l=1}^{n_2} u_{j,l}^{n_2} \sum_{k=m(l-1)+1}^{ml} u_{i,k}^{n_1} \left(\frac{\sin n_2(j+n_1-im-in_2)\pi}{2(1+n_1)(1+n_2)}\right).$$

Moreover,

$$\Omega^{-1} = \begin{pmatrix} (U^{n_1})' \\ (U^{n_2})' \end{pmatrix} V^{-1} \begin{pmatrix} U^{n_1} \\ U^{n_2} \end{pmatrix},$$

where by the Woodbury formula:

$$V^{-1} = \begin{pmatrix} V_{11} & V_{12} V_{22}^{-1} V_{12}' & V_{12}' V_{22}^{-1} & -V_{12} V_{22}^{-1} V_{12}' V_{22}^{-1} \\ -V_{22} V_{12} V_{22}^{-1} V_{12}' & V_{22}^{-1} + V_{22}^{-1} V_{12} V_{22}^{-1} V_{12}' V_{22}^{-1} \\ V_{11} V_{22}^{-1} V_{12}' & -V_{11} V_{22}^{-1} V_{12}' V_{22}^{-1} V_{12}' V_{22}^{-1} & V_{11} V_{22}^{-1} V_{12}' V_{22}^{-1} \end{pmatrix}$$

Then, for any $\theta_1, \theta_2 \in \{\Sigma_{11}, \Sigma_{12}, \Sigma_{22}\}$,

$$\frac{\partial \log(\det \Omega)}{\partial \theta_1} = \text{tr}(\Omega^{-1} \frac{\partial \Omega}{\partial \theta_1}) = \text{tr}(V^{-1} \frac{\partial V}{\partial \theta_1})$$
Therefore, $\approx i \approx i > n$ and decrease rapidly once $n \to \infty$.

On the other hand, $V$ is close to 0. As $n \to \infty$, we can prove that the dominant terms in the second brackets of the last summation are contributed by those such that either $\sin \frac{(j + jn_1 - im - in_1)\pi}{2(1 + n_1)(1 + n_2)}$ or $\sin \frac{(j + jn_1 + im + in_1)\pi}{2(1 + n_1)(1 + n_2)}$ is close to 0. As $n \to \infty$, for each $1 \leq j \leq n_2^{1/2+\delta}$, there exists $m$ different $i \in [1, n_1]$ such that $\sin \frac{(j + jn_1 - im - in_1)\pi}{2(1 + n_1)(1 + n_2)} \approx 0$, or $\sin \frac{(j + jn_1 + im + in_1)\pi}{2(1 + n_1)(1 + n_2)} \approx 0$.

On the other hand, $\left(\frac{\sin \frac{im\pi}{2(1+n_1)}}{\sin \frac{im\pi}{2(1+n_1)}}\right)^2 \approx m^2$, when $i \in [1, n_2^{1/2+\delta}]$ and decrease rapidly once $i > n_2$. Hence, the dominant term when $j \in [1, n_2^{1/2+\delta}]$ is the one with $i \approx j$, in which case

$$\sin \frac{n_2(j + jn_1 - im - in_1)\pi}{2(1 + n_1)(1 + n_2)} \cos \frac{(i-j)\pi}{2} \approx n_2, \quad \text{and} \quad \sin \frac{n_2(j + jn_1 + im + in_1)\pi}{2(1 + n_1)(1 + n_2)} \cos \frac{(i+j)\pi}{2} \approx -1$$

Therefore,

$$\sum_{i=1}^{n_1} (v_{i,j}^2) \mu_{ii}^{-1} = n_1^{-1} n_2^{-1} \mu_{ij}^{-1} \Sigma_{i,j}^{-2} (1 + o(1))$$
so as $n_1 \to \infty$, $n_2 \to \infty$ and $m \to \infty$,

$$
\frac{1}{\sqrt{n_2}} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{22}} \right) = \frac{\Delta_2}{\sqrt{n_2}} \sum_{j=1}^{n_2} \left( \mu_j - \sum_{i=1}^{n_1} (v_{i,j}^2)^{-1} \right) \to \int_0^\infty \frac{\Sigma_{11} T}{\Sigma_{11} T (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12} T^2} \, dx
$$

Similarly, we can derive

$$
\text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{11}} \right) = \Delta_1 \left( \text{tr} \left( V_{11}^{-1} \right) + \text{tr} \left( V_{11}^{-1} V_{12} (V_{22} - V_{12} V_{11}^{-1} V_{12})^{-1} V_{12} V_{11}^{-1} \right) \right)
$$

$$
= \Delta_1 \left( \sum_{i=1}^{n_1} \mu_i^{-1} + \sum_{j=1}^{n_2} \left( \mu_j - \sum_{i=1}^{n_1} (v_{i,j}^2)^{-1} \right) \right) (1 + o(1))
$$

$$
= \left( \int_0^\infty \frac{\Sigma_{12} T}{\Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12} T} \, dx \right) (1 + o(1))
$$

and

$$
\text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{12}} \right) = \frac{2 \Sigma_{12}}{\Sigma_{11}} \text{tr} \left( - V_{11}^{-1} V_{12} (V_{22} - V_{12} V_{11}^{-1} V_{12})^{-1} V_{12} \right)
$$

$$
= \left( \int_0^\infty \frac{2 \Sigma_{12} T}{\Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12} T} \, dx \right) \sqrt{n_2} (1 + o(1)).
$$

Notice that because $n_1 \gg n_2$, $\text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{11}} \right)$ is dominated by the $O(\sqrt{n_1})$ term, hence the convergence rate of the liquid asset is not affected by the illiquid asset.

The Fisher information matrix can then be constructed by calculating the following derivatives multiplied by appropriate rates:

$$
I_{11}^\Sigma = \lim_{n_1 \to \infty, n_2 \to \infty} - \frac{1}{2 \sqrt{n_2}} \frac{\partial}{\partial \Sigma_{11}} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{11}} \right) = \int_0^\infty \frac{T^2}{2 (\Sigma_{11} T + \Lambda_{11} \pi^2 x^2) dx},
$$

$$
I_{22}^\Sigma = \lim_{n_1 \to \infty, n_2 \to \infty} - \frac{1}{2 \sqrt{n_2}} \frac{\partial}{\partial \Sigma_{12}} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{12}} \right) = \int_0^\infty \frac{T (\Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) + \Sigma_{12}^2 T)}{\left( \Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T \right)} dx,
$$

$$
I_{23}^\Sigma = \lim_{n_1 \to \infty, n_2 \to \infty} - \frac{1}{2 \sqrt{n_2}} \frac{\partial}{\partial \Sigma_{22}} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{22}} \right) = \int_0^\infty \frac{\Sigma_{11} T^2}{\left( \Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T \right)} dx,
$$

$$
I_{33}^\Sigma = \lim_{n_1 \to \infty, n_2 \to \infty} - \frac{1}{2 \sqrt{n_2}} \frac{\partial}{\partial \Sigma_{22}} \text{tr} \left( V^{-1} \frac{\partial V}{\partial \Sigma_{22}} \right) = \int_0^\infty \frac{\Sigma_{11} T^4}{2 (\Sigma_{11} (\Sigma_{22} T + \Lambda_{22} \pi^2 x^2) - \Sigma_{12}^2 T^2)} dx.
$$

$$
I_{12} = I_{13} = I_{23} = 0.
$$

Hence,

$$
\Pi_A = \begin{pmatrix}
I_{11}^\Sigma & 0 & 0 \\
0 & I_{22}^\Sigma & I_{23}^\Sigma \\
0 & I_{33}^\Sigma & I_{33}^\Sigma
\end{pmatrix}^{-1},
$$

and this concludes the proof.
Multivariate realised kernels: Consistent positive semi-definite estimators of the covariance of equity prices with noise and non-synchronous trading

Ole E. Barndorff-Nielsen a,b, Peter Reinhard Hansen c, Asger Lunde d,b,* Neil Shephard e,f

a The T.N. Thiele Centre for Mathematics in Natural Science, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark
b CREATE, University of Aarhus, Denmark
c Department of Economics, Stanford University, Landau Economics Building, 579 Serra Mall, Stanford, CA 94305-6072, USA
d School of Economics and Management, Aarhus University, Bartholins Allé 10, DK-8000 Aarhus C, Denmark
e Oxford-Man Institute, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6ED, UK
f Department of Economics, University of Oxford, UK

A R T I C L E   I N F O

Article history:
Received 14 July 2010
Received in revised form 14 July 2010
Accepted 28 July 2010
Available online 27 January 2011

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

A B S T R A C T

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 min intervals. We show that the new estimator is substantially more precise.

© 2011 Elsevier B.V. All rights reserved.

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 Eq. (1).

We study a d-dimensional log price process $X = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ 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 $t_i^{(0)}, t_i^{(1)}, \ldots$. This means the available database of prices is $X^{(i)}(t_i^{(j)})$, for $j = 1, 2, \ldots, N^{(i)}(1)$, and $i = 1, 2, \ldots, d$. 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 ($Y \in \mathcal{B} \mathcal{S} \mathcal{M}$) defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$,

$$Y(t) = \int_0^t a(u)du + \int_0^t \sigma(u)dW(u),$$

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 et al. (1996). If $Y \in \mathcal{B} \mathcal{S} \mathcal{M}$ then its ex-post covariation, which we will focus on for reasons explained in a moment, is

$$[Y](1) = \int_0^1 \Sigma(u)du,$$

where

$$[Y](1) = \lim_{n \to \infty} \sum_{j=1}^n \{Y(t_j) - Y(t_{j-1})\} \{Y(t_j) - Y(t_{j-1})\}'$$

** The second and fourth author are also affiliated with CREATE, 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 at: School of Economics and Management, Aarhus University, Bartholins Allé 10, DK-8000 Aarhus C, Denmark.

