

Econometric analysis of multivariate realised QML: efficient positive semi-definite estimators of the covariation of equity prices

NEIL SHEPHARD

*Nuffield College, New Road, Oxford OX1 1NF, UK,
Department of Economics, University of Oxford*
✉

Oxford-Man Institute, University of Oxford,
neil.shephard@economics.ox.ac.uk

DACHENG XIU

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

First full draft: February 2012

This version: 24th April 2012

Abstract

Estimating the covariance and correlation between assets using high frequency data is challenging due to market microstructure effects and Epps effects. In this paper we extend Xiu's univariate QML approach to the multivariate case, carrying out inference as if the observations arise from an asynchronously observed vector scaled Brownian model observed with error. Under stochastic volatility the resulting 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 achieve the non-parametric efficiency bound for this problem. The estimator is also analysed using Monte Carlo methods and applied on equity data that are distinct in their 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 semi-definite and deals with both market microstructure effects such as bid/ask bounce and non-synchronous recording of data (the so-called Epps (1979) effect). Positive-definiteness is important for it allows us to define a coherent realised QML estimator of correlations and betas, objects of central importance in applied 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. The estimator is also analysed using Monte Carlo methods and applied on equity data in a string of two dimensional cases and a high dimensional case.

Our results suggest that our methods deliver particularly strong gains over existing methods in problems where the data is unbalanced: that is where some assets do not trade a great deal 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 particular line of research and opening up many more areas of potential 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 scaled Brownian motion, delivering the obvious QML estimator which is the realised variance. Multivariate versions of these estimators, based on rotated multivariate Brownian motion, 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 by other researchers.

The second is the Hayashi and Yoshida (2005) estimator, which is the QML estimator for the corresponding multivariate problem but where there is irregularly spaced non-synchronous data. Again the underlying log-price is modelled as a rotated vector Brownian motion.

Neither of the above estimators dealt with noise. Xiu (2010) studied the univariate QML estimator where the scaled Brownian motion is observed with Gaussian noise. He called this the univariate “realised QML estimator” and showed this was an effective estimator under some general semimartingale assumptions and under non-Gaussian noise. We also note the related Zhou (1996), Zhou (1998), Andersen, Bollerslev, Diebold, and Ebens (2001) and Hansen, Large, and Lunde (2008).

Our paper brings together all of 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

As well as this QML theme, a number of authors have approached this sophisticated multivariate problem using a variety of non-QML 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. These are sampled sufficiently sparsely that they could roughly ignore the effect of noise and non-synchronous trading. Related to those papers 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 preaveraging estimator of Christensen, Kinnebrock, and Podolskij (2010). Both use so-called refresh time sampling, suggested first in this context by the former paper. Neither converges at the optimal rate.

Two other estimators have been suggested which rely on a polarisation property 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. Such nonparametric estimators are heavily influenced by bandwidth or tuning parameter selection and so can be tricky to use in practice.

Finally, we note that related univariate work includes, for example, Zhou (1996), Zhou (1998), Hansen and Lunde (2006), Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), Jacod, Li, Mykland, Podolskij, and Vetter (2009), Andersen, Bollerslev, Diebold, and Labys (2000), Bandi and Russell (2008), Kalnina and Linton (2008), Li and Mykland (2007), Gloter and Jacod (2001a), Gloter and Jacod (2001b), Kunitomo and Sato (2009), Reiss (2011), Large (2011) and Hansen and Horel (2009).

1.4 More details on our paper

In this paper we use a QML estimator in the multivariate case where we model efficient prices as following correlated Brownian motion observed at very irregularly spaced and asynchronously recorded datapoints. Each observation is cloaked in noise. We provide some 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 simple smoothing algorithm. This has the virtue that the resulting estimator of the integrated covariance is positive-semidefinite. In practice it can be computed rapidly, even in significant dimensions.

1.5 Three noteworthy papers

There are three papers which are closest to our approach.

Aït-Sahalia, Fan, and Xiu (2010) employed the univariate estimator of Xiu (2010) applied to the multivariate case using polarisation. That is, for example, if they wish to estimate the covariance between x_1 and x_2 , then they applied 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. For d -dimensional cases they repeatedly used the univariate methods on a variety of linear combination of assets and then combine the results to produce an estimator of the covariance matrix. Unfortunately the result 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: extending Xiu (2010) to the multivariate case.

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. But their results rely on exact synchronisation and so cannot be realistically applied to all of the data in our available samples.

1.6 Structure of the paper

The structure of our paper is as follows. In Section 2 we define our model which generates the quasi-likelihood and 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. This is followed by a section containing various additional remarks. 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.

2 Models

2.1 Notation

We consider a d -dimensional log-price process $x = (x_1, \dots, x_d)'$. 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

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

where we have ordered the times so that $0 \leq t_1 < \dots < t_i < \dots < t_n \leq T$. Notice that the t_i times must, thus, be distinct. Of course 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 are seen 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.

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 $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$,

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

where a is a vector of elements which are predictable locally bounded drifts, σ is a càdlàg volatility matrix process and W is a vector of independent Brownian motions. For reviews of the econometrics of this type of process see, for example, Ghysels, Harvey, and Renault (1996). Then the ex-post covariation is

$$[y, y]_T = \int_0^T \Sigma(u) du, \quad \text{where } \Sigma = \sigma\sigma',$$

where

$$[y, y]_T = \text{plim}_{n \rightarrow \infty} \sum_{j=1}^n \{y(\tau_j) - y(\tau_{j-1})\} \{y(\tau_j) - y(\tau_{j-1})\}',$$

(e.g. Protter (2004, p. 66–77) and Jacod and Shiryaev (2003, p. 51)) for any sequence of deterministic synchronized partitions $0 = \tau_0 < \tau_1 < \dots < \tau_n = T$ with $\sup_j \{\tau_{j+1} - \tau_j\} \rightarrow 0$ for $n \rightarrow \infty$. This is the quadratic variation of y .

Our interest is in estimating $[y, y]_T$.

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 will mean 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 . We write our data arrival as

$$x_i = Z_i y(t_i) + Z_i \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where ε_i is a vector of potential market microstructure effects. Throughout we will assume $E(\varepsilon_i) = 0$ and write $\text{Cov}(\varepsilon_i) = \Lambda$. It is conventional to assume Λ is diagonal, but we will not do this as yet¹. General time series discussions of missing data includes Harvey (1989, Ch. 6.4), Durbin and Koopman (2001, Ch. 2.7), Ljung (1989) and Basu and Reinsel (1996).

2.3 A Gaussian quasi-likelihood

We follow Xiu (2010) in proxying the Brownian semimartingale by Brownian motion, this time a multivariate version which is non-synchronously observed. The approach is to model

$$y(t) = \sigma W(t).$$

¹Corsi, Peluso, and Audrino (2012) use a slightly different approach. They update at equally spaced time points Ti/n , like every second, whether there is new data or not. At each time point they use a linear Gaussian state space model $x_i = Z_i y(Ti/n) + \varepsilon_i$, 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 no new prices are available at all then Z_i is a matrix of zeros.

Then writing

$$\Sigma = \sigma\sigma' = [y, y]_1,$$

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

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

Then we can think of the time series of observations $x_{1:n} = (x_1, \dots, x_n)'$ as a Gaussian state space model. A discussion of the corresponding literature is available in, for example, Harvey (1989), West and Harrison (1989), Durbin and Koopman (2001) and Cappe, Moulines, and Ryden (2009).

2.4 ML estimation via EM algorithm

Our goal is to develop positive semidefinite estimators of Σ , noting for us that Λ 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,

$$e_i = x_i - Z_i y(t_i), \quad u_i = y(t_i) - y(t_{i-1}),$$

of the form, writing $y_{1:n} = (y_1, \dots, y_n)'$,

$$\begin{aligned} & \log f(x_{1:n}|y_{1:n}; \Lambda) + \log f(y_{1:n}; \Sigma) \\ = & c - \frac{1}{2} \sum_{i=1}^n \log |Z_i \Lambda Z_i'| - \frac{1}{2} \sum_{i=1}^n e_i' (Z_i \Lambda Z_i')^{-1} e_i \\ & - \frac{1}{2} \sum_{i=2}^n \log |\Sigma| - \frac{1}{2} \sum_{i=2}^n \frac{1}{\Delta_i^n} u_i' \Sigma^{-1} u_i. \end{aligned}$$

Then the EM algorithm works with the

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

$$\begin{aligned}
&= c - \frac{1}{2} \sum_{i=1}^n \log |Z_i \Lambda Z_i'| - \frac{1}{2} \sum_{i=1}^n \mathbb{E} \left\{ e_i' (Z_i \Lambda Z_i')^{-1} e_i | x_{1:n}; \Lambda, \Sigma \right\} \\
&\quad - \frac{1}{2} \sum_{i=2}^n \log |\Sigma| - \frac{1}{2} \sum_{i=2}^n \frac{1}{\Delta_i^n} \mathbb{E} \left\{ u_i' \Sigma^{-1} u_i | x_{1:n}; \Lambda, \Sigma \right\}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E} \left\{ e_i' (Z_i \Lambda Z_i')^{-1} e_i | x_{1:n} \right\} &= \text{tr} \left\{ (Z_i \Lambda Z_i')^{-1} \mathbb{E}(e_i e_i' | x_{1:n}) \right\} \\
&= \text{tr} \left[(Z_i \Lambda Z_i')^{-1} \left\{ \hat{e}_{i|n} \hat{e}_{i|n}' + D_{i|n} \right\} \right],
\end{aligned}$$

and, writing $\hat{u}_{i|n} = \mathbb{E}(u_i | x_{1:n})$ and $N_{i|n} = \text{Mse}(u_i | x_{1:n})$, then

$$\mathbb{E} \left\{ u_i' \Sigma^{-1} u_i | x_{1:n} \right\} = \text{tr} \left\{ \Sigma^{-1} \mathbb{E}(u_i u_i' | x_{1:n}) \right\} = \text{tr} \left[\Sigma^{-1} \left\{ \hat{u}_{i|n} \hat{u}_{i|n}' + N_{i|n} \right\} \right].$$

Then the EM update is

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=2}^n \frac{1}{\Delta_i^n} \left\{ \hat{u}_{i|n} \hat{u}_{i|n}' + N_{i|n} \right\}, \quad \hat{\Lambda} = \frac{1}{n} \sum_{i=1}^n Z_i' \left\{ \hat{e}_{i|n} \hat{e}_{i|n}' + D_{i|n} \right\} Z_i.$$

As these updates are iterated, the sequence of $(\hat{\Sigma}, \hat{\Lambda})$ converges to a maximum in the quasi-likelihood function.

2.5 Recalling the disturbance smoother

Computing $\hat{e}_{i|n}$, $\hat{u}_{i|n}$, $D_{i|n}$ and $N_{i|n}$ is routine and rapid, if rather tedious to write down. It is carried out computationally efficiently using the ‘‘disturbance smoother’’. This starts out with the Kalman filter (e.g. Durbin and Koopman (2001, p. 67)), which is run forward in time $i = 1, 2, \dots, n$ through the data. In our case it takes on the form

$$v_i = x_i - Z_i \hat{y}_i, \quad F_i = Z_i (P_i + \Lambda) Z_i', \quad K_i = P_i Z_i' F_i^{-1}, \quad L_i = I - K_i Z_i$$

then

$$\hat{y}_{i+1} = \hat{y}_i + K_i v_i, \quad P_{i+1} = P_i L_i' + \Delta_i^n \Sigma.$$

Here $\hat{y}_i = \mathbb{E}(y_i | x_{1:i-1})$ 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 Σ or Λ . A typical selection for \hat{y}_1 is the opening auction price, 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 $d \times d_i$, $\hat{y}_{i+1|i}$ 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, which can be monitored as

$$\log f(x_{1:n}; \Lambda, \Sigma) = c - \frac{1}{2} \sum_{i=1}^n \log |F_i| - \frac{1}{2} \sum_{i=1}^n v_i' F_i^{-1} v_i.$$

The disturbance smoother (e.g. Durbin and Koopman (2001, p. 76)) is run backwards $i = n, n-1, \dots, 1$ through the data. It takes the form, writing $H_i = Z_i \Lambda Z_i'$, a $d_i \times d_i$ matrix

$$\begin{aligned} \widehat{e}_{i|n} &= H_i(F_i^{-1}v_i - K_i' r_i), & \widehat{u}_{i|n} &= \Delta_i^n \Sigma r_{i|n}, \\ D_{i|n} &= H_i - H_i(F_i^{-1} + K_i' M_i K_i) H_i, & N_{i|n} &= \Delta_i^n \Sigma - (\Delta_i^n)^2 \Sigma M_{i-1} \Sigma, \end{aligned}$$

where we recursively compute

$$r_{i-1} = Z_i' F_i^{-1} v_i + L_i' r_i, \quad M_{i-1} = Z_i' F_i^{-1} Z_i + L_i' M_i L_i,$$

starting out with $r_n = 0$, $M_n = 0$. Here $\widehat{e}_{i|n}$ is $d_i \times 1$ and $D_{i|n}$ is $d_i \times d_i$. While $\widehat{u}_{i|n}$ 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.

2.6 Sparse and subsampled realised QML

A virtue of the realised QML is that it is applied to all of the high frequency data, to form a quasi-likelihood. However, this estimator may have challenges if the noise has more complicated dynamics than i.i.d. one used here. This opens up the opportunity to define a ‘‘sparse realised QML’’ estimator, which corresponds nicely to the so-called sparse sampling realised variance. This uses the same realised QML approach but only employs a subset of the data: every k -th trade, where k could be for example, 2 or 10. By using every other trade price, for example, the ratio of the variability of efficient price moves to noise increases and so sensitivity to the i.i.d. assumption in the noise should reduce. Such sparse sampling raises no new theoretical or computational issues, we simply subset the data.

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 it employs all of the data in the sample while being less sensitive to the i.i.d. assumption.

We have been experimenting with these two approaches but do not have sufficient results on it yet to present them here. Hopefully they will appear in the next version.

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 (e.g. data is used every 20 seconds) in the quasi-likelihood. We can regard this the “sparse sampling” case as it does not potentially use all the data in the actual sample. Finally we discuss the impact of non-synchronised data on a fully non-synchronised quasi-likelihood.

3.1 Benchmark bivariate MLE

3.1.1 The model

We start with the constant covariance matrix case with equidistant observations. This means we have synchronised trading. We also assume the market microstructure effects are independent and initially normal. For simplicity of exposition we assume the dimension of the system $d = 2$.

Then we observe returns

$$r_{j,i} = x_{j,i} - x_{j,i-1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2,$$

where

$$\begin{aligned} x_{j,i} &= y_j(i/n) + \varepsilon_{j,i}, \quad i = 0, 1, 2, \dots, n, \quad j = 1, 2, \\ y(i/n) &= y((i-1)/n) + \sqrt{T/nu_i}, \end{aligned}$$

where

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \overset{i.i.d.}{\sim} N \left(\begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}, \begin{pmatrix} 0 \\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \right).$$

In the discrete time series literature x would be called a bivariate Gaussian “local level model” or “random walk plus noise model” (e.g. Harvey (1989) and Durbin and Koopman (2001, Ch. 2)). Here it will be analysed using “infill” asymptotics.

Suppose the observed returns are collected in r :

$$r = (r_{1,1}, r_{1,2}, \dots, r_{1,n}, r_{2,1}, \dots, r_{2,n})',$$

then the likelihood can be rewritten as

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

where $\Omega = \Delta \Sigma \otimes I_n + \Lambda \otimes J_n$. Here \otimes denotes the Kronecker product and the 2×2 matrices

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix},$$

and J_n is a $n \times n$ matrix

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

This is attractive as we know the eigenvalues and eigenvectors of J_n and so Ω .

3.1.2 The asymptotic theory

Before we give the bivariate case we just recall the univariate one, which allows for non-Gaussianity in the error,

$$\begin{pmatrix} n^{\frac{1}{4}} \left(\widehat{\Sigma}_{11} - \Sigma_{11} \right) \\ n^{\frac{1}{2}} \left(\widehat{\Lambda}_{11} - \Lambda_{11} \right) \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} 8\Lambda_{11}^{1/2} \Sigma_{11}^{3/2} T^{-1/2} & 0 \\ 0 & 2\Lambda_{11}^2 \end{pmatrix} \right),$$

see Stein (1987), Gloter and Jacod (2001a), Gloter and Jacod (2001b), Ait-Sahalia, Mykland, and Zhang (2005) and Xiu (2010).

We now go onto the Gaussian bivariate case, and deliver basically, the Fisher Information.

Theorem 1 (Bivariate MLE) *The maximum likelihood estimators $\widehat{\Sigma}$ and $\widehat{\Lambda}$ satisfy the central limit theorem as $n \rightarrow \infty$*

$$n^{\frac{1}{4}} \begin{pmatrix} \widehat{\Sigma}_{11} - \Sigma_{11} \\ \widehat{\Sigma}_{12} - \Sigma_{12} \\ \widehat{\Sigma}_{22} - \Sigma_{22} \end{pmatrix} \xrightarrow{d} N(0, V), \quad n^{\frac{1}{2}} \begin{pmatrix} \widehat{\Lambda}_{11} - \Lambda_{11} \\ \widehat{\Lambda}_{22} - \Lambda_{22} \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} 2\Lambda_{11}^2 & 0 \\ 0 & 2\Lambda_{22}^2 \end{pmatrix} \right).$$

Here for a 3×1 vector $\theta = \text{vech}(\Sigma) = (\Sigma_{1,1}, \Sigma_{1,2}, \Sigma_{2,2})'$ the V matrix is such that $V^{-1} = \frac{\partial \Psi_{\theta}}{\partial \Sigma_{\theta'}}$ 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 $\Lambda_{ij}^* = \Lambda_{ij}/T$,

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

Proof. Given in the Appendix.

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

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

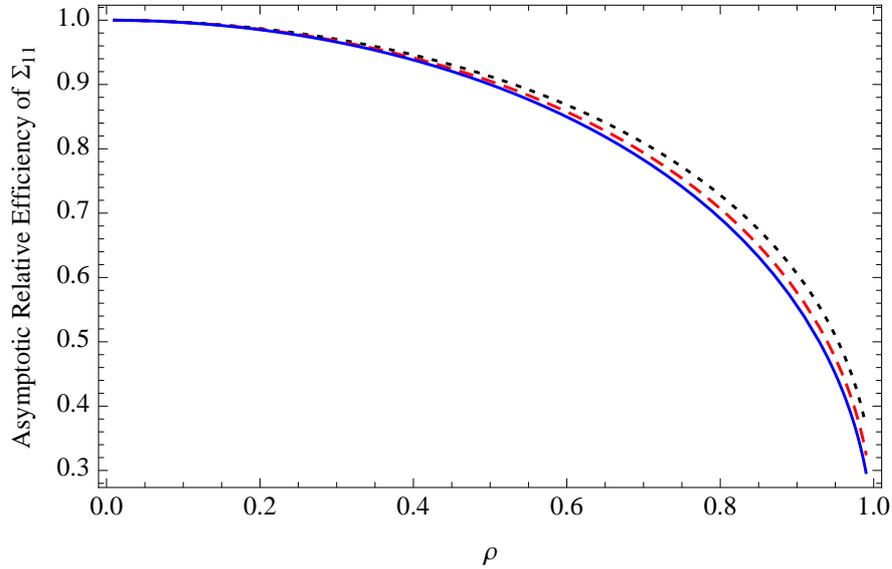


Figure 1: The figure plots the relative efficiency for bivariate MLE of Σ_{11} over the univariate alternative, against the correlation. $\Sigma_{11} = 0.25^2$. As the number falls below zero the gains from bivariate MLE become greater. The blue line, the red dashed line, and the black dotted line correspond to the cases with $\Sigma_{22} = 0.3^2, 0.25^2$, and 0.2^2 , respectively.

3.2 Bivariate QMLE with equidistant observations

3.2.1 Assumptions

In order to understand the estimators under stochastic volatility, we need to be clear about our assumptions on the volatility process and the noise.

Assumption 1. The underlying latent d -dimensional log-price process satisfies

$$dy(t) = \mu(t)dt + \sigma(t)dW(t),$$

where the drift is predictable locally bounded, the $d \times d$ volatility process is locally bounded Itô semimartingales and W is a d -dimensional Brownian motion.

Assumption 2. The noise ε_i is a vector random variable which is independent and identically distributed, and independent of t_i , W , σ and has fourth moments.

3.2.2 The asymptotic theory

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

$$\begin{pmatrix} n^{\frac{1}{4}} \left(\widehat{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_t^2 dt \right) \\ n^{\frac{1}{2}} \left(\widehat{\Lambda}_{11} - \Lambda_{11} \right) \end{pmatrix} \xrightarrow{d} MN \left(0, \begin{pmatrix} (5R_T+3) \Lambda_{11}^{1/2} \left(\frac{1}{T} \int_0^T \sigma_t^2 dt \right)^{3/2} T^{-1/2} & 0 \\ 0 & 2\Lambda_{11}^2 + cum_4[\varepsilon] \end{pmatrix} \right),$$

which is a rewrite of the result due to Xiu (2010). This shows that the asymptotic variance of the estimator increases with R_T keeping $\frac{1}{T} \int_0^T \sigma_t^2 dt$ fixed. Here cum_4 denotes the fourth cumulant, recalling this is zero under Gaussianity.

We now extend this to the multivariate case. In the presence of stochastic volatility, we apply the same estimator, misspecifying the model intentionally. The asymptotic theory for the resulting realised QML estimator $\widehat{\Sigma}$ is given below.

Theorem 2 (Bivariate QMLE) *In the presence of stochastic volatility, we have*

$$n^{\frac{1}{4}} \begin{pmatrix} \widehat{\Sigma}_{11} - \frac{1}{T} \int_0^T \Sigma_{11,t} dt \\ \widehat{\Sigma}_{12} - \frac{1}{T} \int_0^T \Sigma_{12,t} dt \\ \widehat{\Sigma}_{22} - \frac{1}{T} \int_0^T \Sigma_{22,t} dt \end{pmatrix} \xrightarrow{\mathcal{L}_X} MN(0, V_Q), \quad (2)$$

$$n^{\frac{1}{2}} \begin{pmatrix} \widehat{\Lambda}_{11} - \Lambda_{11} \\ \widehat{\Lambda}_{22} - \Lambda_{22} \end{pmatrix} \xrightarrow{d} N \left(0, \begin{pmatrix} 2\Lambda_{11}^2 + cum_4[\varepsilon_1] & 0 \\ 0 & 2\Lambda_{22}^2 + cum_4[\varepsilon_2] \end{pmatrix} \right), \quad (3)$$

where V_Q is

$$\begin{aligned} V_Q &= \frac{1}{4} \left(\frac{\partial \Psi_\theta}{\partial \Sigma_{\theta'}} \right)^{-1} \left\{ Avar^{(2)} + Avar^{(3)} + Avar^{(4)} \right\} \left\{ \left(\frac{\partial \Psi_\theta}{\partial \Sigma_{\theta'}} \right)^{-1} \right\}', \\ Avar^{(2)} &= 2 \sum_{l,s,u,v=1}^2 \int_0^\infty \frac{\partial \omega^{v,u}(\Sigma, \Lambda, x)}{\partial \Sigma_\theta} \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma'_\theta} dx \left(\frac{1}{T} \int_0^T \Sigma_{sv,t} \Sigma_{ul,t} dt \right), \\ Avar^{(3)} &= 4 \sum_{l,s,v=1}^2 \Lambda_{ll}^* \int_0^\infty \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma_\theta} \frac{\partial \omega^{l,v}(\Sigma, \Lambda, x)}{\partial \Sigma'_\theta} \pi^2 x^2 dx \left(\frac{1}{T} \int_0^T \Sigma_{sv,t} dt \right), \\ Avar^{(4)} &= 2 \sum_{l,s=1}^2 \Lambda_{ll}^* \Lambda_{ss}^* \int_0^\infty \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma_\theta} \frac{\partial \omega^{l,s}(\Sigma, \Lambda, x)}{\partial \Sigma'_\theta} \pi^4 x^4 dx. \end{aligned}$$

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 there derivation is different.

3.3 Bivariate QMLE with synchronised but irregularly space observations

When it comes to irregular spaced observations, the asymptotic variance could change. Mykland and Zhang (2006) has studied this in the context of realised variance estimation, employing a concept called the quadratic variation of the sampling times.

Here we study a somewhat different problem. We now use a quasi-likelihood based upon some synchronised times $\{t_1, t_2, \dots, t_n\}$, which we assume accurately reflects the times of the irregularly spaced but synchronised data. These times imply a collection of irregular time increments $\{\Delta_i^n = t_i - t_{i-1}, 1 \leq i \leq n\}$. We then make the following assumption.

Assumption 3. We assume that

$$\Delta_i^n = \bar{\Delta} (1 + \xi_i), \quad i = 1, 2, \dots, n, \quad \bar{\Delta} = \frac{T}{n}$$

where $E(\xi_i) = 0$, $\text{Var}(\xi_i) < \infty$ and $\{\xi_i, 1 \leq i \leq n\}$ are *i.i.d.*. Further we assume Y and $\{\xi_i\}$ are independent.

Assumption 3 means that $\xi_i = O_p(1)$. The implication is that there is modelled heterogeneity in the time gap between observations, but the individual gaps shrink at rate $O_p(n^{-1})$.

The following Corollary shows that the heterogeneity has no impact on the asymptotic distribution of the realised QML estimator. The reason for this is that the effect is less important than the presence of noise.