E-mail addresses: oebn@imf.au.dk (O.E. Barndorff-Nielsen),
peter.hansen@stanford.edu (P.R. Hansen), alunde@asb.dk (A. Lunde),
niel.shephard@economics.ox.ac.uk (N. Shephard).
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 realised kernel estimator converges at rate \( n^{1/3} \). 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 \( 390^{1/4} < 2000^{1/3} \). So the rates of convergence will not (alone) tell us which estimators will be most accurate in practice — even for the univariate estimation problem. In addition to being robust to noise with a general form of dependence, the \( n^{1/3} \) 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 causes 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 synchronise 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. Synchronising 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 and Yoshida (2005), Voev and Lunde (2007) and Large (unpublished paper). State 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 \( N_j(i, t) \), 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 et al. (1995), and Martens (unpublished paper) 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 \( t_1 = \max(t_1(1), \ldots, t_1(i)) \), and then subsequent refresh times as

\[
\tau_{j+1} = \max(t_{j+1}(1), \ldots, t_{j+1}(i)).
\]
Fig. 1. This figure illustrates Refresh Time in a situation with three assets. The dots represent the times \( t_i \); 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 \( n^{(3)} = 10 \).

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

The \( t_j \) is the time it has taken for all the assets to trade, i.e. all their posted price have been updated. \( t_2 \) is the first time when all the prices are again refreshed. This process is displayed in Fig. 1 for \( d = 3 \).

Our analysis will now be based on this time clock \( \{ t_j \} \). Our approach will be to:

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

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 \( n^{(i)} \). 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 \( p = dN / \sum_{i=1}^{n^{(i)}} n^{(i)} \). For the data in Fig. 1, \( p = 21 / 27 \approx 0.78 \).

2.2. Jittering end conditions

It turns out that our asymptotic theory dictates that we need to average 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 \( X_j = X(t_{n,j+m}), j = 1, 2, \ldots, n - 1 \), and

\[
X_0 = \frac{1}{m} \sum_{j=1}^{m} X(t_{n,j}) \quad \text{and} \quad X_n = \frac{1}{m} \sum_{j=1}^{m} X(t_{n,N-m+1}).
\]

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: \( \delta_j = \delta_j - \delta_{j-1}, j = 1, 2, \ldots, n \), that the realised kernels are built out of.

2.3. Realised kernel

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

\[
K(X) = \sum_{h=-n}^{n} k \left( \frac{h}{H} \right) I_h,
\]

where \( I_h = \sum_{j=0}^{n} x_j^{h} x_{j-h} \), for \( h \geq 0 \).

and \( I_h = I_h \) for \( h < 0 \). Here \( I_h \) is the \( h \)-th realised autocovariance and \( K : \mathbb{R} \rightarrow \mathbb{R} \) is a non-stochastic weight function. We focus on the class of functions, \( \mathcal{K} \), that is characterised by:

**Assumption K.** (i) \( k(0) = 1, k'(0) = 0 \); (ii) \( k \) is twice differentiable with continuous derivatives; (iii) define \( k_{2,0} = \int_0^\infty k''(x)^2 dx, k_{1,1} = \int_0^\infty k'(x)^2 dx \), and \( k_{2,2} = \int_0^\infty k''(x)^2 dx \) then \( k_{0,0}, k_{1,1} k_{2,2} < \infty \); (iv) \( \int_0^\infty k'(x) \exp(ix) dx \) exists for all \( \lambda \in \mathbb{R} \).

The assumption \( k(0) = 1 \) means \( I_0 \) gets unit weight, while \( k'(0) = 0 \) means the kernel gives close to unit weight to \( I_0 \) 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 and West (1987) and Andrews (1991)). But there are a number of important differences. For example, the sum that defines 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 \( \sigma_r = \sup_{x \in (0,1]} |\sigma(t)| \).

This can be relaxed to locally bounded if \((\mu, \sigma)\) is an Ito process — e.g. Mykland and Zhang (forthcoming).

2.4. Some assumptions about refresh time and noise

Having defined the positive semi-definite realised kernel, we will now write our assumptions about the refresh times \( \{ t_{n,j} \} \) and the market microstructure effects \( U \) that govern the properties of the vector returns \( \{ \delta_j \} \) 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 \( t_{N,i} - t_{N,i-1} = \Delta t_{N,i} = \frac{dN}{N} \) for all \( i \).

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

**Assumption D.** (i) \( E(D_{N,i}^{(j)} | F_{N_i-1}) \rightarrow \mathcal{D} \), as \( N \rightarrow \infty \). Here we assume \( \mathcal{D} \) is strictly positive càdlàg processes adapted to \( \{ F_t \} \); (ii) \( \max_{1 \leq j \leq N, j \neq 0} \mathcal{D}_{N,i} = \mathcal{O}_p(\sqrt{N}) \) for any \( j \); (iii) \( t_{N,0} \leq 0 \) and \( t_{N,N+1} \geq 1 \).

**Remark 1.** If we have Poisson sampling the durations are exponential and the max\( \max_{1 \leq j \leq N} \Delta t_{N,i} = \mathcal{O}_p(\log(N)/N) \), so max\( \max_{1 \leq j \leq N} \mathcal{D}_{N,i} = \mathcal{O}_p(\log(N)) \). Note both Barndorff-Nielsen et al. (2008) and Mykland and Zhang (2006) assume that max\( \max_{1 \leq j \leq N} \mathcal{D}_{N,i} = \mathcal{O}_p(1) \). Phillips and Yu (unpublished paper) provided a novel analysis of realised volatility under random times of trades. We use their
Assumption 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 $D_j^{(i)} \sim \exp(\lambda), j = 1, 2, \ldots, d$, so the refresh time durations are $D_{Ni,j} \sim \max(D_1^{(i)}, \ldots, D_d^{(i)})$, and so (e.g. Embrechts et al. (1997, p. 189)) the refresh times have the form of a renewal process $\tau_{Ni,j} - \tau_{Ni,j-1} = 1/N D_{Ni,j}$, $D_{Ni,j} = \sum_{j=1}^d 1/N D_j^{(i)}$. In particular $x_1(t) = \lambda^{-1} \sum_{j=1}^d 1/jx(t) = \lambda^{-1}(\sum_{j=1}^d 1/F_j - (\sum_{j=1}^d 1/F_j)^2)$. Of interest here 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 $\lambda^2/6$ as $d \to \infty$, so $\lim_{d \to \infty} x_1(t) = 1$. For $d = 1$, $x_1(t) = \lambda^{-1}$ and $x_2(t) = 2\lambda^{-1}$, so $x_1(t)/x_2(t) = 2\lambda^{-1}$.

2.4.2. Assumptions about the noise

The assumptions about the noise are stated in observations time — that is only model the noise at exactly the times where there are trades or quote updates. This follows, for example, Zhou (1998), Bandi and Russel (unpublished paper), Zhang et al. (2005), Barndorff-Nielsen et al. (2008) and Hansen and Lunde (2006).

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

Assumption A. Assume the component model $U_{Ni,j} = \psi_{Ni,j} + \varepsilon_{Ni,j}$, where $\psi_{Ni,j} = \sum_{h=0}^{\infty} \psi(h)(\tau_{Ni,j-1-h})N_{Ni,j-h}$, with $\varepsilon_{Ni,j} = \Delta^{-1/2}[\tilde{W}(t_{Ni,j}) - \tilde{W}(t_{Ni,j-1})]$. Here $\tilde{W}$ is a standard Brownian motion and $\{\varepsilon_{Ni,j}\}$ is a sequence of independent random variables, with $E(\varepsilon_{Ni,j}|F_{Ni-1}) = 0$ and $\text{var}(\varepsilon_{Ni,j}|F_{Ni-1}) = \Sigma(\varepsilon_{Ni,j}|F_{Ni-1})$. Further, $\psi_{Ni,j}$ and $\varepsilon_{Ni,j}$ are assumed to be independent, while $\{\psi(h), \Sigma(h)\}$ are bounded and adapted to $\{F_i\}$, with $\sum_{h=0}^{\infty} \psi(h) < \infty$ a.s. uniformly in $t$. We also assume that $\Sigma \in \mathbb{G}$.

Remark 2. The auxiliary Brownian motion $W$ facilitates a general form of endogenous noise through correlation between $W$ and the Brownian motion, $W$, that drives the underlying process, $Y$. In fact, the case $W = W$ is permitted under our assumptions.

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

The “local” long run variance of $v$ is given by $\Sigma(v) = \sum_{h=-\infty}^{\infty} y_h(v)$, where $y_h(v) = \sum_{j=1}^d \psi_{Ni,j}(\tau_{Ni,j-1-h})N_{Ni,j-h}$ for $h \geq 0$ and $y_h(v) = y_{-h}(v)$ for $h < 0$, so that the local long run variance of $U$ is given by $\Sigma(U) = \Sigma(v) + \Sigma(\varepsilon(v))$.

3 The earliest research on this that we know of is Jacod (1994), Mykland and Zhang (2006) and Barndorff-Nielsen et al. (2008) provided an analysis based on the assumption that $D_{Ni,j} = 0_p(1)$, which is perhaps too strong an assumption here (see Remark 1). Barndorff-Nielsen and Shephard (2005) allowed very general spacing, but assumed times and prices were independent. More recent important contributions include Hayashi et al. (unpublished paper), Li et al. (2009), Mykland and Zhang (forthcoming), and Jacod (unpublished paper) provide insightful analysis. It is convenient to define the average long run variance of $U$ by

$$\Sigma(U) = \int_0^1 \Sigma(u)du,$$

which is a $d \times d$ matrix. When $d = 1$ we write $o^2$ in place of $\Sigma(U)$, $\sigma^2(t)$ in place of $\Sigma(t)$, etc. $o^2$ 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

$K(X) = K(Y) + K(U, Y) + K(U) + K(U, Y) + K(U)$,

where $K(Y, U) = \sum_{h=-\infty}^{\infty} k(h) \sum_j y_j u_j - y_j$. 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 $K(Y) = \Omega_p(1)$. Then $K(X) = K(U) + \Omega_p(1)$.

Theorem 1 is a very powerful result for dealing with endogenous noise. Note whatever the relationship between $Y$ and $U$, if $K(U) = \Omega(0)$ then $K(X) = \Omega(Y) = \Omega_p(1)$, so also $K(Y) = \Omega_p(1)$ then $K(X) = \Omega_p(1)$. Hansen and 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 $m/n \to 0, H = c_0 n^{1/3}, c_0 > 0,$ and $\eta \in (0, 1)$

$$H^2/n \to |K''(0)|\Omega, \quad \text{if } \eta < 1/2,$$

$$K(X) = \frac{1}{0} \sum(u)du + c_0^2 |K''(0)|\Omega + \Omega_p(1), \quad \text{if } \eta = 1/2,$$

$$K(X) \to \frac{1}{0} \sum(v)du, \quad \text{if } \eta > 1/2.$$
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 $n \to \infty$, $H = c_0 n^{1/5}$ and $m^{-1} = \sigma(\sqrt{H/n}) = o(n^{-1/5})$ we have that

\[
\frac{1}{n^{1/5}} \left( K(X) - \int_0^1 \sigma^2(u) du \right) \xrightarrow{L} \text{MN} \left( c_0^{-2} k''(0) |\omega|^2, 4c_0 k_0 \int_0^1 \sigma^4(u) \frac{x_2(u)}{x_1(u)} du \right).
\]

The notation $\xrightarrow{L} \text{MN}$ 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 $\int_0^1 \sigma^4(u) \frac{x_2(u)}{x_1(u)} du$ is random, and stable convergence guarantees joint convergence that is needed here. Stable convergence is discussed, for example, in Mykland and 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 $c_0 = c_0^* n^{2/5}$ so $H = c_0^* n^{3/5} n^{5/3}$ where

\[
c_0^* = \left\{ \frac{k''(0)^2}{k_0^2} \right\}^{1/5}, \quad \xi^2 = \frac{\omega^2}{\sqrt{IQ}}, \quad \text{IQ} = \int_0^1 \sigma^4(u) \frac{x_2(u)}{x_1(u)} du.
\]

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

\[
c_0 k_0^* \text{IQ} = \kappa^2, \quad \frac{|k''(0)|}{c_0^*} \omega^2 = \kappa,
\]

where

\[
\kappa = k_0 \sqrt{\text{IQ} \omega^{2/5}}, \quad \kappa_0 = (|k''(0)| (k_0^*)^2)^{1/5}.
\]

Then

\[
\frac{1}{n^{1/5}} \left( K(X) - \int_0^1 \sigma^2(u) du \right) \xrightarrow{L} \text{MN}(\kappa, 4\kappa^2).
\]

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 $\text{Var}(D_{m,n}(u) | \mathcal{F}_{m-1,n}) \overset{d}{\to} x_2(t) - x_1(t)$, which is non-negative. Thus we have the inequality $x_2(t) - x_1(t) \geq x_1(t)$, which means that $\int_0^1 \sigma^4(u) \frac{x_2(u)}{x_1(u)} du \geq \int_0^1 \sigma^4(u) x_1(u) du$. 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 Barndoff-Nielsen et al. (2008) with $k'(0) = 0$. They had the nicer result that

\[
\frac{1}{n^{1/4}} \left( K^f(X) - \int_0^1 \sigma^2(u) du \right) \xrightarrow{L} \text{MN} \left( 0, 4c_0^4 k_0^4 \omega^4 + \frac{8}{c_0^4} k_0^2 \omega^2 \right),
\]

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 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 $K$ 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 $H \propto n^{1/3}$ 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 $U$ and a kernel function with $k'(0) \neq 0$ but otherwise satisfies $K$. Then, as $n, H, m \to \infty$ we have that $\frac{H}{n} K(U) \xrightarrow{D} 2|k'(0)| \int_{-\infty}^\infty \left( \sum_j u_j + y_0(u) \right) du$.

**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 $|k'(0)| (k_0^*)^2$ 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 minimises $|k'(0)| (k_0^*)^2$, 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.

---

4 See also Zhang (2006) who independently obtained a $n^{1/3}$ consistent estimator using a multiscale approach.
### 3.3. Multivariate asymptotic distribution

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

$$\text{IQ} = \int_{0}^{1} \left( \Sigma(u) \otimes \Sigma(u) \right) x_{2}(u) x_{1}(u) \, du,$$

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

**Theorem 3.** Suppose $H = c_{0} n^{3/4}, m^{-1} = o(n^{-1/2}), K, \text{SH}, D, \text{and U}$ then

$$n^{1/5} \left( K(X) - \int_{0}^{1} \Sigma(u) \, du \right) \overset{L}{\to} \text{MN}(c_{0}^{2} |k''(0)| \Omega, 4c_{0}k^{(0)}_{\bullet} |\text{IQ}|).$$

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 quantity.

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

$$n^{1/5} a^{t} \left( K(X) - \int_{0}^{1} \Sigma(u) \, du \right) b \overset{L}{\to} \text{MN}(c_{0}^{2} |k''(0)| a^{t} \Omega b, 4c_{0}k^{(0)}_{\bullet} |\text{IQ}| v_{ab}).$$

where $v_{ab} = \text{vec} \left( a^{t} \Sigma b \right)$. For two different elements, $a^{t} K(X) b$ and $c^{t} K(X) d$ say, their asymptotic covariance is given by $4c_{0}k^{(0)}_{\bullet} |\text{IQ}| v_{ab}$.

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

$$n^{1/5} \left( K(X^{(i)}) - \int_{0}^{1} \Sigma_{0} \, du \right) \overset{L}{\to} \text{MN}(A, B),$$

where

$$A = c_{0}^{2} |k''(0)| \left( \begin{array}{c} \Omega_{ii} \\ \Omega_{ij} \\ \Omega_{ji} \\ \Omega_{jj} \end{array} \right)$$

and

$$B = 2c_{0}k^{(0)}_{\bullet} \int_{0}^{1} \left( \begin{array}{cccc} 2 \Sigma_{ii} & 2 \Sigma_{ij} & 2 \Sigma_{ij} & 2 \Sigma_{jj} \\ 2 \Sigma_{ij} & 2 \Sigma_{ii} & 2 \Sigma_{ij} & 2 \Sigma_{jj} \end{array} \right) x_{2} \, du.$$
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 $\beta(1,2)$. For $\beta(1,2)$ the trade-off is between $c_0^{-1}|k(0)|^2(\Omega_{12} - \Omega_{22} \beta(1,2))^2$, and

$$2c_0\beta_{1,2} \int_0^1 (\Sigma_{11} \Sigma_{22} + \Sigma_{12}^2 - 4\beta(1,2) \Sigma_{11} \Sigma_{22} + 2\beta(1,2)^2 \Sigma_{22})d\nu.$$ 

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

$$dY^{(i)} = \mu^{(i)}dt + dV^{(i)} + dB^{(i)}, \quad dV^{(i)} = \rho^{(i)}\sigma^{(i)}dB^{(i)},$$

$$dF^{(i)} = \sqrt{1 - (\rho^{(i)})^2} \sigma^{(i)}dW$$

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

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 $\sigma^{(i)} = \exp(\beta_0^{(i)} + \beta_1^{(i)} \omega^{(i)})$ with $d\omega^{(i)} = a^{(i)} \omega^{(i)}dt + dW^{(i)}$. Thus there is perfect statistical leverage (correlation between their innovations) between $Y^{(i)}$ and $\sigma^{(i)}$, while the leverage between $Y^{(i)}$ and $\sigma^{(i)}$ is $\rho^{(i)}$. The correlation between $Y^{(1)}(t)$ and $Y^{(2)}(t)$ is $\sqrt{1 - (\rho^{(1)})^2}$.

5. Empirical illustration

We analyse high-frequency assets prices for thirty assets. In the analysis the main focus will be on the empirical properties of $30 \times 30$ realised covariance matrices and compare it to the standard realised covariance. In the no noise case of $\xi = 0$ 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 $\xi$ is quite large, but it is small compared to the estimator’s variance. In that situation the realised kernel is far more precise than the realised variance. None of these results is 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.

5.1. 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 and Lunde (2006), Brownless and Gallow (2006) and Barndorff-Nielsen et al. (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.–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) Delete rows for which the spread is more than 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 centred median (excluding the observation under consideration) of 50 observations.

Trade data only. (T1) Delete entries with corrected trades. (Trades with a Correction Indicator, CORR ≠ 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 timestamp: 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 and 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 calendar time sampling. Specifically, we consider 15 s, 5 min, and 30 min 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 stocks 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 realised 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 × 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 than half the observations for most frequently traded assets. For the least active assets we typically lose between 30% and 40% of the observations.