To highlight this point, it is sufficient to consider the case of constant covariance structure. We follow Ait-Sahalia, Mykland, and Zhang (2005) and Ait-Sahalia and Mykland (2003) in expanding the Fisher Information with respect to the average length of the sampling intervals.

Corollary 1 *Suppose Σ is constant, then under Assumption 3 the asymptotic variance of the MLE is the same as that in Theorem 1.*

Proof. Given in the Appendix.

3.4 Bivariate QMLE with non-synchronous observations

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 Ait-Sahalia, Fan, and Xiu (2010). See also the earlier more informal papers by Harris, McMcInish, Shoesmith, and Wood (1995) and Martens (2003).

We can apply our ML approach to a synchronised series of datapoints, ignoring the effect of staleness. Such synchronisation does change the limit theory results. However, the consistency is not affected using these refreshed pairs, which can be shown following the same argument in Ait-Sahalia, Fan, and Xiu (2010, Theorem 3). A full discussion of this will be given in a later version of this paper.

In this paper, synchronisation is not needed for our approach as it is applied directly to the data we have in our hand. In fact, our approach is a generalization of the Hayashi and Yoshida (2005) approach to the case where microstructure noise is present.

4 Additional developments

4.1 QML correlation and regression estimator

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

$$\widehat{\rho}_{12} = \frac{\widehat{\Sigma}_{12}}{\sqrt{\widehat{\Sigma}_{11}\widehat{\Sigma}_{22}}} \in [-1, 1],$$

which estimates

$$\rho_{12} = \frac{\frac{1}{T} \int_0^T \Sigma_{12,t} dt}{\sqrt{\left(\frac{1}{T} \int_0^T \Sigma_{11,t} dt\right) \left(\frac{1}{T} \int_0^T \Sigma_{22,t} dt\right)}} \in [-1, 1].$$

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

$$\widehat{\beta}_{1|2} = \frac{\widehat{\Sigma}_{12}}{\widehat{\Sigma}_{22}},$$

which estimates

$$\beta_{1|2} = \frac{\frac{1}{T} \int_0^T \Sigma_{12,t} dt}{\frac{1}{T} \int_0^T \Sigma_{22,t} dt}.$$

The corresponding limit theory follows by the application of the delta method:

$$\text{Avar}(\widehat{\rho}_{12}) = \nu_\rho V_Q \nu'_\rho, \text{ and } \text{Avar}(\widehat{\beta}_{1|2}) = \nu_\beta V_Q \nu'_\beta,$$

where

$$\nu_\rho = \left(-\frac{1}{2} \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}^3 \Sigma_{22}}}, \frac{1}{\sqrt{\Sigma_{11} \Sigma_{22}}}, -\frac{1}{2} \frac{\Sigma_{12}}{\sqrt{\Sigma_{11} \Sigma_{22}^3}} \right)', \text{ and } \nu_\beta = \left(0, \frac{1}{\Sigma_{22}}, -\frac{\Sigma_{12}}{\Sigma_{22}^2} \right)'.$$

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

4.2 Miniature QML based estimation

So far we have carried out 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

$$0 = b_0 < b_1 < \dots < b_B = T.$$

Then we can compute a 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 QML’’ estimator and write it as $\widehat{\Sigma}_i$. For fixed block sizes the resulting estimator, as the sample goes to infinity, is

$$n_i^{1/4} \left(\widehat{\Sigma}_{11,i} - \frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right) \xrightarrow{\mathcal{L}_X} MN \left(0, \frac{(5R_{ii}+3) \Lambda_{11}^{1/2}}{(b_i - b_{i-1})^{1/2}} \left(\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^{3/2} \right),$$

where $n_i = n(b_i - b_{i-1})/T$,

$$R_{ii} = \frac{\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^4 dt}{\left(\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^2} \geq 1.$$

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

Corollary 2 *Define the unblocked QML estimator of Σ_{11} ,*

$$\widetilde{\Sigma}_{11} = \frac{1}{T} \sum_{i=1}^B (b_i - b_{i-1}) \widehat{\Sigma}_{11,i}$$

then for fixed b_i and B we have as $n \rightarrow \infty$

$$n^{1/4} \left(\widetilde{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_t^2 dt \right) \xrightarrow{\mathcal{L}_X} MN \left(0, \frac{1}{T^{3/2}} \sum_{i=1}^B (b_i - b_{i-1}) (5R_{ii}+3) \Lambda_{11}^{1/2} \left(\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} \sigma_t^2 dt \right)^{3/2} \right).$$

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 key virtue though is that $\int_0^t \sigma_t^2 dt$ and $\int_0^t \sigma_t^4 dt$ are of bounded variation as a function of t and so are both $O_p(t)$ as $t \downarrow 0$. This means that $R_{ii} \simeq 1 + O_p(b_i - b_{i-1})$ 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 this is achieved by allowing $b_i - b_{i-1}$ to get

small. This is achieved by allowing the gaps $\max_i (b_i - b_{i-1}) = o(1)$ 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

$$n^{1/4} \left(\tilde{\Sigma}_{11} - \frac{1}{T} \int_0^T \sigma_t^2 dt \right) \xrightarrow{\mathcal{L}_X} MN \left(0, 8\Lambda_{11}^{1/2} \left(\frac{1}{T} \int_0^T \sigma_t^3 dt \right) T^{-1/2} \right).$$

This approach is also efficient when the variance of the noise is time-varying, for Λ_{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 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 ρ_{12} keeping Σ_{11}, Σ_{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.

5 Monte Carlo experiments

5.1 Monte Carlo design

Throughout we quite closely follow the design of Ait-Sahalia, Fan, and Xiu (2010), which is a bivariate model. Throughout each day will be taken as lasting $T = 1/252$ units of time, so $T = 1$ would represent a year. Here we recall the structure of their model

$$\begin{aligned} dy_{it} &= \alpha_{it} dt + \sigma_{it} dW_{it} \\ d\sigma_{it}^2 &= \kappa_i (\bar{\sigma}_i^2 - \sigma_{it}^2) dt + s_i \sigma_{it} dB_{it} + \sigma_{it} J_{it}^V dN_{it} \end{aligned}$$

where $E(dW_{it} dB_{jt}) = \delta_{ij} \rho_i dt$ and $E(dW_{1t} dW_{2t} | \rho^*) = \rho^* dt$. Here $\kappa_i > 0$.

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

We now depart slightly from their setup.

For each day we draw independently $\sigma_{i0}^2 \sim \Gamma(2\kappa_i \bar{\sigma}_i^2 / s_i^2, s_i^2 / 2\kappa_i)$ over $i = 1, 2$, which means each replication will be independent. Likewise for each separate day we simulate independently

$$\rho^* \sim \rho_0 \text{Beta}(\rho_1^*, \rho_2^*),$$

where $\rho_0 = \sqrt{(1 - \rho_1^2)(1 - \rho_2^2)}$, guarantees the positive-definiteness of the covariance matrix of the four Brownian motions (W_1, W_2, B_1, B_2) . This means $E(\rho^*) = \rho_0 \rho_1^* / (\rho_1^* + \rho_2^*)$ and $sd(\rho^*) = \rho_0 \sqrt{\rho_1^* \rho_2^*} / \{(\rho_1^* + \rho_2^*) \sqrt{\rho_1^* + \rho_2^* + 1}\}$. 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 ρ_2^* are given in Table 1. To check our limit theory calculations, Figure 2 plots the histograms of the

	a_i	α_i	ρ_i	κ_i	θ_i	μ_i	λ_i	$\bar{\sigma}_i^2$	s_i	
$i = 1$	0.005	0.05	-0.6	3	-5	0.8	12	0.16	0.8	$\rho_1^* = 2$ $\rho_0 = 0.529$ $\rho_2^* = 1$ $E(\rho^*) = 0.176$ $sd(\rho^*) = 0.125$
$i = 2$	0.001	0.01	-0.75	2	-6	1.2	36	0.09	0.5	

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

standardized estimators (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.

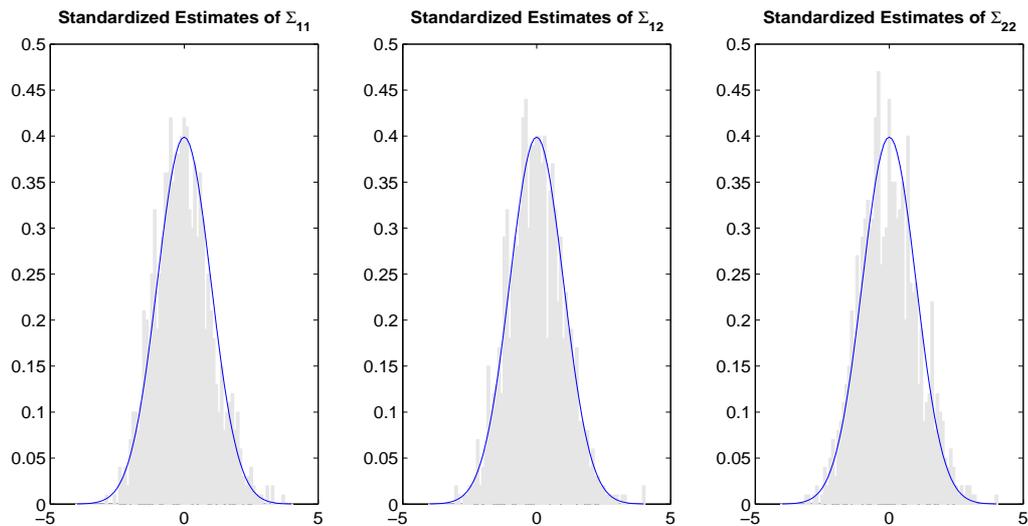


Figure 2: The figure plots the histograms of the standardized estimates, which verify the asymptotic theory developed in Theorem 2. The standardisation is carried out using the infeasible true random asymptotic variance for each replication.

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 as n we randomly and uniform scatter those 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 Σ_{11} , Σ_{22} , Σ_{12} , $\beta_{1|2}$, $\beta_{2|1}$ and $\rho_{1,2}$. The six are: (i) multivariate QML, (ii) multistep QML estimator, (iii) blocked multivariate QML², (iv) multivariate QML but using the reduced data synchronised by Refresh Time, (v) the multivariate realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) which uses Refresh Time, (vi) Ait-Sahalia, Fan, and Xiu (2010) which uses polarisation and Refresh time. We will use the notation θ_{QML} , θ_{Step} , θ_{Bloc} , θ_{RT} , θ_{Kern} , θ_{Pol} , respectively, where θ is some particular parameter. We write this generically as θ_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 in the estimation, the others are based 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 $\left| n^{1/4} (\hat{\theta}_L - \theta_L) \right|$ for various values of n , L and F . The table also shows $n^{1/4}$, which allows us to see the actual speed by which the quantiles for $\left| \hat{\theta}_L - \theta_L \right|$ contract.

The results indicate that all six estimators perform roughly similarly for Σ_{11} and Σ_{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 QML being the best by a small margin. In this setup Bloc was a little disappointing and needs more study. QML outperformed Step when $F = 0.1$ 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 Σ_{12} there are signs that the QML type estimators ‘‘QML’’, ‘‘Step’’ and ‘‘Bloc’’ perform better than ‘‘RT’’. All of these estimators seem to perform more strongly than the existing ‘‘Kern’’ and ‘‘Pol’’ estimators. The differences are less important in the case where $F = 0.5$.

²The number of blocks was taken to be $\sqrt{n}/3$. Sometimes this yielded very unequally sized blocks in which case we selected $n/6$ but requiring at least two observations for each asset.

When we move onto $\rho_{1,2}$ the differences become more significant, although we recall that 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 the same general impression holds.

Before we progress to the regression case it is helpful to calibrate how accurately we have estimated $\rho_{1,2}$ in the QML case. When $F = 0.1$ the quantile is 1.98 with $n = 117$, so the corresponding quantile for $|\hat{\theta}_{QML} - \theta_{QML}|$ is 0.602. When $n = 1,170$ it is 0.265. When $n = 11,700$ it is 0.142. In the balanced case $F = 0.5$ the corresponding results are 0.459, 0.262 and 0.148. Hence balanced data helps, but not by very much as long as n is moderately large and the QML method is used. Balancing is much more important for RT, Kern and Pol. We think this makes the QML approach distinctly promising. A final point is worth noting. Even though $n = 11,700$ the quantiles in the balanced case of 0.148 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 performing around an order of magnitude better than RT, Kern and Pol. The results for QML and Bloc are roughly the same with perhaps QML being very slightly better.