For the 10 × 1 case the data loss is more pronounced. Still, on average more that 25% of the observations remain in the sample. For transaction data the average number of Refresh Time observations is 1470, whereas the corresponding number is 4491 for the quote data. So in most cases we have an observation on average more often than every 5 s for quote data and 15 s for trade data. We observed that the data loss levels off as the dimension increases. For the 30 × 1 case we have on average more that 17% of the observations remaining in the sample. For transaction data the average number of Refresh Time observations is 966 and 2978 for the quote data. This gives an observation on average more often than every 8 s for quote data and 24 s for trade data.

5.3. Analysis of the covariance estimators: Cov, CovHV, CovOtoC and CovAmo.

Throughout this subsection the target which we wish to estimate is \( \{Y^{(1)}, Y^{(10)}\} \), \( j = 1, 2, \ldots, d, \in \mathbb{N} \). 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 \( \{Y^{(1)}, Y^{(10)}\} \), denoted by \( \text{Cov}_{s}^{K_{30 \times 30}}, \text{Cov}_{s}^{K_{30 \times 10}} \) and \( \text{Cov}_{s}^{K_{10 \times 10}} \), 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: \( \text{Cov}_{s}^{HV} \), the Hayashi and Yoshida (2005) covariance estimator. \( \text{Cov}_{s}^{OtoC} \), the realised covariance based on intraday returns that span a interval of length \( \Delta \), e.g. 5 or 30 min (the previous-tick method is used). \( \text{Cov}_{s}^{Amo} \), 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 and Yoshida (2005) estimator \( \text{Cov}_{s}^{HV} \) and the open to close estimator \( \text{Cov}_{s}^{OtoC} \). 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 \( \text{Cov}_{s}^{HV} \) estimates are typically much lower than the corresponding \( \text{Cov}_{s}^{OtoC} \) number. Bold font indicate estimates that are significantly different from \( \text{Cov}_{s}^{OtoC} \) at the one percent level. This assessment is carried out in the following way.

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 \( s_{\bar{e}}^{2} = \gamma_{h} + 2 \sum_{h=1}^{H} (1 - \frac{h}{P}) \gamma_{h} \), where \( \gamma_{h} = \frac{1}{\sum_{i=1}^{d} n_{i} n_{j-i}} \). Here \( n_{j} \approx \bar{e} - \hat{\bar{e}} \) and \( q = \text{int}(4(T/100)^{2}) \).


<table>
<thead>
<tr>
<th>Trades</th>
<th>Quotes</th>
<th>8460</th>
<th>9270</th>
<th>8626</th>
<th>8553</th>
<th>10291</th>
<th>10809</th>
<th>8026</th>
<th>10254</th>
<th>8521</th>
<th>15973</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trades</td>
<td>Quotes</td>
<td>3442</td>
<td>4228</td>
<td>3461</td>
<td>3529</td>
<td>4544</td>
<td>5480</td>
<td>3330</td>
<td>4845</td>
<td>3307</td>
<td>5412</td>
</tr>
</tbody>
</table>

Average over daily number of high frequency observations available before the refresh time transformation

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 \( p = \sum_{t=1}^{T} \frac{1}{\sum_{i=1}^{d} n_{i} n_{j-i}} \). The upper panel display this in the 2 × 2 case. The upper panel is based on transaction prices, whereas the lower panel is based on mid-quotes. In the lower panel we average over the daily number of high frequency observations.

We now move on to more successful estimators. The upper panel of Table 5 presents the time series average estimates for \( \text{Cov}_{s}^{K_{30 \times 30}}, \text{Cov}_{s}^{K_{10 \times 10}} \), and the lower panel gives 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% level. These results are quite encouraging for all three estimators. The average levels of the three estimators are roughly the same.

A much tighter comparison is to replace the noisy \( e_{s} = \text{Cov}_{s}^{K_{2 \times 2}} - \text{Cov}_{s}^{OtoC} \), with \( e_{s} = \text{Cov}_{s}^{K_{2 \times 2}} - \text{Cov}_{s}^{K_{2 \times 2}} \), where the two estimates come from applying the realised 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 \( e_{s} = \text{Cov}_{s}^{K_{2 \times 2}} - \text{Cov}_{s}^{K_{2 \times 2}} \), we find that the realised kernels and the realised covariances based on 5 min 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 \( \text{Cov}_{s}^{HV} \) 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.
are boldfaced if the bias is significant at the 1% level. The upper panel presents average estimates for \( \text{Cov}^{(\text{H})} \), and the lower panel displays these for \( \text{Cov}^{(\text{NC})} \). 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% level.

### Table 5

<table>
<thead>
<tr>
<th>AA</th>
<th>AIG</th>
<th>AXE</th>
<th>BA</th>
<th>BAC</th>
<th>C</th>
<th>CAT</th>
<th>CVX</th>
<th>DD</th>
<th>SPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.338</td>
<td>0.945</td>
<td>1.033</td>
<td>0.808</td>
<td>0.865</td>
<td>1.118</td>
<td>1.055</td>
<td>0.818</td>
<td>1.047</td>
<td>0.862</td>
</tr>
<tr>
<td>0.933</td>
<td>2.725</td>
<td>1.300</td>
<td>0.722</td>
<td>1.254</td>
<td>1.556</td>
<td>0.856</td>
<td>0.533</td>
<td>0.820</td>
<td>0.878</td>
</tr>
<tr>
<td>1.003</td>
<td>1.273</td>
<td>2.580</td>
<td>0.780</td>
<td>1.326</td>
<td>1.656</td>
<td>0.974</td>
<td>0.578</td>
<td>0.900</td>
<td>0.930</td>
</tr>
<tr>
<td>0.792</td>
<td>0.730</td>
<td>0.773</td>
<td>2.194</td>
<td>0.654</td>
<td>0.831</td>
<td>0.765</td>
<td>0.488</td>
<td>0.702</td>
<td>0.670</td>
</tr>
<tr>
<td>0.850</td>
<td>1.226</td>
<td>1.305</td>
<td>0.658</td>
<td>2.057</td>
<td>1.681</td>
<td>0.805</td>
<td>0.485</td>
<td>0.767</td>
<td>0.811</td>
</tr>
<tr>
<td>1.102</td>
<td>1.528</td>
<td>1.627</td>
<td>0.840</td>
<td>1.649</td>
<td>2.942</td>
<td>1.011</td>
<td>0.638</td>
<td>0.986</td>
<td>1.057</td>
</tr>
<tr>
<td>0.258</td>
<td>0.841</td>
<td>0.939</td>
<td>0.760</td>
<td>0.792</td>
<td>1.001</td>
<td>2.225</td>
<td>0.603</td>
<td>0.776</td>
<td></td>
</tr>
<tr>
<td>0.822</td>
<td>0.534</td>
<td>0.570</td>
<td>0.500</td>
<td>0.489</td>
<td>0.640</td>
<td>0.607</td>
<td>0.655</td>
<td>0.576</td>
<td>0.589</td>
</tr>
<tr>
<td>1.049</td>
<td>0.813</td>
<td>0.890</td>
<td>0.706</td>
<td>0.759</td>
<td>0.984</td>
<td>0.874</td>
<td>0.589</td>
<td>1.810</td>
<td>0.736</td>
</tr>
<tr>
<td>0.962</td>
<td>0.876</td>
<td>0.917</td>
<td>0.678</td>
<td>0.808</td>
<td>1.058</td>
<td>0.775</td>
<td>0.602</td>
<td>0.744</td>
<td>0.745</td>
</tr>
<tr>
<td>3.400</td>
<td>0.953</td>
<td>1.024</td>
<td>0.808</td>
<td>0.863</td>
<td>1.121</td>
<td>1.040</td>
<td>0.825</td>
<td>1.046</td>
<td>0.868</td>
</tr>
<tr>
<td>0.931</td>
<td>2.766</td>
<td>1.298</td>
<td>0.734</td>
<td>1.248</td>
<td>1.553</td>
<td>0.860</td>
<td>0.544</td>
<td>0.829</td>
<td>0.885</td>
</tr>
<tr>
<td>0.995</td>
<td>1.271</td>
<td>2.593</td>
<td>0.778</td>
<td>1.315</td>
<td>1.642</td>
<td>0.957</td>
<td>0.581</td>
<td>0.900</td>
<td>0.926</td>
</tr>
<tr>
<td>0.798</td>
<td>0.739</td>
<td>0.783</td>
<td>2.222</td>
<td>0.656</td>
<td>0.845</td>
<td>0.722</td>
<td>0.501</td>
<td>0.712</td>
<td>0.679</td>
</tr>
<tr>
<td>0.844</td>
<td>1.204</td>
<td>1.290</td>
<td>0.660</td>
<td>2.080</td>
<td>1.669</td>
<td>0.799</td>
<td>0.493</td>
<td>0.765</td>
<td>0.811</td>
</tr>
<tr>
<td>1.101</td>
<td>1.518</td>
<td>1.616</td>
<td>0.854</td>
<td>1.625</td>
<td>2.985</td>
<td>1.013</td>
<td>0.653</td>
<td>0.992</td>
<td>1.062</td>
</tr>
<tr>
<td>1.019</td>
<td>0.843</td>
<td>0.931</td>
<td>0.767</td>
<td>0.785</td>
<td>1.003</td>
<td>2.234</td>
<td>0.613</td>
<td>0.879</td>
<td>0.779</td>
</tr>
<tr>
<td>0.684</td>
<td>0.548</td>
<td>0.583</td>
<td>0.512</td>
<td>0.501</td>
<td>0.656</td>
<td>0.619</td>
<td>1.682</td>
<td>0.592</td>
<td>0.603</td>
</tr>
<tr>
<td>1.045</td>
<td>0.819</td>
<td>0.892</td>
<td>0.719</td>
<td>0.755</td>
<td>0.991</td>
<td>0.876</td>
<td>0.603</td>
<td>1.850</td>
<td>0.745</td>
</tr>
<tr>
<td>0.965</td>
<td>0.876</td>
<td>0.917</td>
<td>0.688</td>
<td>0.806</td>
<td>1.061</td>
<td>0.779</td>
<td>0.616</td>
<td>0.752</td>
<td>0.755</td>
</tr>
</tbody>
</table>

The upper panel presents average estimates for \( \text{Cov}^{(\text{H})} \), the middle panel for \( \text{Cov}^{(\text{NC})} \), and the lower panel gives results for \( \text{Cov}^{(\text{NC})} \). In both panels the upper diagonal is based on transaction prices, whereas the lower diagonal is based on mid-quotes. The diagonal element are computed with transaction prices. Outside the diagonals numbers are boldfaced if the bias is significant at the 1% level.
so seven underestimates the efficiency gain of using $\text{Cov}_{t-3}^{K_{2t+2}}$. If volatility is close to being persistent then $\text{Cov}_{t}^{K_{10t+30}}$ is at least 16.1 times more informative than the cross product of daily returns. The same observation holds for mid-quotes. $\text{Cov}_{t}^{K_{5t}}$ and $\text{Cov}_{t}^{K_{10t}}$ are very precise estimates of the wrong quantity. $\text{Cov}_{t}^{K_{10t}}$ is quite close to $\text{Cov}_{t}^{K_{10t+30}}$, $\text{Cov}_{t}^{K_{10t+10}}$, and $\text{Cov}_{t}^{K_{2t+2}}$, with $\text{Cov}_{t}^{K_{5t}}$ and $\text{Cov}_{t}^{K_{10t+30}}$ having a correlation of 0.942. We note that realised kernel results seem to show some bias compared to $\text{Cov}_{t}^{\text{Basic}}$, the difference is however statistically insignificantly different than zero, as $\text{Cov}_{t}^{\text{Basic}}$ 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 the resulting estimates to be biased. The most persistent estimator is $\text{Corr}_{t}^{K_{10t}}$ 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 $\text{Corr}_{t}^{K_{2t+2}}$, which suggest that this is most effective estimate of the correlation.

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

In this subsection we will focus on estimating \( \hat{\rho}^{(i,j)} = [Y^{(i)}, Y^{(j)}] \), by the realised kernel correlation \( \hat{\rho}^{(i,j)K} = K^{(i,j)}_{4,4}/K^{(i,j)}_{4,4} \) and the corresponding realised correlation \( \hat{\rho}^{(i,j)K}_{4,4} \). A table in our web appendix reports the average estimates for \( \hat{\rho}^{(i,j)K}_{4,4} \) and \( \hat{\rho}^{(i,j)K}_{4,4} \). It shows the expected result that \( \hat{\rho}^{(i,j)K}_{4,4} \) is more precise than \( \hat{\rho}^{(i,j)K}_{4,4} \). 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 \( \hat{\rho}^{(i,j)K}_{4,4} \) are measured with noise and so when we form \( \hat{\rho}^{(i,j)K}_{4,4} \) it will be downward biased.

5.6. Analysis of the beta estimates

Here we will focus on estimating \( \hat{\beta}^{(i,j)} = [Y^{(i)}] \), by the realised kernel beta \( \hat{\beta}^{(i,j)K} = K^{(i,j)}_{4,4}/K^{(i,j)}_{4,4} \). Fig. 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 \( \hat{\beta}^{(i,j)K} \) to \( \hat{\beta}^{(i,j)K}_{4,4} \). The results are not very different in these two cases.

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

\[
\hat{\beta}^{K} = 1.20 + 0.923 \hat{\beta}^{K}_{4,4} + u_{t} - 0.726 u_{t-1}, \quad \text{adj}-R^{2} = 0.213,
\]

and for 5 min based realised betas

\[
\hat{\beta}^{K}_{4,4} = 1.16 + 0.950 \hat{\beta}^{K,5}_{4,4} + u_{t} - 0.821 u_{t-1}, \quad \text{adj}-R^{2} = 0.145.
\]

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.

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

\[
\hat{\beta}^{K}_{4,4} = 0.084 + 0.858 \hat{\beta}^{K}_{4,4} + 0.074 \hat{\beta}^{K,5}_{4,4} + u_{t} - 0.726 u_{t-1}, \quad \text{adj}-R^{2} = 0.215,
\]

with the corresponding 5 min based realised betas

\[
\hat{\beta}^{K,5}_{4,4} = 0.056 + 0.879 \hat{\beta}^{K,5}_{4,4} + 0.069 \hat{\beta}^{K,5}_{4,4} + u_{t} - 0.822 u_{t-1}, \quad \text{adj}-R^{2} = 0.150.
\]

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, 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 and Gallo (2006). Here we follow this route, fitting multivariate GARCH models with \( \hat{E}(r_t|\mathcal{F}_{t-1}) = 0 \), \( \text{Cov}(r_t|\mathcal{F}_{t-1}) = H_t \), where \( r_t \) is the \( d \times 1 \) vector of daily close to close returns, \( \mathcal{F}_{t-1} \) is the information available at time \( s < t \) to predict \( r_t \). A standard Gaussian quasi-likelihood \(-\frac{1}{2} \sum_{t=1}^{T} (\log |H_t| + r_t'H_t^{-1}r_t)\) is used to make inference. The model we fit is a variant on the scalar BEKK (e.g. Engle and Kroner (1995))

\[
H_t = C'C + \beta H_{t-1} + \alpha r_{t-1}'r_{t-1}' + \gamma K_{t-1}, \quad \alpha, \beta, \gamma \geq 0.
\]

Here we follow the literature and use \( H_t \) 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 and Mezrich (1996). The general idea is to estimate the intercept matrix by an auxiliary estimator that is given by

\[
\hat{C}'\hat{C} = \hat{S} \odot (1 - \alpha - \beta - \gamma \kappa), \quad \hat{S} = \frac{1}{T} \sum_{t=1}^{T} r_t'r_t,
\]

Fig. 2. Scatter plots for daily realised kernel betas for the AA and SPY asset combination.
where $\odot$ denotes the Hadamard product. There is a slight deviation from the situation considered by Engle and Mezrich (1996) because $K_{s-1}$ is only estimated for the part of the day where the NYSE is open. To accommodate this we follow Shephard and Sheppard (2010) that introduce the scaling matrix $K$, in (7) which we estimate by

$$
\hat{K}_t = \left( \frac{\bar{\mu}_K}{\mu} \right)_t, \quad \bar{\mu} = T^{-1} \sum_{s=1}^T r_{s-1} r'_s \quad \text{and} \quad \hat{\mu}_K = T^{-1} \sum_{s=1}^T K_s.
$$

Having $\tilde{S}$ and $\hat{K}$ at hand the remaining parameters are simply estimated by maximising the concentrated quasi-log-likelihood, with

$$
H_t = \tilde{S} \odot (1 - \alpha - \beta - \gamma \hat{K}) + \beta H_{t-1} + \alpha r_{t-1} r'_{t-1} + \gamma K_{t-1},
$$

$\alpha, \beta, \gamma \geq 0$.

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 $RV_{s-1}^m$.

We estimate scalar BEKK models for the $30 \times 30$, $10 \times 10$, and the $45 \times 2$ 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 $K_{s-1}$ reduces the size of the estimated $H_{s-1}$ term. It is also interesting to note that including $K_{s-1}$ in general gives a higher log-likelihood than including $RV_{s-1}^m$. 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 $RV_{s-1}^m$ from our most general specification is 0.66, where as the corresponding average for removing $K_{s-1}$ is 11.9. These tests can be interpreted as encompassing tests, and provide an overwhelming evidence that the information in $RV_{s-1}^m$ is contained in $K_{s-1}$.

---

**Table 7**

<table>
<thead>
<tr>
<th>Panel A: $30 \times 30$ case</th>
<th>Panel B: $10 \times 10$ case</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{s-1}$ $r_{s-1}(r_{s-1})'$ $K_{s-1}$ $RV_{s-1}^m$ $\log L$ $H_{s-1}$ $r_{s-1}(r_{s-1})'$ $K_{s-1}$ $RV_{s-1}^m$ $\log L$</td>
<td></td>
</tr>
<tr>
<td>0.943 (0.006) 0.005 (0.004) 0.040 (0.034) – – 27,029.9 0.768 (0.007) 0.015 (0.003) 0.151 (0.011) – – 7920.7</td>
<td></td>
</tr>
<tr>
<td>0.742 (0.014) 0.013 (0.001) – 0.115 (0.006) – 27,077.7 0.687 (0.022) – – 1.60 (0.013) – – 7935.9</td>
<td></td>
</tr>
<tr>
<td>0.984 (0.001) 0.008 (0.001) – – 28,477.5 0.965 (0.023) – – – – 8307.5</td>
<td></td>
</tr>
<tr>
<td>0.777 – 0.076 (0.004) 0.061 (0.003) – 26,948.3 0.705 – – 0.126 (0.014) 0.067 – 7923.3</td>
<td></td>
</tr>
<tr>
<td>0.784 (0.013) 0.009 (0.004) 0.067 (0.003) 0.059 (0.003) – 26,904.3 0.716 (0.023) 0.017 (0.003) 0.106 (0.014) 0.065 – 7903.0</td>
<td></td>
</tr>
</tbody>
</table>