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 a suggestion that only data from a single exchange is used. 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

			$F = 0.9$						$F = 0.5$						$F = 0.1$					
	n	$n^{1/4}$	QML	Step	Bloc	RT	Kern	Pol	QML	Step	Bloc	RT	Kern	Pol	QML	Step	Bloc	RT	Kern	Pol
$\widehat{\Sigma}_{11}$	117	3.29	0.46	0.46	0.53	0.90	0.81	0.79	0.50	0.54	0.59	0.59	0.56	0.55	0.62	0.79	0.75	0.95	0.69	0.76
	1,170	5.84	0.39	0.39	0.46	0.75	0.85	0.78	0.44	0.46	0.56	0.50	0.74	0.49	0.65	0.76	1.18	0.73	0.85	0.76
	11,700	10.3	0.36	0.36	0.49	0.64	0.96	0.67	0.43	0.44	0.61	0.44	0.95	0.44	0.64	0.70	1.60	0.67	0.97	0.70
$\widehat{\Sigma}_{12}$	117	3.29	0.24	0.22	0.22	0.32	0.38	0.34	0.20	0.17	0.21	0.19	0.22	0.23	0.24	0.22	0.23	0.32	0.36	0.36
	1,170	5.84	0.18	0.18	0.19	0.30	0.35	0.31	0.18	0.17	0.18	0.18	0.22	0.20	0.24	0.23	0.24	0.27	0.35	0.32
	11,700	10.3	0.17	0.17	0.19	0.24	0.36	0.27	0.16	0.16	0.17	0.17	0.22	0.18	0.21	0.21	0.26	0.25	0.34	0.27
$\widehat{\Sigma}_{22}$	117	3.29	0.26	0.35	0.33	0.31	0.38	0.33	0.24	0.19	0.23	0.21	0.24	0.20	0.21	0.16	0.19	0.29	0.41	0.34
	1,170	5.84	0.22	0.28	0.27	0.36	0.45	0.29	0.17	0.16	0.16	0.17	0.26	0.16	0.15	0.14	0.15	0.34	0.45	0.28
	11,700	10.3	0.21	0.22	0.26	0.22	0.42	0.22	0.14	0.14	0.14	0.14	0.26	0.14	0.11	0.11	0.12	0.24	0.45	0.23
$\widehat{\rho}_{1,2}$	117	3.29	1.98	1.98	1.92	3.43	2.95	4.41	1.51	1.51	1.37	2.26	1.59	2.31	1.98	1.98	1.59	3.47	2.83	4.59
	1,170	5.84	1.55	1.55	1.71	2.90	2.54	3.01	1.53	1.53	1.39	1.75	1.71	1.89	2.14	2.14	1.85	2.91	2.50	2.96
	11,700	10.3	1.46	1.46	1.92	2.23	2.61	2.56	1.40	1.40	1.66	1.59	2.10	1.70	1.82	1.82	2.30	2.15	2.57	2.47
$\widehat{\beta}_{1 2}$	117	3.29	4.45	6.50	4.03	6.19	7.71	11.2	2.93	3.08	2.87	3.63	3.41	4.11	4.34	4.40	3.67	6.03	7.63	10.8
	1,170	5.84	3.16	3.11	3.50	4.21	4.83	4.76	2.58	2.68	2.75	3.08	3.02	3.09	3.38	3.54	3.89	4.41	5.01	5.04
	11,700	10.3	2.76	2.71	4.85	4.31	5.58	4.39	2.69	2.60	3.05	2.83	3.43	3.20	3.47	3.56	4.45	4.02	5.09	4.73
$\widehat{\beta}_{2 1}$	117	3.29	2.09	2.16	1.89	7.70	2.77	5.55	1.91	2.14	1.50	4.81	1.70	3.36	2.18	3.08	1.54	8.76	3.26	5.75
	1,170	5.84	1.78	1.89	2.07	5.05	2.94	3.99	1.93	2.07	1.97	2.56	2.55	2.53	2.58	3.31	2.51	5.18	3.11	4.77
	11,700	10.3	1.82	1.63	2.82	3.27	3.21	3.28	1.77	1.71	3.08	1.84	3.61	2.13	2.83	2.30	4.27	2.93	3.40	3.17

Table 2: Monte Carlo results for the volatility, covariance, correlation and beta estimation. Throughout we report the 0.9 quantiles of $\left|n^{1/4} \left(\widehat{\theta}_L - \theta_L\right)\right|$ 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.

median of multiple trades which occur in the same second. This blurring will mean for some TAQ time stamps multiple assets will indeed trade at the same time. Of course this feature would disappear if lower latency data was used. When there are multiple trades in the same second, we record the average trade price and can be thought of as a particular type 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. The aim of the selection is to have 2 infrequently traded and 11 highly traded assets. This will allow us to look at a variety of different types of data environments on the estimators.

The assets we study are the Spyder (SPY), an S&P 500 exchange traded fund, along with some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) 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). That makes 13 series in all. 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 intensity figure 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) and we name it after its creators Asger Lunde and Kevin Sheppard.

The Lunde-Sheppard 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 Lunde-Sheppard trading intensity graph demonstrates that the Refresh time approach is

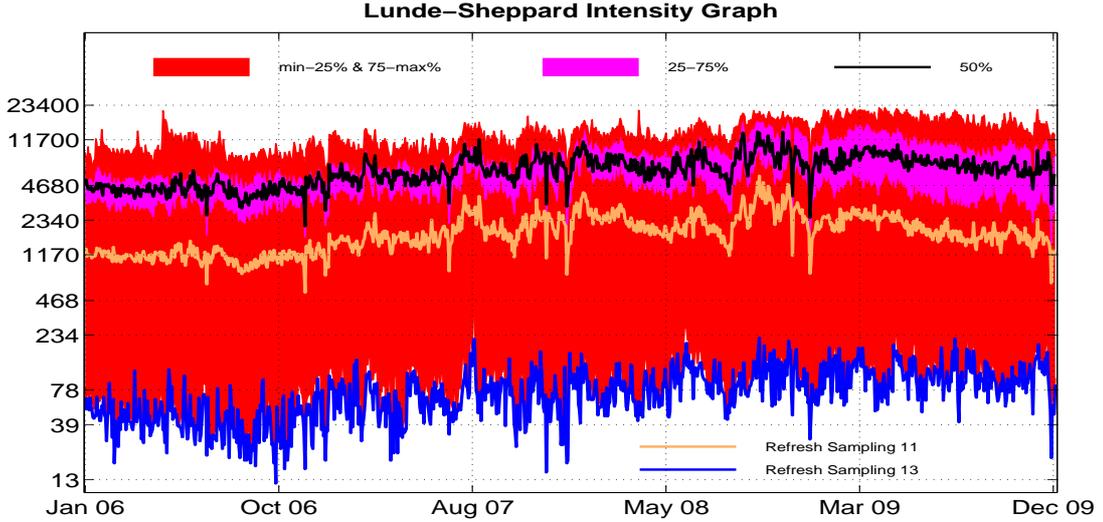


Figure 3: The Lunde-Sheppard trading intensity graph for our 13 assets. The figure plots the min, max, 25%, 50%, and 75% quantiles of the number of observations for the cleaned dataset. The number of refresh sampling for the 11 asset (RefT1) and 13 asset (RefT2) databases are also plotted.

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 the covariances.

6.2.2 Individual volatilities

Effect of infrequent assets Next we will focus on the estimation of the individual 11 daily volatilities $\sqrt{\Sigma_{ii}}$, $i = 1, 2, \dots, 11$ for the frequently traded stocks. We will carry this out three ways:

- using a univariate realised QML;
- from a multivariate realised QML based on 11 assets;
- from a multivariate realised QML based on 13 assets.

We will write these as QML_1 , QML_{11} and QML_{13} . We might worry that the infrequently traded stocks will upset the estimates of the volatility of the frequently stocks, so QML_1 and QML_{11} will differ a great deal from QML_{13} . Is that worry justified?

Figure 4 gives a time series plot of the cross-sectional of the average daily volatilities for the three methods, where the cross-sectional averaging is across the 11 stocks. It is difficult to see much difference, which is encouraging. The corresponding realised kernel results, denoted RK_{13} show quite a lot of variability as we would expect from such a small sample size.

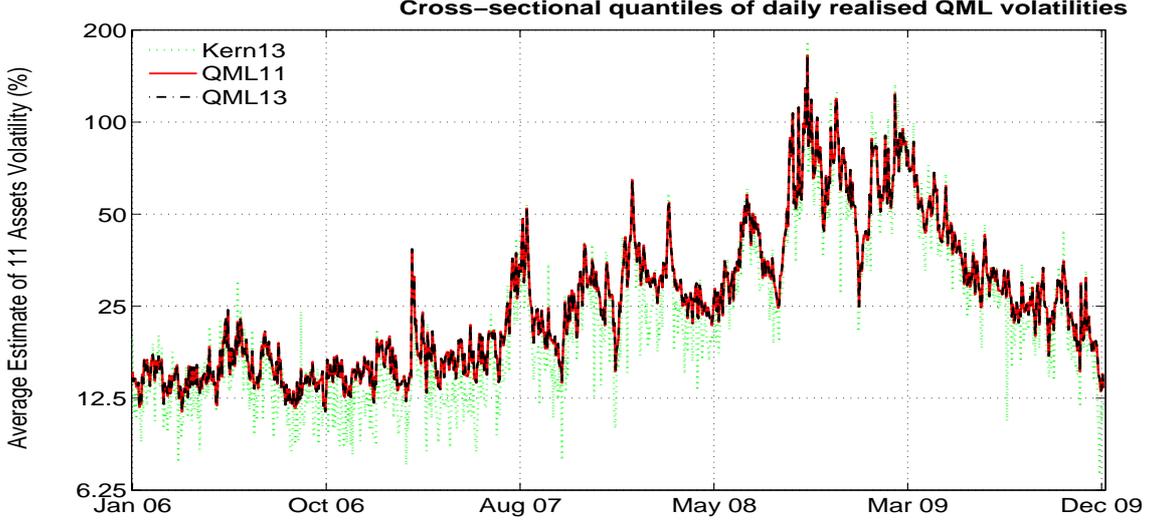


Figure 4: The figure plots the average volatility across 11 liquid assets based on three different estimators, including QML using 11 assets, QML using 13 assets and Multivariate Kernel using refresh time with all 13 assets.

Table 3 looks closer at this issue. It shows for the temporal average value of the percentage absolute differences between QML_1 and QML_{11} , QML_1 and QML_{13} and QML_{11} and QML_{13} :

$$Q_{1,11} = 100 \frac{|QML_1 - QML_{11}|}{QML_1}, \quad Q_{1,13} = 100 \frac{|QML_1 - QML_{13}|}{QML_1},$$

$$Q_{11,13} = 100 \frac{|QML_{11} - QML_{13}|}{QML_{11}}$$

for all of the individual 11 stocks, and $Q_{1,13}$ for the two infrequently traded stocks. We find that the proportional differences are larger when moving from 1 asset to 11 assets, since more useful information is added. However, moving to 13 stocks does not change the accuracy for the 11 stocks very much, as the newly added assets are highly illiquid and may provide little information for the variance of other liquid assets.

	Frequently traded											Infrequent	
	AA	AXP	BAC	DD	KO	GE	IBM	JPM	MSFT	SPY	XOM	BRK-B	WPO
$Q_{1,11}$	5.44	7.90	6.10	8.48	5.81	7.51	6.94	5.88	6.35	8.10	7.55		
$Q_{1,13}$	5.44	7.89	6.10	8.46	5.81	7.49	6.93	5.86	6.36	8.09	7.55	3.08	5.46
$Q_{11,13}$	0.006	0.015	0.002	0.026	0.001	0.018	0.011	0.017	0.001	0.003	0.007		
$RK_{1,11}$	0.92	1.38	2.73	1.38	5.89	2.11	1.42	4.53	6.69	0.55	1.95		
$RK_{1,13}$	13.51	12.50	14.26	15.26	19.20	17.04	13.57	19.48	20.22	8.62	14.14	7.58	1.49
$RK_{11,13}$	12.71	11.28	11.85	14.08	14.14	15.25	12.32	15.67	14.50	8.11	12.43		

Table 3: Volatility estimation: temporal average percentage difference differences between QML_1 (univariate realised QML) and QML_{11} (multivariate realised QML using frequently traded assets), QML_1 and QML_{13} (multivariate methods using all data) and QML_{11} and QML_{13} . Large numbers are bad. The RK results are the corresponding results for realised kernels.

These results can be compared to those from a multivariate realised kernel, also reported in Table 3, which replaces all of the above by RK rather than QML. These suggest the volatilities from the multivariate realised kernel is more affected by the number of data involved in the calculation, as information from other assets may only affect the estimation through the bandwidth, due to the linear structure of this estimator. The bandwidth can hardly make any difference in small sample. Clearly, when one includes some infrequently traded assets, the deterioration in the results are quite severe.

6.2.3 Individual correlations

Validating the correlations As we are computing a new correlation measure it is important that it is validated. One way of doing is to assess the correlation by comparing it to the less efficient but consistent multivariate realised kernel. The result is given in Figure 5. This shows the time series of the cross-sectional average daily correlations in the QML_{11} and QML_{13} cases, as well as the corresponding realised kernel.