Panel C: $2 \times 2$ cases

<table>
<thead>
<tr>
<th>AIG-CAT</th>
<th>BA-SPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{s-1}$ $r_{s-1}(r_{s-1})'$ $K_{s-1}$ $RV_{s-1}^m$ $\log L$ $H_{s-1}$ $r_{s-1}(r_{s-1})'$ $K_{s-1}$ $RV_{s-1}^m$ $\log L$</td>
<td></td>
</tr>
<tr>
<td>0.837 (0.025) 0.038 (0.030) 0.126 (0.037) – – 2584.9 0.844 (0.031) 0.031 (0.031) 0.094 – – 1516.9</td>
<td></td>
</tr>
<tr>
<td>0.863 (0.023) 0.044 (0.030) – 0.098 (0.025) – 2591.2 0.843 (0.032) – – 0.091 (0.030) – – 1517.8</td>
<td></td>
</tr>
<tr>
<td>0.951 (0.006) 0.045 (0.005) – – 2629.7 0.958 (0.036) 0.036 (0.005) – – – – 1544.4</td>
<td></td>
</tr>
<tr>
<td>0.764 – 0.236 (0.003) 0 – 2592.4 0.717 – – 0.125 (0.061) 0.083 – 1521.1</td>
<td></td>
</tr>
<tr>
<td>0.837 (0.028) 0.038 (0.028) 0.126 (0.020) 0 – 2584.9 0.837 (0.031) 0.031 (0.009) 0.068 (0.047) 0.031 – 1516.6</td>
<td></td>
</tr>
</tbody>
</table>

Estimation results for scalar BEKK models for close-to-close d = 30, 10, 2 dimensional return vectors.
6. Additional remarks

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

In the univariate case the realised kernel $K(X) = \sum_{h=-\infty}^{\infty} k(h)$

$$I_h^f = \sum_{j=1}^{n} x_j \gamma_{j-h},$$

is used. We are grateful to Kevin Sheppard for pointing out these negative days.

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 may 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 in the bias-variance trade-off and reduces the best rate of convergence from $n^{1/4}$ to $n^{1/3}$. This resembles the effects seen in the literature on density estimation with kernel functions. The property, \( \int u^2 k(u) du = 0 \), reduces the order of the asymptotic bias, but kernel functions that satisfy \( \int u^2 k(u) du = 0 \) 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. The most obvious is the use of a kernel function that does not satisfy, \( \int_{-\infty}^{\infty} k(x) \exp(i\lambda x) dx \geq 0 \) 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 $y_1$ and $y_2$, which can mean $K^f(X)$ may be negative. This can be verified by rewriting the estimator as a quadratic form estimator, \( X'Mx \), where $M$ is a symmetric band matrix \( M = \text{band}(1,1,k(\frac{1}{h}), k(\frac{2}{h}), \ldots) \). The determinant of the upper-left matrix is given by \( -k(\frac{1}{h}) - 1)^2 \), so that $k(h) \neq 0$ is needed to avoid negative eigenvalues. Repeating this argument leads to $k(\frac{h}{2}) = 1$ for all $h$, which violates the condition that $k(\frac{h}{2}) \to 0$, as $h \to \infty$. Finally, the third reason that the flat-top kernel could produce a negative estimate was due to the construction of realised autocovariances, $\gamma_h = \sum_{j=1}^{n} x_j x_{j-h}$. This requires the use of “out-of-period” intraday returns, such as $x_1 - x_H$. This formulation was chosen because it makes $E(K(U)) = 0$ 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 $x_1 - x_H$ (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 $E[n^{1/4}(K(X) - [Y])^2/\omega]$ and $E[n^{1/4}(K^f(X) - [Y])^2/\omega]$, the mean square normalised by the rate of convergence of $K^f(X)$ (which is the flat-top realised kernel using the Parzen weight function). An implication is that the scaled MSE for the $K(X)$ and $K^f(X)$ 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 ($K'_B(X)$) and Parzen ($K'_P(X)$) 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 square bias. For small $n$ there is not much difference between the three estimators, but by the time $n = 4096$ (which is realistic for our applications) the flat-top $K^f(X)$ has roughly half the MSE of $K(X)$ in the univariate case. Hence in ideal (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 $U$ is modelled as a dependent process. We consider the moving average specification, $U_t = \epsilon_t - \theta_1 \epsilon_{t-1}$, with $\theta = \pm 0.5$ and the autoregressive specification, $U_t = \psi U_{t-1} + \epsilon_t$, with $\psi = \pm 0.5$, where $\epsilon_t$ 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 $H_{16}^f = 2.280 n^{1/3}$, $H_5^f = 4.77\text{ton}^{1/3}$, and $H_5^p = 3.51 n^{1/3} n^{1/3}$ where $\omega^2 = \sum_{h=1}^{\infty} \text{cov}(U_t, U_{t-h})$. The results show the robustness of $K(X)$ and the strong asymptotic bias of $K^f$ and $K^p$ under the non-white noise assumption. The specifications, $\theta = 0.5$ and $\psi = -0.5$ induce a negative first-order autocorrelation while $\theta = -0.5$ and $\psi = 0.5$ induce positive autocorrelation. Negative first-order autocorrelation can be the product of bid-ask bound 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.

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 fresh time would look like. We define it as $\tilde{V} = \sum_{j=1}^{n-h} x_j x_{j-h}$, where $x_j = (\psi_j H)^{-1/2} \sum_{h=1}^{\infty} \zeta(h) x_{j+h}$, $\psi_2 = \int_{0}^{1} \rho^2(u) du$. Here $g(u), u \in [0,1]$ is a non-negative, continuously differentiable weight function, with the properties that $g(0) = g(1) = 1$ and $\psi_2 > 0$. Now if we set $H = \eta n^{1/5}$, then the univariate result in Vetter (2008) would suggest that $\tilde{V}$ converges at rate $n^{-1/4}$, like the univariate version of our multivariate realised kernel. There is no simple guidance, even in the univariate case, as to how to choose $\theta$. 

---

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.
In the univariate bias corrected form, Jacod et al. (2009) show that $\tilde{V}$ is asymptotically equivalent to using a $K^2(X)$ with $k(x) = \psi^{-1}_0 \int g(u)g(u-x)du$ and $H \propto n^{1/2}$. It is clear the same result will hold for the relationship between $\tilde{V}$ and $K(X)$ in the multivariate case when $H = \theta n^{1/2}$. A natural choice of $g(x) = (1-x) \times \chi$, which delivers $\int g(u)^2(u)du = 1/12$ 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 $\tilde{V}$ when applied to refresh time based high frequency returns.

Following the initial draft of this paper Christensen et al. (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 is not guaranteed to be positive semi-definite.

6.2.2. Estimating integrated quantility

In order to construct feasible confidence intervals for our realised quantities (see Barndorff-Nielsen and 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 and Shephard (2004) bipower type estimator applied to suboptimal preaveraged data taking $H = \theta n^{1/3}$. This is not an optimal estimator, it will converge at rate $n^{1/5}$, but it will be positive semi-definite. The proposed (positive semi-definite) estimator of vec(IQ) is $\hat{Q} = n \sum_{j=1}^{n-H-1} [c_j c_j']$, where $c_j = \text{vec}(\bar{x}_j \bar{x}_j')$. That the elements of $\hat{Q}$ are 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 $n^{1/5} \left\{ \text{log}(K(X)) - \log(\int_0^1 \sigma^2(u)du) \right\}$ to improve its finite sample performance. When the data is regularly spaced and the volatility is constant then $\kappa \sigma^2 = (\omega/\theta)^{5/2} |k'(0)|^{1/2} (\sigma_{\bar{X}}^2)^{5/2}$, 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 $U_m U_{m} + U_{m} U_{m}$, which is given by $4m^{-2} \omega^2 + 4m^{-2} \omega^2 = 8 \omega^2 m^{-2}$, see the proof of the Proposition A.2. The asymptotic variance (abstracting from end-effects) is $5k^2 n^{-2/5} = 5k^2 \sigma^2 \omega^2 / 2 \int k'(0)^{5/2} (\sigma_{\bar{X}}^2)^{5/2}$, which lies between contributions from end-effects and asymptotic variance is given by $8 \omega^2 \omega_{\text{IQ}}(m) = m^{-2} \omega^2 + 5 \int k''(0) \omega^2 / 5 (\sigma_{\bar{X}}^2)^{5/2} / (N - m)^{-2/5}$. This function is plotted in Fig. 4 for the case where $N = 1000$ and 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 Fig. 4, where the optimal value of $m$ has increased a little from Fig. 4 as $n$ has increased to 5000. However, the optimal amount of jittering is still rather modest.
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 BSM. The analysis could be extended to the situation where $Y$ is a pure 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 utilise high frequency data to significantly improve our predictive models as well as providing a better understanding of asset pricing and management of risk in financial markets.

In the appendices that follow, we give some proofs and the errors induced by stale prices. 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 synchronised data. This is shown in Appendix A, where Propositions A.1–A.5, Theorems 3 and A.4 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 $i, j$,

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

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

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

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 is to prove the following proposition.

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

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

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

$$V_h = \sum_{j=h+1}^{n-1} U_j U_{j-h} + U_{j-h} U_j', $$

and

$$Z_h = (U_0 U_h' + U_h U_0') + (U_h U_{n-h} + U_{n-h} U_h'). $$

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

$$\Gamma_0(U) = V_0 - V_1 + \frac{1}{2} Z_0 - Z_1 $$

(A.1)

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

so with $k_h = k(h^2)$ we have

$$K(U) = (k_0 - k_1)V_0 - \sum_{h=1}^{n-1} (k_{h+1} - 2k_h + k_{h-1})V_h $$

$$+ \frac{1}{2} Z_0 - \sum_{h=1}^{n-1} (k_h - k_{h-1}) Z_h. $$

(A.3)

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

$$\Gamma_0(U) = \sum_{j=1}^{n} (U_j - U_{j-1}) (U_j - U_{j-1})' $$

$$= U_0 U_0' + U_n U_n' + \sum_{j=1}^{n-1} (U_j U_j' + U_j U_j') $$

$$- \sum_{j=2}^{n-1} (U_{j-1} U_j' + U_{j-1} U_j'), $$

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 $V_h$ for $h = 0, 1, \ldots$.