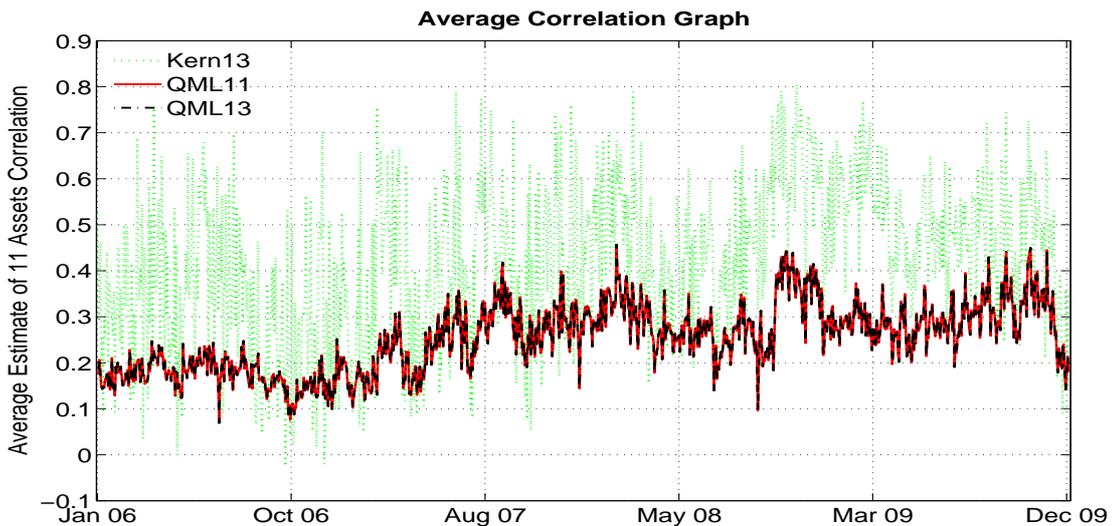


Figure 5: The figure plots the average correlation across 11 liquid assets based on three different estimators, including QML using 11 assets, QML using 13 assets and Multivariate Kernel using refresh time with all 13 assets.

A concern here is that the general level of the QML estimators is a little lower than the corresponding realised kernel, which may be due to unmodelled leads and lags in the returns. A strong feature here is that the QML_{11} and QML_{13} estimators are very similar, so the typical results are not upset by the presence of some infrequently traded stocks.

Other material Figure 4 and Figure 5 show a time series of the QML daily estimates of the volatility and correlations between 11 liquid assets, computed with all the series all at once and 11 liquid series respectively. The results indicates that the inclusion of illiquid assets has little impact on the estimation of the rest 11 liquid assets, while the kernel estimates using refresh time based on all assets suffer from a huge information loss, leading to inaccurate estimation.

Figure 6 shows a time series of the QML daily estimates of the correlations $\rho_{i,j}$ between the assets, computed with all the series all at once. Of course such a time series is quite noisy, but it shows a clumping of the correlations during the financial crisis. The heavy line through the middle shows the daily cross-sectional median of all the correlations.

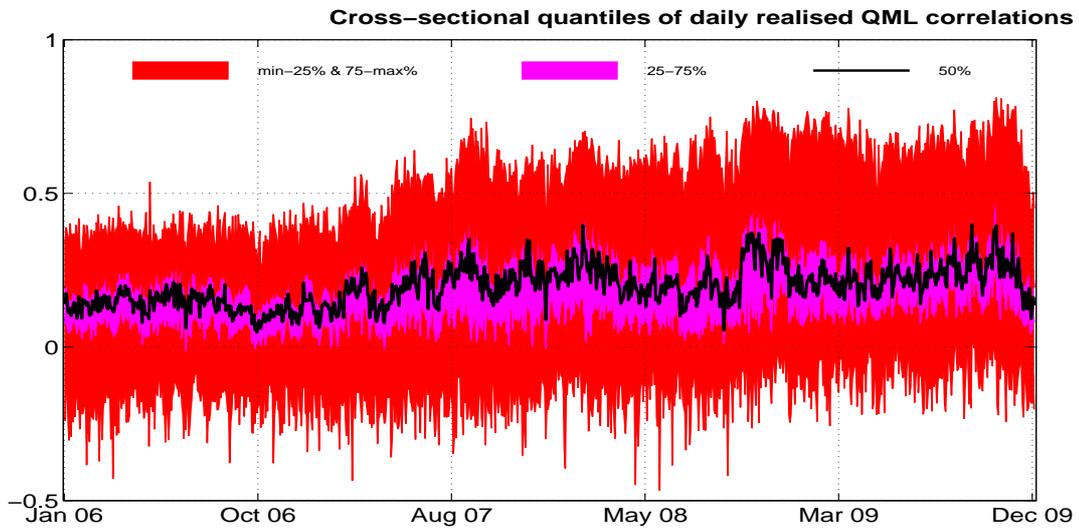


Figure 6: The figure plots the min, max, 25%, 50%, and 75% quantiles of all correlations.

Table 4 shows the unconditional covariance of open to close daily returns and the unconditional mean of the QML estimator, averaged over the 4 years of data.

These results are very preliminary but they indicate the method is underestimating the unconditional covariance to an important degree. More work needs to be done to calibrate if this is due to the i.i.d. assumption or some numerical issues. If it is the i.i.d. assumption then the use of sparse QML or subsampling QML may improve the situation.

6.3 A string of bivariate examples

Finally, we compare the QML with polarisation QML and realised kernel for 3 pairs of assets. These correlation estimates seem in agreement very well for the JPM and SPY pair, but less so for the pair of BAC and SPY.

	AA	AXP	BAC	BRKB	DD	GE	IBM	JPM	KO	MSFT	SPY	WPO	XOM
AA		0.042	0.042	0.004	0.046	0.029	0.025	0.039	0.018	0.024	0.029	0.012	0.031
AXP	0.105		0.065	0.003	0.041	0.032	0.031	0.068	0.022	0.027	0.036	0.013	0.032
BAC	0.132	0.209		0.004	0.036	0.034	0.026	0.078	0.018	0.026	0.033	0.011	0.030
BRKB	0.043	0.049	0.069		0.003	0.003	0.001	0.002	0.001	0.002	0.002	0.003	0.002
DD	0.085	0.088	0.096	0.034		0.025	0.025	0.038	0.019	0.022	0.030	0.012	0.029
GE	0.080	0.096	0.147	0.043	0.058		0.018	0.030	0.015	0.019	0.022	0.009	0.021
IBM	0.050	0.054	0.061	0.019	0.040	0.039		0.029	0.016	0.019	0.023	0.007	0.021
JPM	0.097	0.172	0.260	0.065	0.088	0.114	0.063		0.021	0.026	0.036	0.010	0.032
KO	0.037	0.039	0.038	0.012	0.028	0.027	0.021	0.037		0.013	0.016	0.006	0.017
MSFT	0.059	0.065	0.068	0.022	0.044	0.040	0.035	0.065	0.028		0.020	0.008	0.018
SPY	0.066	0.077	0.093	0.027	0.051	0.052	0.035	0.081	0.027	0.041		0.009	0.026
WPO	0.047	0.060	0.055	0.021	0.045	0.041	0.027	0.055	0.020	0.030	0.033		0.008
XOM	0.069	0.062	0.065	0.019	0.046	0.044	0.032	0.057	0.028	0.041	0.044	0.030	

Table 4: This table compares the QML estimate (upper triangle) of average of annualized covariances among 13 assets with the estimate based on daily open-to-close returns (lower triangle).

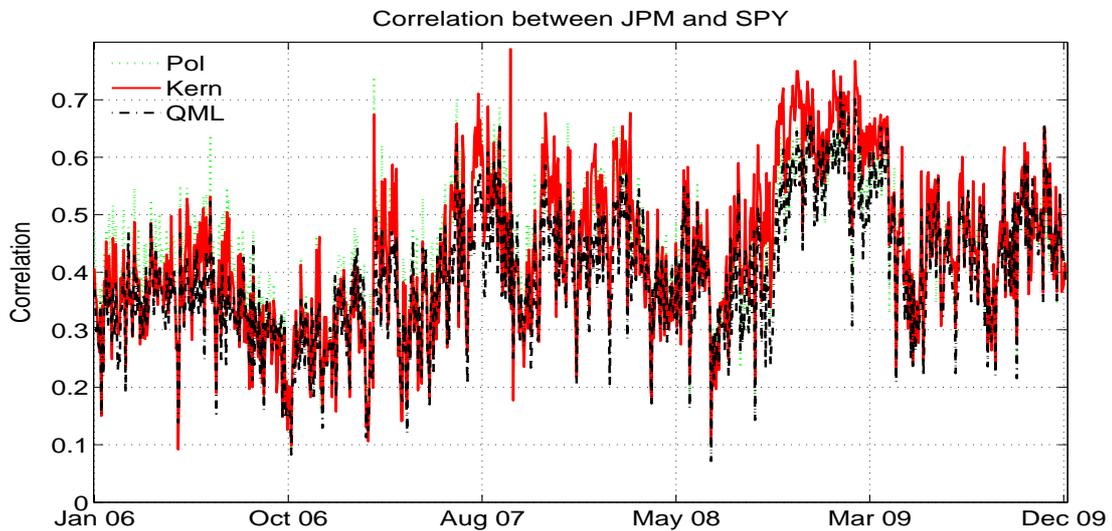


Figure 7: The figure plots the time series estimates of the correlation between JPM and SPY

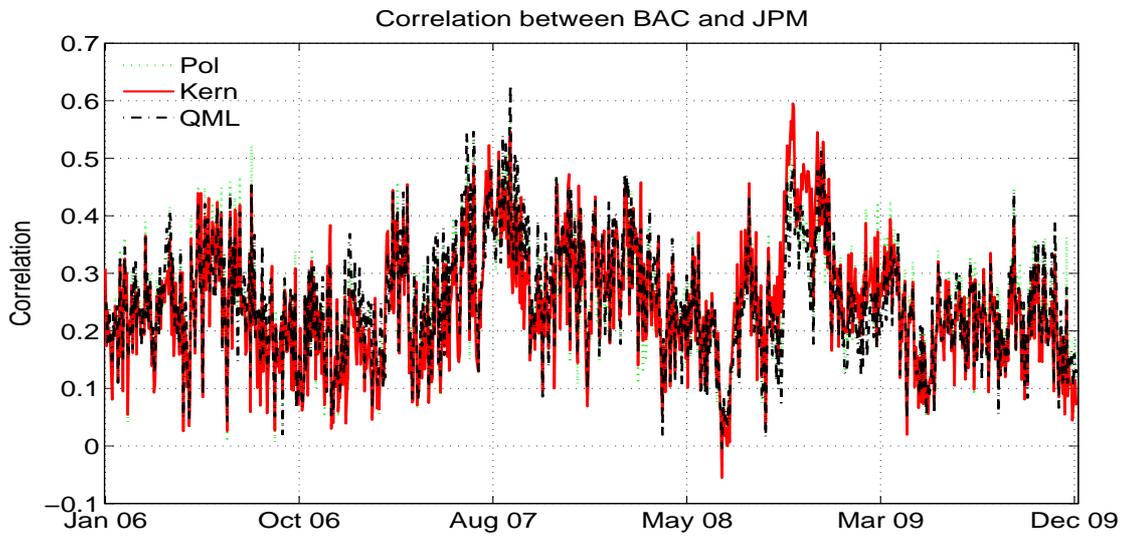


Figure 8: The figure plots the time series estimates of the correlation between BAC and JPM

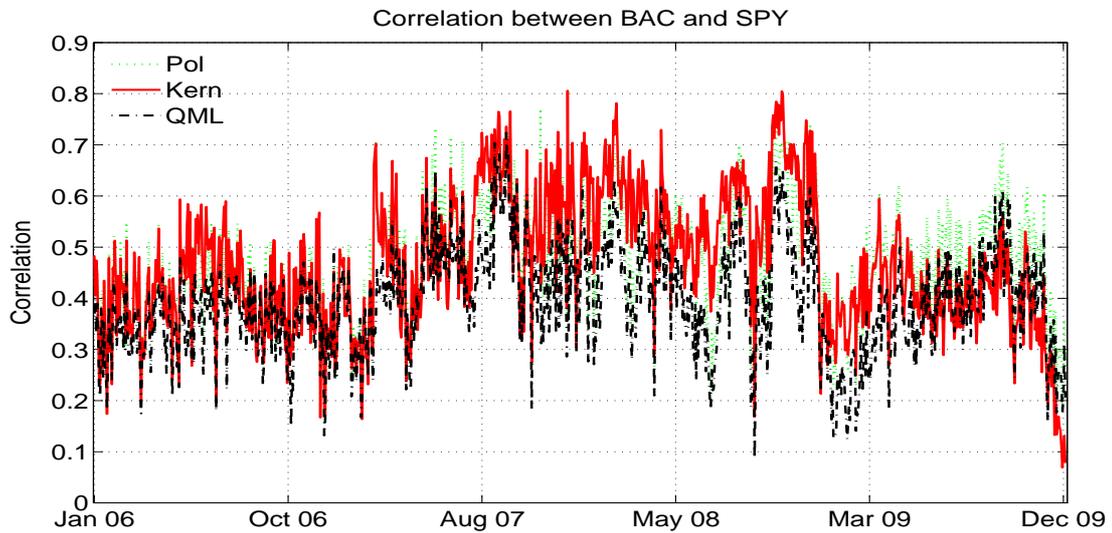


Figure 9: The figure plots the time series estimates of the correlation between BAC and SPY

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. 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 for this to be as polished as the rest of this paper. The volatilities seem to be robust to the presence of slowly trading stocks in the dataset. The results for measures for dependence are mixed, for some individual results show rather sensible time series of correlation estimates, but generally the realised QML is currently underestimating long-run dependence in these empirical experiments. This underestimation does not appear in our Monte Carlo experiments. There are various explanations for this, but we need to investigate this further and we will report on this in a later revision during the summer.