Proposition A.2. Given $U$ then

$$\frac{1}{n} V_h \xrightarrow{p} \begin{cases} \int_0^1 \{ \Sigma_h(u) + \gamma_0(u) \} du & \text{for } h = 0, \\ \int_0^1 \{ \gamma_h(u) + \gamma_0(u) \} du & \text{for } h > 0, \end{cases} $$

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

$$m Z_h \xrightarrow{p} \begin{cases} 2 [\Sigma_0(0) + \Sigma_0(1)] & \text{for } h = 0, \\ \sum_{j=0}^{\infty} \{ \gamma_{j+h}(0) + \gamma_{j+h}(0)' + \gamma_{j+h}(1) + \gamma_{j+h}(1)' \} & \text{for } h > 0. \end{cases} $$

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

Proof of Proposition A.2. The first result follows by the definition of $V_0$ and $U$. Next, since $U_0 = m^{-1} \sum_{j=0}^{m-1} T(t_j)$ it follows that $Z_0$ is
stochastic for any $m < \infty$, and

$$m U_0' U_0' = m^{-1} \sum_{j=0}^{m-1} U(t_j) U(t_j)' \xrightarrow{p} \Sigma_0(0), $$

and similar $m U_h' U_h' \sim \Sigma_h(1)$. So the result for $h = 0$ follows from

$$Z_0 = 2(U_0' U_0 + U_n' U_n), $$

Next, for $h > 0$,

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

$$= \sum_{j=0}^{\infty} \gamma_{j+h}(0)' $$

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

Proof of Theorem A.4. Since $k'(0) = 0$ and $k''(\epsilon)$ 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

$$(k_0 - k_1)V_0 - \sum_{h=1}^{n-1} (k_{h+1} - 2k_h + k_{h-1}) V_h $$

$$= H^{-2} \sum_{h=n+1}^{n} a_h \sum_{j} U_j U_{j-h} $$

$$= H^{-2} \sum_{h>\sqrt{n}} a_h \sum_{j} U_j U_{j-h} + H^{-2} \sum_{h<\sqrt{n}} a_h \sum_{j} U_j U_{j-h}. $$

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

$$\sup_{h \leq \sqrt{n}} |H^2 a_h n + k''(0)| \to 0, $$

as $H, n \to \infty$ with $H/n = o(1)$, so that the first term $\frac{H^2}{n} \sum_{h \leq \sqrt{n}} a_h n \sum_{j} U_j U_{j-h} = -k''(0) \frac{n}{H^2} \Omega + o(\frac{n}{H^2})$. The second term vanishes because

$$\frac{n}{H^2} \left| \sum_{h \leq \sqrt{n}} a_h n \sum_{j} U_j U_{j-h} \right| $$

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

and $\sup_{h \leq \sqrt{n}} \frac{n}{H^2} \sum_{j} U_j U_{j-h} = O_p(1)$.

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

$$\sum_{h=1}^{n-1} (k_h - k_{h-1}) Z_h = \frac{1}{m H} \sum_{h=1}^{n-1} \{ K(h/H) + o(1) \} m Z_h $$

$$= O_p(m^{-1}). $$

Proof of Lemma 2. When $k'(0) \neq 0$ we see that the first term of
(A.3) is such that $\frac{H}{n} (k_0 - k_1)V_0 \xrightarrow{p} -k'(0)2 \int_0^1 \{ \Sigma_h(u) + \gamma_0(u) \} du$. From the proof of Theorem A.4 it follows that the other terms in
(A.3) are of lower order. \Box
A.2. Results concerning $K(Y)$

The aim of this subsection is to 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

$$
\sqrt{\frac{H}{n}} \left( K(Y) - \int_0^1 \sigma^2(u) \, du \right) 
\rightarrow \text{MN} \left( 0, 4\rho_0^2 \int_0^1 \sigma^4(u) \frac{\xi_2(u)}{\xi_1(u)} \, du \right).
$$

(A.4)

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

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

where $\eta_{N,i} = Y(\tau_{N,i}) - Y(\tau_{N,i-1})$ and $\hat{\eta}_{N,i} = \sigma(\tau_{N,i-1})(W_{N,i} - W_{N,i-1})$ and

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

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

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

$$
y_{N,i} = \int_{\tau_{N,i-1}}^{\tau_{N,i}} \mu(u) \, du + \int_{\tau_{N,i-1}}^{\tau_{N,i}} \sigma(u) \, dW(u) = \hat{y}_{N,i} + \int_{\tau_{N,i-1}}^{\tau_{N,i}} \hat{\sigma}(u) \, d\hat{W}(u) \quad \text{(A.5)}
$$

Jacobian (unpublished paper, (6.25)) and Phillips and Yu (unpublished paper, Eq. 66). Let $\hat{\epsilon}_{N,i} = \sqrt{\hat{\lambda}_{N,i}/(W_{N,i} - W_{N,i-1})}$ so that $\hat{\epsilon}_{N,i} \sim \text{iid } N(0, 1)$ and note that $\hat{\eta}_{N,i} = N^{-1/2} \sigma(\tau_{N,i-1})D_{N,i}^2 \epsilon_{N,i}$. We use $\hat{y}_{N,i}$ as our estimate of $y_{N,i}$ throughout, later showing no impact on the result.

Note that $y_{N,i} - \hat{y}_{N,i} = \int_{\tau_{N,i-1}}^{\tau_{N,i}} \sigma(u) - \sigma(\tau_{N,i-1}) \, dW(u)$, so with $d[\sigma]_i = \lambda_i \, dt$ we find

$$
\hat{\sigma}_{N,i-1} = \int_{\tau_{N,i-1}}^{\tau_{N,i}} (\sigma(u) - \sigma(\tau_{N,i-1})) \, dW(u) = \frac{\lambda^2}{2} \int_{\tau_{N,i-1}}^{\tau_{N,i}} (u - \tau_{N,i-1}) \frac{dW(u)}{\sqrt{\lambda_i}} \quad \text{and} \quad \hat{\lambda}_{N,i-1} = \frac{\hat{\lambda}_i^2}{\sqrt{N}}
$$

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

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

**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

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

since $\max_{i=1,...,m} D_{N,i} = o_p(m^{1/2})$, $\sigma^2(t)$ is bounded, and $\sum_{i=1}^m \epsilon_{N,i}^2 = o_p(m)$. So we need $\sqrt{\frac{n}{m^{3/2}}} = O(1)$ 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{H}{n}} \left[ K(Y) - \tilde{K}(Y) \right] = o_p(1).

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

$$
\frac{N}{2} \sum_{i=1}^N \hat{\eta}_{N,i} = N \sum_{i=1}^N y_{N,i} \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right)
$$

$$
- \sum_{i=1}^N \hat{y}_{N,i} \left( \sum_{h=0}^{N-1} k_h \hat{y}_{N,i-h} \right)
$$

$$
= N \left( \sum_{i=1}^N (y_{N,i} - \hat{y}_{N,i}) \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right) \right)
$$

$$
+ \sum_{i=1}^N \hat{y}_{N,i} \left( \sum_{h=0}^{N-1} k_h (y_{N,i-h} - \hat{y}_{N,i-h}) \right). \quad \text{(A.6)}
$$

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

$$
V = \sum_{i=1}^N \lambda_{N,i}^2 \frac{D_{N,i}^2}{N} \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right)^2
$$

$$
\leq \max_{i=1,...,N} \lambda_{N,i}^2 \frac{1}{N} \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right)^2
$$

$$
= o_p(N)O_p(1) \frac{1}{N} O_p \left( \frac{H}{N} \right) = o_p(\sqrt{H/N}),
$$

where we have used that $\sum_{h=0}^{N-1} k_h y_{N,i-h} = O_p(\sqrt{H/N})$. The second term is

$$
\sum_{i=1}^N \hat{y}_{N,i} \sum_{h=0}^{N-1} k_h (y_{N,i-h} - \hat{y}_{N,i-h})
$$

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

It has a zero conditional means and its conditional variance is

$$
\frac{1}{N} \sum_{i=1}^N \sigma^2(\tau_{N,i-1})D_{N,i} \sum_{h=0}^{N-1} k_h^2 \int_{\tau_{N,i-h}}^{\tau_{N,i}} \sigma(u) - \sigma(\tau_{N,i-h}) \, dW(u)
$$

$$
= \frac{1}{N} \sum_{i=1}^N \lambda_{N,i}^2 \frac{D_{N,i}^2}{N} \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right)^2 / 2 \left( 1 + o_p(1) \right),
$$

where $d[\sigma] = \lambda_i \, dt$

$$
\leq \frac{1}{N^3} \sum_{i=1}^N \lambda_{N,i}^2 \left( \sum_{h=0}^{N-1} k_h y_{N,i-h} \right)^2 \max_{i=1,...,N} D_{N,i} \frac{1}{2} \lambda_{N,i}.
$$
Suppose \( \varnothing \) variance, which is the sum of the martingale differences:

\[
\sum_{i=1}^{N} \sigma^2(\tau_{N,i-1}) \hat{o}_p(N^{1/2}) \hat{o}_p(1) \hat{o}_p(\hat{q}) \hat{o}_p \left( \frac{\log N}{q} \right) = \hat{o}_p \left( \frac{H}{N} \right), \quad \text{take } q = N^{1/2} / \log(N).
\]