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 Pelusoy 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. Finally, we would like to thank Siem Jan Koopman for various pieces of advice on state space computations.

References

- Aït-Sahalia, Y., J. Fan, and D. Xiu (2010). High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105, 1504–1517.
- Aït-Sahalia, Y. and P. A. Mykland (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71, 483–549.

- Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18, 351–416.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000). Great realizations. *Risk* 13, 105–108.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Bandi, F. M. and J. R. Russell (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339–369.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal* 12, C1–C32.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.
- Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.
- Basu, S. and G. C. Reinsel (1996). Relationship between missing data likelihoods and complete data restricted likelihoods for regression time series models: an application to total ozone data. *Applied Statistics* 45, 63–72.
- Cappe, O., E. Moulines, and T. Ryden (2009). *Inference in Hidden Markov Models*. Springer.
- Christensen, K., S. Kinnebrock, and M. Podolskij (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics* 159, 116–133.
- Corsi, F., S. Peluso, and F. Audrino (2012). Missing asynchronicity: a Kalman-EM approach to multivariate realized covariance estimation. Unpublished paper: University of St. Gallen.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- Engle, R. F. and J. R. Russell (1998). Forecasting transaction rates: the autoregressive conditional duration model. *Econometrica* 66, 1127–1162.
- Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* 74, 291–296.
- Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.
- Gloter, A. and J. Jacod (2001a). Diffusions with measurement errors. I — local asymptotic normality. *ESAIM: Probability and Statistics* 5, 225–242.
- Gloter, A. and J. Jacod (2001b). Diffusions with measurement errors. II — measurement errors. *ESAIM: Probability and Statistics* 5, 243–260.
- Hansen, P. R. and G. Horel (2009). Quadratic variation by Markov chains. Unpublished paper: Department of Economics, Stanford University.
- Hansen, P. R., J. Large, and A. Lunde (2008). Moving average-based estimators of integrated variance. *Econometric Reviews* 27, 79–111.
- Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics* 24, 127–218.
- Harris, F. H. d., T. H. McMcInish, G. Shoesmith, and R. A. Wood (1995). Cointegration, error correction and price discovery on informationally linked security markets. *Journal of Financial and Quantitative Analysis* 30, 563–579.

- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.
- Jacod, J. (2012). Statistics and high frequency data. In M. Kessler, A. Lindner, and M. Sorensen (Eds.), *Statistical methods for stochastic differential equations*, pp. 191–310. Chapman and Hall.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications* 119, 2249–2276.
- Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes* (2 ed.). Springer: Berlin.
- Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of correlated measurement error. *Journal of Econometrics* 147, 47–59.
- Kunitomo, N. and S. Sato (2009). Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise. Unpublished paper: Graduate School of Economics, University of Tokyo.
- Large, J. (2011). Estimating quadratic variation when quoted prices jump by a constant increment. *Journal of Econometrics* 160, 2–11.
- Li, Y. and P. Mykland (2007). Are volatility estimators robust to modelling assumptions? *Bernoulli* 13, 601–622.
- Li, Y., P. Mykland, E. Renault, L. Zhang, and X. Zheng (2009). Realized volatility when endogeneity of time matters. Working Paper, Department of Statistics, University of Chicago.
- Liu, C. and C. Y. Tang (2012). A quasi-maximum likelihood approach to covariance matrix with high frequency data. Unpublished paper: Department of Statistics and Applied Probability, National University of Singapore.
- Ljung, G. M. (1989). A note on the estimation of missing values in time series. *Communications in Statistics - Simulation and Computation* 18, 459–465.
- Lunde, A., N. Shephard, and K. K. Sheppard (2012). Econometric analysis of vast covariance matrices using composite realized kernels. Unpublished paper: Department of Economics, University of Oxford.
- Martens, M. (2003). Estimating unbiased and precise realized covariances. Unpublished paper: Department of Finance, Erasmus School of Economics, Rotterdam.
- Mykland, P. A., N. Shephard, and K. K. Sheppard (2012). Efficient and feasible inference for the components of financial variation using blocked multipower variation. Unpublished paper: Department of Economics, Oxford University.
- Mykland, P. A. and L. Zhang (2006). ANOVA for diffusions and Ito processes. *Annals of Statistics* 34, 1931–1963.
- Mykland, P. A. and L. Zhang (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–1455.
- Protter, P. (2004). *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.
- Reiss, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Annals of Statistics* 39, 772–802.
- Stein, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association* 82, 765–772.
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions* (3 ed.). New York: Springer-Verlag.
- Voev, V. and A. Lunde (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics* 5, 68–104.
- West, M. and J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
- Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235–250.
- Zhang, L. (2011). Estimating covariation: Epps effect and microstructure noise. *Journal of Econometrics* 160, 33–47.

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

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

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

Appendices

A Mathematical Proofs

A.1 Proof of Theorem 1

There exists an orthogonal matrix $U = (u_{ij})$, 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

$$\Omega_{12} = \Sigma_{12}\Delta \otimes I,$$

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

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

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

Since $U' = U^{-1}$, we have

$$\Omega^{-1} = \begin{pmatrix} U' & \\ & U' \end{pmatrix} V^{-1} \begin{pmatrix} U & \\ & U \end{pmatrix}$$

where

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

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

Note that

$$\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \theta} = -\frac{1}{2\sqrt{n}} \left(\text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \theta} \right) - \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \theta} \Omega^{-1} z z' \right) \right)$$

where for $\theta = \Sigma_{11}$ and Σ_{12} ,

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

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

and

$$\begin{aligned} & \text{tr}\left(\Omega^{-1}\frac{\partial\Omega}{\partial\Sigma_{11}}\Omega^{-1}zz'\right) \\ &= \text{tr}\left(\begin{pmatrix} \text{diag}\left(\frac{\mu_{2i}^2\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) & \text{diag}\left(\frac{-\Sigma_{12}\mu_{2i}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) \\ \text{diag}\left(\frac{-\Sigma_{12}\mu_{2i}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) & \text{diag}\left(\frac{\Sigma_{12}^2\Delta^3}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) \end{pmatrix} \begin{pmatrix} U & \\ & U \end{pmatrix} zz' \begin{pmatrix} U' & \\ & U' \end{pmatrix}\right) \\ & \text{tr}\left(\Omega^{-1}\frac{\partial\Omega}{\partial\Sigma_{12}}\Omega^{-1}zz'\right) \\ &= \text{tr}\left(\begin{pmatrix} \text{diag}\left(\frac{-2\mu_{2i}\Sigma_{12}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) & \text{diag}\left(\frac{\Sigma_{12}^2\Delta^3 + \mu_{1i}\mu_{2i}\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) \\ \text{diag}\left(\frac{\Sigma_{12}^2\Delta^3 + \mu_{1i}\mu_{2i}\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) & \text{diag}\left(\frac{-2\mu_{1i}\Sigma_{12}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\right) \end{pmatrix} \begin{pmatrix} U & \\ & U \end{pmatrix} zz' \begin{pmatrix} U' & \\ & U' \end{pmatrix}\right) \end{aligned}$$

Therefore, by direct calculations and using symmetry, we have

$$\begin{aligned} E\left(\frac{-\partial^2 L}{\partial\Sigma_{11}^2}\right) &= \frac{1}{2} \sum_{i=1}^n \frac{\mu_{2i}^2\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}, \\ &\sim \frac{1}{2} \int_0^\infty \frac{(\Sigma_{22}T + \Lambda_{22}\pi^2x^2)^2T^2}{((\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2)^2} dx := I_{11}^\Sigma, \\ E\left(\frac{-\partial^2 L}{\partial\Sigma_{11}\partial\Sigma_{12}}\right) &= \frac{1}{2} \sum_{i=1}^n \frac{-2\Sigma_{12}\mu_{2i}\Delta^3}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}, \\ &\sim \int_0^\infty \frac{-\Sigma_{12}(\Sigma_{22}T + \Lambda_{22}\pi^2x^2)T^3}{((\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2)^2} dx := I_{12}^\Sigma, \\ E\left(\frac{-\partial^2 L}{\partial\Sigma_{11}\partial\Sigma_{22}}\right) &= \frac{1}{2} \sum_{i=1}^n \frac{\Sigma_{12}^2\Delta^4}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}, \\ &\sim \frac{1}{2} \int_0^\infty \frac{\Sigma_{12}^2T^4}{((\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2)^2} dx := I_{13}^\Sigma, \\ E\left(\frac{-\partial^2 L}{\partial\Sigma_{12}^2}\right) &= \frac{1}{2} \sum_{i=1}^n \frac{2(\Sigma_{12}^2\Delta^4 + \Delta^2\mu_{1i}\mu_{2i})}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}, \\ &\sim \int_0^\infty \frac{(\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2)T^2 + \Sigma_{12}^2T^4}{((\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2)^2} dx := I_{22}^\Sigma. \end{aligned}$$

By symmetry, we can obtain I_{23}^Σ , I_{33}^Σ , and I_{22}^Σ by simply switching the index 1 and 2 in I_{12}^Σ and I_{11}^Σ .

The asymptotic variance for $(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{12}, \widehat{\Sigma}_{22})$ is given by

$$V = \begin{pmatrix} I_{11}^\Sigma & I_{12}^\Sigma & I_{13}^\Sigma \\ \cdot & I_{22}^\Sigma & I_{23}^\Sigma \\ \cdot & \cdot & I_{33}^\Sigma \end{pmatrix}^{-1}$$

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

$$\begin{pmatrix} I_{11}^{\Lambda} & \\ & I_{22}^{\Lambda} \end{pmatrix}^{-1} = \begin{pmatrix} 2\Lambda_{11}^2 & \\ & 2\Lambda_{22}^2 \end{pmatrix}.$$

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 Σ, Λ , 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 (1), we have

$$L = -n \log(2\pi) - \frac{1}{2} \log(\det \Omega) - \frac{1}{2} z' \Omega^{-1} z$$

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,

$$\Omega_0 = \begin{pmatrix} \Omega_0^{11} & \Omega_0^{12} \\ \Omega_0^{21} & \Omega_0^{22} \end{pmatrix}$$

with $\Omega_{0,ii}^l = \int_{t_{i-1}}^{t_i} \Sigma_{l,t} dt + 2\Lambda_{0,l}$, $\Omega_{0,i,i+1}^l = \Omega_{0,i,i-1}^l = \Lambda_{0,l}$, and $\Omega_{0,i,i}^{12} = \Omega_{0,i,i}^{21} = \int_{t_{i-1}}^{t_i} \Sigma_{12,t} dt$.

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

$$L - \bar{L} = \frac{1}{2} \text{tr} \left(\Omega^{-1} (zz' - \Omega_0) \right) = \frac{1}{2} \text{tr} \left(\begin{pmatrix} U' & \\ & U' \end{pmatrix} V^{-1} \begin{pmatrix} U & \\ & U \end{pmatrix} (zz' - \Omega_0) \right)$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \omega_{ii}^{ls} \left((\Delta_i y_l) (\Delta_i y_s) - \int_{t_{i-1}}^{t_i} \Sigma_{ls,t} dt \right) + \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j<i}^n \omega_{ij}^{ls} \Delta_i^n y_l \Delta_j^n y_s \\
&\quad + \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}^{ls} \Delta_i^n \varepsilon_l \Delta_j^n y_s + \frac{1}{2} \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij}^{ls} \left(\Delta_i^n \varepsilon_l \Delta_j^n \varepsilon_s - E(\Delta_i^n \varepsilon_l \Delta_j^n \varepsilon_s) \right)
\end{aligned}$$

where ω_{ij}^{ls} is the (i, j) element of the (l, s) block of the matrix:

$$\Omega^{-1} = \begin{pmatrix} U' & \\ & U' \end{pmatrix} V^{-1} \begin{pmatrix} U & \\ & U \end{pmatrix}$$

Consider $\omega_{i,j}^{11}$ first.

$$\begin{aligned}
\omega_{i,j}^{11} &= \frac{2}{n+1} \sum_{k=1}^n \frac{\mu_{2k}}{\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2} \sin\left(\frac{ki}{n+1}\pi\right) \sin\left(\frac{kj}{n+1}\pi\right) \\
&= \frac{1}{n+1} \sum_{k=1}^n \frac{\mu_{2k}}{\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2} \left(\cos\left(\frac{k(i-j)}{n+1}\pi\right) - \cos\left(\frac{k(i+j)}{n+1}\pi\right) \right)
\end{aligned}$$

Therefore, we can separate the two components in the sum and analyze the following generic form:

$$\begin{aligned}
&\frac{1}{n+1} \sum_{k=1}^n \frac{\mu_{2k}}{\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2} \cos\left(\frac{kl}{n+1}\pi\right) \\
&= \frac{1}{2(n+1)} \sum_{k=1}^n \frac{\mu_{2k}}{\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2} \frac{\sin\left(\frac{(k+1)l}{n+1}\pi\right) - \sin\left(\frac{(k-1)l}{n+1}\pi\right)}{\sin\left(\frac{l}{n+1}\pi\right)} \\
&= \frac{1}{2(n+1)} \sum_{k=0}^n \frac{\sin\left(\frac{(k+1)l}{n+1}\pi\right)}{\sin\left(\frac{l}{n+1}\pi\right)} \left(\frac{\mu_{2k}}{\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2} - \frac{\mu_{2,k+2}}{\mu_{1,k+2}\mu_{2,k+2} - \Sigma_{12}^2 \Delta^2} \right) \\
&= \frac{1}{2(n+1)} \sum_{k=0}^n \frac{\sin\left(\frac{(k+1)l}{n+1}\pi\right)}{\sin\left(\frac{l}{n+1}\pi\right)} \frac{\mu_{2,k+2}\mu_{2,k}(\mu_{1,k+2} - \mu_{1,k}) + (\mu_{2,k+2} - \mu_{2,k})\Sigma_{12}^2 \Delta^2}{(\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2)(\mu_{1,k+2}\mu_{2,k+2} - \Sigma_{12}^2 \Delta^2)} \\
&= \frac{2}{n+1} \sum_{k=0}^n \frac{\sin\left(\frac{(k+1)l}{n+1}\pi\right) \sin\left(\frac{\pi}{n+1}\right) \sin\left(\frac{k+1}{n+1}\pi\right)}{\sin\left(\frac{l}{n+1}\pi\right)} \frac{(\mu_{2,k+2}\mu_{2,k}\Lambda_{11} + \Lambda_{22}\Sigma_{12}^2 \Delta^2)}{(\mu_{1k}\mu_{2k} - \Sigma_{12}^2 \Delta^2)(\mu_{1,k+2}\mu_{2,k+2} - \Sigma_{12}^2 \Delta^2)}
\end{aligned}$$

Clearly, $\omega_{i,j}^{11} = \omega_{j,i}^{11} = \omega_{n+1-i, n+1-j}^{11}$. For any $n^{\frac{1}{2}+\delta} \leq l \leq \lfloor \frac{n+1}{2} \rfloor$, we have

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

hence, for any $n^{\frac{1}{2}+\delta} \leq i \leq n - n^{\frac{1}{2}+\delta}$,

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

Similarly, we can derive

$$\begin{aligned}
\omega_{i,i}^{22} &= \left(\int_0^\infty \frac{\Sigma_{11}T + \Lambda_{11}\pi^2 x^2}{(\Sigma_{11}T + \Lambda_{11}\pi^2 x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2 x^2) - \Sigma_{12}^2 T^2} dx \right) \cdot \sqrt{n}(1 + o(1)) \\
\omega_{i,i}^{12} &= \left(\int_0^\infty \frac{-\Sigma_{12}T}{(\Sigma_{11}T + \Lambda_{11}\pi^2 x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2 x^2) - \Sigma_{12}^2 T^2} dx \right) \cdot \sqrt{n}(1 + o(1))
\end{aligned}$$

To simplify our notation, let

$$\begin{aligned}\omega^{11}(\Sigma, \Lambda, x) &= \frac{\Sigma_{22}T + \Lambda_{22}\pi^2x^2}{(\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2} \\ \omega^{22}(\Sigma, \Lambda, x) &= \frac{\Sigma_{11}T + \Lambda_{11}\pi^2x^2}{(\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2} \\ \omega^{12}(\Sigma, \Lambda, x) &= \frac{-\Sigma_{12}T}{(\Sigma_{11}T + \Lambda_{11}\pi^2x^2)(\Sigma_{22}T + \Lambda_{22}\pi^2x^2) - \Sigma_{12}^2T^2}\end{aligned}$$

We define the score vectors and their targets as

$$\Psi_\Sigma = -\frac{1}{\sqrt{n}} \frac{\partial L}{\partial \Sigma}, \bar{\Psi}_\Sigma = -\frac{1}{\sqrt{n}} \frac{\partial \bar{L}}{\partial \Sigma}, \Psi_\Lambda = -\frac{1}{n} \frac{\partial L}{\partial \Lambda}, \text{ and } \bar{\Psi}_\Lambda = -\frac{1}{n} \frac{\partial \bar{L}}{\partial \Lambda},$$

where

$$\frac{\partial}{\partial \Sigma} = \begin{pmatrix} \frac{\partial}{\partial \Sigma_{11}} \\ \frac{\partial}{\partial \Sigma_{12}} \\ \frac{\partial}{\partial \Sigma_{22}} \end{pmatrix}, \text{ and } \frac{\partial}{\partial \Lambda} = \begin{pmatrix} \frac{\partial}{\partial \Lambda_{11}} \\ \frac{\partial}{\partial \Lambda_{22}} \end{pmatrix}.$$

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

$$\begin{aligned}M_1^{(\Sigma)} &= \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \frac{\partial \omega_{ii}^{ls}}{\partial \Sigma} \left((\Delta_i y_l)(\Delta_i y_s) - \int_{\tau_{i-1}}^{\tau_i} \Sigma_{ls,t} dt \right) \\ M_2^{(\Sigma)} &= \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j < i} \frac{\partial \omega_{ij}^{ls}}{\partial \Sigma} \Delta_i^n y_l \Delta_j^n y_s \\ M_3^{(\Sigma)} &= \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \omega_{ij}^{ls}}{\partial \Sigma} \Delta_i^n \varepsilon_l \Delta_j^n y_s \\ M_4^{(\Sigma)} &= \sum_{l=1}^2 \sum_{s=1}^2 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \omega_{ij}^{ls}}{\partial \Sigma} \left(\Delta_i^n \varepsilon_l \Delta_j^n \varepsilon_s - E(\Delta_i^n \varepsilon_l \Delta_j^n \varepsilon_s) \right)\end{aligned}$$

Following the same argument in Xiu (2010) and Theorem 7.1 in Jacod (2012), we can show

$$n^{-\frac{1}{4}} (M_1^{(\Sigma)} + 2M_2^{(\Sigma)}) \xrightarrow{\mathcal{L}_X} MN(0, \text{Avar}^{(2)}(\Sigma))$$

where

$$\text{Avar}^{(2)}(\Sigma) = \lim_{n \rightarrow \infty} 4n^{-\frac{1}{2}} \sum_{l,s,u,v=1}^2 \sum_{i=1}^n \sum_{j < i} \frac{\partial \omega_{ij}^{v,u}}{\partial \Sigma} \frac{\partial \omega_{ij}^{l,s}}{\partial \Sigma'} \Sigma_{sv,t_i} \Sigma_{ul,t_j} \Delta^2 \quad (\text{A.1})$$

$$= 2T \sum_{l,s,u,v=1}^2 \int_0^\infty \frac{\partial \omega^{vu}(\Sigma, \Lambda, x)}{\partial \Sigma} \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Sigma'} dx \int_0^T \Sigma_{sv,t} \Sigma_{ul,t} dt \quad (\text{A.2})$$

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

$$\begin{aligned}
& \text{Avar}^{(3)}(\Sigma) \\
&= \lim_{n \rightarrow \infty} 4 \sum_{j=1}^n n^{-\frac{1}{2}} \sum_{l,s,v=1}^2 \sum_{i=1}^n \Lambda_{0,ul} \frac{\partial \omega_{ij}^{ls}}{\partial \Sigma} \left(2 \frac{\partial \omega_{ij}^{lv}}{\partial \Sigma'} - \frac{\partial \omega_{i,j-1}^{lv}}{\partial \Sigma'} - \frac{\partial \omega_{i,j+1}^{lv}}{\partial \Sigma'} \right) \Sigma_{sv,t_j} \Delta \\
&= 4 \sum_{l,s,v=1}^2 \Lambda_{0,ul} \int_0^\infty \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Sigma} \frac{\partial \omega^{lv}(\Sigma, \Lambda, x)}{\partial \Sigma'} \pi^2 x^2 dx \int_0^T \Sigma_{sv,t} dt
\end{aligned} \tag{A.3}$$

Finally, we have

$$n^{-\frac{1}{4}} M_4^{(\Sigma)} \xrightarrow{\mathcal{L}} N(0, \text{Avar}^{(4)}(\Sigma))$$

where

$$\begin{aligned}
& \text{Avar}^{(4)}(\Sigma) \\
&= \lim_{n \rightarrow \infty} n^{-\frac{1}{2}} \sum_{i,j,k,l=1}^n \left(\frac{\partial \omega_{i,j}^{11}}{\partial \Sigma} \frac{\partial \omega_{k,l}^{11}}{\partial \Sigma'} K_{11}^{ij,kl} + 4 \frac{\partial \omega_{i,j}^{12}}{\partial \Sigma} \frac{\partial \omega_{k,l}^{12}}{\partial \Sigma'} K_{11}^{i,k} K_{22}^{j,l} + \frac{\partial \omega_{i,j}^{22}}{\partial \Sigma} \frac{\partial \omega_{k,l}^{22}}{\partial \Sigma'} K_{22}^{ij,kl} \right) \\
&= \lim_{n \rightarrow \infty} n^{-\frac{1}{2}} \left(V_1 \left(\frac{\partial \omega^{11}}{\partial \Sigma}, \frac{\partial \omega^{11}}{\partial \Sigma'} \right) + V_2 \left(\frac{\partial \omega^{11}}{\partial \Sigma}, \frac{\partial \omega^{11}}{\partial \Sigma'} \right) + 2V_2 \left(\frac{\partial \omega^{12}}{\partial \Sigma}, \frac{\partial \omega^{12}}{\partial \Sigma'} \right) \right. \\
&\quad \left. + V_1 \left(\frac{\partial \omega^{22}}{\partial \Sigma}, \frac{\partial \omega^{22}}{\partial \Sigma'} \right) + V_2 \left(\frac{\partial \omega^{22}}{\partial \Sigma}, \frac{\partial \omega^{22}}{\partial \Sigma'} \right) \right),
\end{aligned}$$

and

$$\begin{aligned}
V_1 \left(\frac{\partial \omega^{ll}}{\partial \Sigma}, \frac{\partial \omega^{ll}}{\partial \Sigma'} \right) &= \left(\sum_{i=1}^{n-1} \left(-8 \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i+1,i+1}^{ll}}{\partial \Sigma'} + 2 \frac{\partial \omega_{i,i}^{ll}}{\partial \Sigma} \frac{\partial \omega_{i+1,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) \right. \\
&\quad \left. + 2 \sum_{i=1}^n \left(\frac{\partial \omega_{ii}^{ll}}{\partial \Sigma} \right) \left(\frac{\partial \omega_{ii}^{ll}}{\partial \Sigma'} \right) \right) \text{cum}_4[\varepsilon_l] \\
&\sim O(1) \\
V_2 \left(\frac{\partial \omega^{ll}}{\partial \Sigma}, \frac{\partial \omega^{ll}}{\partial \Sigma'} \right) &= 2(\Lambda_{0,ul})^2 \sum_{i,j=1}^n \left(\frac{\partial \omega_{i,j}^{ll}}{\partial \Sigma} \left(\frac{\partial \omega_{j-1,i-1}^{ll}}{\partial \Sigma'} + \frac{\partial \omega_{j-1,i+1}^{ll}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j-1,i}^{ll}}{\partial \Sigma'} + \frac{\partial \omega_{j+1,i-1}^{ll}}{\partial \Sigma'} \right. \right. \\
&\quad \left. \left. + \frac{\partial \omega_{j+1,i+1}^{ll}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j+1,i}^{ll}}{\partial \Sigma'} - 2 \left(\frac{\partial \omega_{j,i-1}^{ll}}{\partial \Sigma'} + \frac{\partial \omega_{j,i+1}^{ll}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j,i}^{ll}}{\partial \Sigma'} \right) \right) \right) \\
&\sim 2(\Lambda_{0,ul})^2 \left(\int_0^\infty \frac{\partial \omega^{ll}}{\partial \Sigma} \frac{\partial \omega^{ll}}{\partial \Sigma'} \pi^4 x^4 dx \right) n^{\frac{1}{2}} \\
V_2 \left(\frac{\partial \omega^{12}}{\partial \Sigma}, \frac{\partial \omega^{12}}{\partial \Sigma'} \right) &= 2(\Lambda_{0,11})(\Lambda_{0,22}) \sum_{i,j=1}^n \left(\frac{\partial \omega_{i,j}^{12}}{\partial \Sigma} \left(\frac{\partial \omega_{j-1,i-1}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{j-1,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j-1,i}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{j+1,i-1}^{12}}{\partial \Sigma'} \right. \right. \\
&\quad \left. \left. + \frac{\partial \omega_{j+1,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j+1,i}^{12}}{\partial \Sigma'} - 2 \left(\frac{\partial \omega_{j,i-1}^{12}}{\partial \Sigma'} + \frac{\partial \omega_{j,i+1}^{12}}{\partial \Sigma'} - 2 \frac{\partial \omega_{j,i}^{12}}{\partial \Sigma'} \right) \right) \right) \\
&\sim 2(\Lambda_{0,11})(\Lambda_{0,22}) \left(\int_0^\infty \frac{\partial \omega^{12}}{\partial \Sigma} \frac{\partial \omega^{12}}{\partial \Sigma'} \pi^4 x^4 dx \right) n^{\frac{1}{2}}
\end{aligned}$$

Here, $K_{11}^{i,j}, K_{22}^{i,j}, K_{11}^{ij,kl}$ and $K_{22}^{ij,kl}$ are the corresponding cumulants for $\Delta_i^n \varepsilon_1$ and $\Delta_i^n \varepsilon_2$, and $\text{cum}_4[\varepsilon_1]$ and $\text{cum}_4[\varepsilon_2]$ are the fourth cumulants of ε_1 and ε_2 .