Here we have used that \( \max_{1\leq i \leq m} D_{N,i} = o_p(N^{1/2}) \), that 
\[ \frac{1}{N} \sum_{i=1}^{N} k_i h \approx N \] is at most of order \( o(j^{-1}) \), since \[ \frac{1}{N} \sum_{i=1}^{N} k_i (\frac{1}{N})^2 \approx \sum_{i=1}^{N} k_i (\frac{1}{N})^2 \] is convergent, and that \[ \sum_{i=1}^{m} \frac{1}{j} = O(\log m). \]

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

\[
\sqrt{N} \left( \hat{N} - \int_0^1 \sigma^2(u) du \right) \xrightarrow{L} \text{MN} \left( 0, 4k_0 \int_0^1 \sigma^4(u) \frac{X_2(u)}{X_1(u)} du \right).
\]

**Proof.** We have \( \hat{N} = \sum_{i=1}^{N} \sigma^2(\tau_{N,i-1}) \Delta_{N,i} + \hat{\eta}_{N,i}^{(1)} + \hat{\eta}_{N,i}^{(2)} \). Phillips and Yu (unpublished paper) imply that \( \sqrt{N} \left( \sum_{i=1}^{N} \sigma^2(\tau_{N,i-1}) \Delta_{N,i} - \int_0^1 \sigma^2(u) du \right) = o_p \left( \frac{1}{\sqrt{N}} \right) \). This means that

\[
\sqrt{N} \left( \hat{N} - \int_0^1 \sigma^2(u) du \right) = \sqrt{N} \left( \sum_{i=1}^{N} \left( \hat{\eta}_{N,i}^{(1)} + \hat{\eta}_{N,i}^{(2)} \right) + o_p(1) \right),
\]

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

The first term, \( \hat{\eta}_{N,i}^{(1)} \), is the sampling error of the well-known realized variance. In the present context, it was studied in Phillips and Yu (unpublished paper), and it follows that \( \sqrt{N} \left( \sum_{i=1}^{N} \hat{\eta}_{N,i}^{(1)} \right) = O_p(H^{-1/2}) \). This means that unless \( H = O(1) \) this term will be asymptotically irrelevant for the realised kernel. Next

\[
\sqrt{N} \left( \hat{\eta}_{N,i}^{(2)} \right)^2 = 4N^2 \left( \frac{\Delta_{N,i}}{\sigma^2(\tau_{N,i-1})} \right)^2 \frac{1}{N} \sum_{i=1}^{N} k_i \hat{\eta}_{N,i-1}^2 \frac{1}{N} \sum_{h=1}^{H-1} k_h \hat{\eta}_{N,i-h}^2 \frac{1}{N} \sum_{h=1}^{H-1} k_h \hat{\eta}_{N,i-h}^2,
\]

where \( (H-1) \sum_{h=1}^{H-1} k_h \hat{\eta}_{N,i-h}^2 \approx \sum_{h=1}^{H-1} k_h \hat{\eta}_{N,i-h}^2 \approx \sigma^2(\tau_{N,i-1}) \Delta_{N,i} \). Since \( N^2 \hat{\eta}_{N,i}^2 = D_{N,i} \sigma^2(\tau_{N,i-1}) \Delta_{N,i} \), we have

\[
N \int \limits_0^1 \sigma^2(u) du \int \limits_0^1 \sigma^2(u) du = N \int \left( \sum_{i=1}^{N} \hat{\eta}_{N,i}^{(2)} \right)^2 = \sum_{i=1}^{N} \sigma^4(\tau_{N,i-1}) \hat{E}(D_{N,i}^2 | \mathcal{F}_{N,i-1}) + o_p(1).
\]

Now we follow Phillips and Yu (unpublished paper) and write

\[
E(\hat{D}_{N,i}^2 | \mathcal{F}_{N,i-1}) = D_{N,i}^2 \frac{E(D_{N,i}^2 | \mathcal{F}_{N,i-1})}{E(D_{N,i} | \mathcal{F}_{N,i-1})} - \left[ D_{N,i} - E(D_{N,i} | \mathcal{F}_{N,i-1}) \right] \frac{E(D_{N,i}^2 | \mathcal{F}_{N,i-1})}{E(D_{N,i} | \mathcal{F}_{N,i-1})}.
\]

Now

\[
N \int \frac{1}{N} \sum_{i=1}^{N} \sigma^4(\tau_{N,i-1}) \left[ D_{N,i} - E(D_{N,i} | \mathcal{F}_{N,i-1}) \right] \frac{E(D_{N,i}^2 | \mathcal{F}_{N,i-1})}{E(D_{N,i} | \mathcal{F}_{N,i-1})} = \hat{o}_p(1),
\]

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

\[
\frac{N}{H} \int \limits_0^1 \sigma^4(\tau_{N,i-1}) \hat{E}(D_{N,i}^2 | \mathcal{F}_{N,i-1}) + o_p(1).
\]

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

**Proof of Theorem A.5.** Follows by combining the results of Propositions A.3–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 \( K(U) \xrightarrow{P} \frac{-E[U]}{\sigma^2(U)} \) when \( H = c_0 n^{1/2} \). □

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

\[
E[\exp(\text{itr}(AK(X)))] = E \left[ \sum_{j=1}^{d} \lambda_j \text{tr}(K(X)a_j a_j^T) \right],
\]

where \( A = \sum_{j=1}^{d} \lambda_j a_j a_j^T \) 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 \( K(a_j'Y) = K(a_j'X) \), the univariate kernel applied to the process \( X_i \). 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 \( H \to n^\eta \) for \( \eta \to (0, 1) \) and \( m/n \to 0 \), we will establish that

\[
\sqrt{N} \left( \hat{N} - \int_0^1 \sigma^2(u) du \right) \xrightarrow{L} \text{MN} \left( 0, 4k_0 \int_0^1 \sigma^2(u) du \right),
\]

which \( \sigma^2(u) = \Sigma(u) \otimes \Sigma(u) \). This will then complete the theorem. The univariate proof implies that we can replace \( K(a_j'Y) \) by \( \hat{K}(a_j'Y) \)

\[
= \sum_{i=1}^{n} x_{i,j}^2 + 2 \sum_{i=1}^{n} x_{i,j} x_{i,j}^T \sum_{i=1}^{n} k_h x_{i,j-h} \hat{\epsilon}_n, \quad \text{where } x_{i,j} = a_j' \tau_{N,i-1} \Delta_{N,i} \hat{\epsilon}_n \text{ and } \hat{\epsilon}_n = W_{N,i} - W_{T_{1,i-1}}.
\]

But this raises no new principles and so we can see that using the same method as before

\[
\sqrt{N} \left[ K(a_j'Y) - \int_0^1 \sigma^2(u) du \right] \xrightarrow{L} \text{MN} \left( 0, 4k_0 \int_0^1 \left( \sigma^2(u) \right)^2 \frac{X_2(u)}{X_1(u)} du \right).
\]

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

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

---

8 It is well known that the distribution is characterized by the matrix characteristic function \( E[\exp(\text{itr}(AK(X)))] \). Without loss of generality we can assume \( A = K(X) \) is symmetric and \( tr(K(X)) = \sum_{j=1}^{d} tr(a_j K(X)) = \sum_{j=1}^{d} \hat{o}_p(1) + \sum_{i=1}^{d} (a_i + a_j) \hat{o}_p(1) = \frac{1}{2} tr((A + A^T) K(X)). \)
A.4. Optimal choice of bandwidth

The problem is simply to minimise the squared bias plus the contribution from the asymptotic variance with respect to \( c_0 \). Set
\[
IQ = \int_0^\infty \sigma^2(u)\,du.
\]
The first order conditions of \( \min_{c_0} \{ - c_0 k'(0)^2 \omega^4 + c_0 \} \) yield the optimal value for \( c_0 \)
\[
c_0 = \left( \frac{k'(0)^2}{k'^2IQ} \right)^{1/5} = c^* \left( \frac{k'(0)^2}{k'^2IQ} \right)^{1/5}.
\]
With \( H^* = c^* \theta^{4/5}n^{1/5} \) the asymptotic bias is given by
\[
- \frac{k'(0)^2\omega^4}{k'^2IQ} \left( \frac{k'(0)^2\omega^2}{k'^2IQ} \right)^{2/5} k'(0)^2n^{-1/5} = k'(0)^2 \omega^2 \left( \frac{k'^2IQ}{k'^2IQ} \right)^{2/5} n^{-1/5},
\]
and the asymptotic variance is
\[
\frac{k'(0)^2\omega^4}{k'^2IQ} \left( 4k'^2IQn^{-2/5} = 4k'(0)^2\omega^2 \left( \frac{k'^2IQ}{k'^2IQ} \right)^{2/5} n^{-2/5}. \right.
\]

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 \( t \) is, in fact, the price recorded at time \( t_i \leq t \), for \( i = 1, \ldots, d \). With Refresh Time we have \( t \geq t_i \rightarrow t \), so that
\[
X(t_i) = Y(t_i) + U(t_i) = Y(t_i) + U(t_i) - Y(t_i) - Y(t_i) = \left( t_i - Y(t_i) \right).
\]

The endogenous component that is induced by refresh time is
\[
Y(t_i) - Y(t_i).
\]
But this is exactly the sort of dependence that Assumption U can accommodate through correlation between \( W \) and \( W \), and the (random) coefficients \( \theta(t) \), \( h = 0, 1, \ldots \).

References