Therefore, we have

$$\begin{aligned} & \text{Avar}^{(4)}(\Sigma) \\ &= 2 \sum_{l,s=1}^2 \Lambda_{0,ll} \Lambda_{0,ss} \int_0^\infty \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Sigma} \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Sigma'} \pi^4 x^4 dx \end{aligned} \quad (\text{A.4})$$

In summary, we have

$$n^{-\frac{1}{4}}(\Psi_\Sigma - \bar{\Psi}_\Sigma) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{4} \left(\text{Avar}^{(2)}(\Sigma) + \text{Avar}^{(3)}(\Sigma) + \text{Avar}^{(4)}(\Sigma) \right)\right). \quad (\text{A.5})$$

Similarly, we can obtain

$$n^{-\frac{1}{2}} \begin{pmatrix} \Psi_{\Lambda_{11}} - \bar{\Psi}_{\Lambda_{11}} \\ \Psi_{\Lambda_{22}} - \bar{\Psi}_{\Lambda_{22}} \end{pmatrix} \xrightarrow{\mathcal{L}} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2(\Lambda_{0,11})^2 + \text{cum}_4[\varepsilon_1]}{4\Lambda_{11}^4} & \\ & \frac{2(\Lambda_{0,22})^2 + \text{cum}_4[\varepsilon_2]}{4\Lambda_{22}^4} \end{pmatrix} \right) \quad (\text{A.6})$$

Further, we need to solve $\bar{\Psi}_\Sigma = 0$ and $\bar{\Psi}_\Lambda = 0$ for the pseudo-true parameters θ^* , and show that the distance between θ^* and the values of interest are negligible asymptotically. In fact, for any $\Sigma_{uv} \in \{\Sigma_{11}, \Sigma_{12}, \Sigma_{22}\}$, we have

$$\begin{aligned} \bar{\Psi}_{\Sigma_{uv}} &= \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \right) + \frac{\partial \text{tr}(\Omega^{-1} \Omega_0)}{\partial \Sigma_{uv}} \right\} \\ &= \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \Sigma_{uv}} \right) + \frac{\partial \text{tr}(\Omega^{-1}(\Omega + J \otimes (\Lambda_0 - \Lambda) + \Gamma))}{\partial \Sigma_{uv}} \right\} \\ &= \frac{1}{2\sqrt{n}} \left\{ \text{tr} \left(\frac{\partial \Omega^{-1}}{\partial \Sigma_{uv}} J \otimes (\Lambda_0 - \Lambda) \right) + \text{tr} \left(\frac{\partial \Omega^{-1}}{\partial \Sigma_{uv}} \Gamma \right) \right\} \\ &= \frac{1}{2\sqrt{n}} \left\{ \sum_{l=1}^2 \sum_{i=1}^n \left(2 \frac{\partial \omega_{ii}^{ll}}{\partial \Sigma_{uv}} - \frac{\partial \omega_{i,i-1}^{ll}}{\partial \Sigma_{uv}} - \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Sigma_{uv}} \right) (\Lambda_{0,ll} - \Lambda_{ll}) + \sum_{l,s=1}^2 \sum_{i=1}^n \frac{\partial \omega_{ii}^{ls}}{\partial \Sigma_{uv}} \Gamma_{ii}^{ls} \right\} \\ &= \frac{1}{2} \left\{ \sum_{l,s=1}^2 \left(\int_0^\infty \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Sigma_{uv}} dx \right) \left(\int_0^T \Sigma_{ls,t} dt - \Sigma_{ls} T \right) (1 + o(1)) \right. \\ &\quad \left. + \sum_{l=1}^2 (\Lambda_{0,ll} - \Lambda_{ll}) \int_0^\infty \left(\frac{\partial \omega^{ll}(\Sigma, \Lambda, x)}{\partial \Sigma_{uv}} \right) \pi^2 x^2 dx \right\} \end{aligned}$$

where Λ_0 denotes the true covariance matrix of noise, Γ is block diagonal matrix, with $\Gamma_{ii}^{ls} = \int_{\tau_{i-1}}^{\tau_i} \Sigma_{ls,t} dt - \Sigma_{ls} \Delta$, 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

$$\begin{aligned} \bar{\Psi}_{\Lambda_{uu}} &= \frac{1}{2n} \left\{ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \right) + \frac{\partial \text{tr}(\Omega^{-1} \Omega_0)}{\partial \Lambda_{uu}} \right\} \\ &= \frac{1}{2n} \left\{ \text{tr} \left(\Omega^{-1} \frac{\partial \Omega}{\partial \Lambda_{uu}} \right) + \frac{\partial \text{tr}(\Omega^{-1}(\Omega + J \otimes (\Lambda_0 - \Lambda) + \Gamma))}{\partial \Lambda_{uu}} \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2n} \left\{ \text{tr} \left(\frac{\partial \Omega^{-1}}{\partial \Lambda_{uu}} J \otimes (\Lambda_0 - \Lambda) \right) + \text{tr} \left(\frac{\partial \Omega^{-1}}{\partial \Lambda_{uu}} \Gamma \right) \right\} \\
&= \frac{1}{2n} \left\{ \sum_{l=1}^2 \sum_{i=1}^n \left(2 \frac{\partial \omega_{ii}^{ll}}{\partial \Lambda_{uu}} - \frac{\partial \omega_{i,i-1}^{ll}}{\partial \Lambda_{uu}} - \frac{\partial \omega_{i,i+1}^{ll}}{\partial \Lambda_{uu}} \right) (\Lambda_{0,l} - \Lambda_{ll}) + \sum_{l,s=1}^2 \sum_{i=1}^n \frac{\partial \omega_{ii}^{ls}}{\partial \Lambda_{uu}} \Gamma_{ii}^{ls} \right\} \\
&= \frac{1}{2} \left\{ \frac{1}{\sqrt{n}} \sum_{l,s=1}^2 \left(\int_0^\infty \frac{\partial \omega^{ls}(\Sigma, \Lambda, x)}{\partial \Lambda_{uu}} dx \right) \left(\int_0^T \Sigma_{ls,t} dt - \Sigma_{ls} T \right) (1 + o(1)) \right. \\
&\quad \left. + \frac{1}{\sqrt{n}} \sum_{l=1}^2 \left(\int_0^\infty \frac{\partial \omega^{ll}(\Sigma, \Lambda, x)}{\partial \Lambda_{uu}} \pi^2 x^2 dx \right) (\Lambda_{0,l} - \Lambda_{ll}) - \sum_{l=1}^2 \frac{\delta_{lu}}{\Lambda_{ll}^2} (\Lambda_{0,l} - \Lambda_{ll}) \right\}
\end{aligned}$$

Therefore, solving for $\bar{\Sigma}$ and $\bar{\Lambda}$, we obtain:

$$\begin{aligned}
\bar{\Lambda}_{ll} &= \Lambda_{0,l} + \frac{\bar{\Lambda}_{ll}^2}{\sqrt{n}} \left\{ \sum_{l,s=1}^2 \left(\int_0^\infty \frac{\partial \omega^{ls}(\bar{\Sigma}, \bar{\Lambda}, x)}{\partial \Lambda_{uu}} dx \right) \left(\int_0^T \Sigma_{ls,t} dt - \bar{\Sigma}_{ls} T \right) \right. \\
&\quad \left. + \sum_{l=1}^2 \left(\int_0^\infty \frac{\partial \omega^{ll}(\bar{\Sigma}, \bar{\Lambda}, x)}{\partial \Lambda_{uu}} \pi^2 x^2 dx \right) (\Lambda_{0,l} - \bar{\Lambda}_{ll}) \right\} (1 + o_p(1)), \text{ for } l = 1, 2, \\
\bar{\Sigma}_{ls} &= \frac{1}{T} \int_0^T \Sigma_{ls,t} dt + O_p(n^{-\frac{1}{2}}) = \Sigma_{0,ls} + O_p(n^{-\frac{1}{2}}), \text{ for } l, s = 1, 2.
\end{aligned}$$

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

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

hence consistency is established.

To find the central limit theorem, we do the usual ‘‘sandwich’’ calculations. Denote,

$$\frac{\partial \bar{\Psi}_\Sigma}{\partial \Sigma} = \begin{pmatrix} \frac{\partial \bar{\Psi}_{\Sigma_{11}}}{\partial \Sigma_{11}} & \frac{\partial \bar{\Psi}_{\Sigma_{11}}}{\partial \Sigma_{12}} & \frac{\partial \bar{\Psi}_{\Sigma_{11}}}{\partial \Sigma_{22}} \\ \frac{\partial \bar{\Psi}_{\Sigma_{12}}}{\partial \Sigma_{11}} & \frac{\partial \bar{\Psi}_{\Sigma_{12}}}{\partial \Sigma_{12}} & \frac{\partial \bar{\Psi}_{\Sigma_{12}}}{\partial \Sigma_{22}} \\ \frac{\partial \bar{\Psi}_{\Sigma_{22}}}{\partial \Sigma_{11}} & \frac{\partial \bar{\Psi}_{\Sigma_{22}}}{\partial \Sigma_{12}} & \frac{\partial \bar{\Psi}_{\Sigma_{22}}}{\partial \Sigma_{22}} \end{pmatrix} \xrightarrow{\mathcal{P}} \frac{\partial \Psi_{\Sigma_0}}{\partial \Sigma},$$

where

$$\frac{\partial \bar{\Psi}_{\Sigma_{0,uv}}}{\partial \Sigma_{ij}} = -\frac{T}{2} \left(\int_0^\infty \frac{\partial \omega^{ij}(\Sigma_0, \Lambda_0, x)}{\partial \Sigma_{uv}} dx \right) - \frac{T}{2} \left(\int_0^\infty \frac{\partial \omega^{ij}(\Sigma_0, \Lambda_0, x)}{\partial \Sigma_{uv}} dx \right) 1_{\{u \neq v\}}$$

and Σ_0 denotes the true parameter value. So, the central limit theorem is:

$$n^{\frac{1}{4}}(\hat{\Sigma} - \Sigma_0) = n^{\frac{1}{4}} \begin{pmatrix} \hat{\Sigma}_{11} - \frac{1}{T} \int_0^T \Sigma_{11,t} dt \\ \hat{\Sigma}_{12} - \frac{1}{T} \int_0^T \Sigma_{12,t} dt \\ \hat{\Sigma}_{22} - \frac{1}{T} \int_0^T \Sigma_{22,t} dt \end{pmatrix} \xrightarrow{\mathcal{L}_X} MN(0, V_Q) \quad (\text{A.7})$$

where

$$V_Q = \frac{1}{4} \left(\frac{\partial \bar{\Psi}_{\Sigma_0}}{\partial \Sigma} \right)^{-1} \left(\text{Avar}^{(2)}(\Sigma_0) + \text{Avar}^{(3)}(\Sigma_0) + \text{Avar}^{(4)}(\Sigma_0) \right) \left(\left(\frac{\partial \bar{\Psi}_{\Sigma_0}}{\partial \Sigma} \right)^{-1} \right)'. \quad (\text{A.8})$$

Note that

$$\frac{\partial \Psi_{\Lambda_0}}{\partial \Lambda} = \begin{pmatrix} -\frac{1}{2\Lambda_{0,11}^2} & \\ & -\frac{1}{2\Lambda_{0,22}^2} \end{pmatrix},$$

hence the CLT for $\hat{\Lambda}$ follows immediately from (A.6). This concludes the proof of Theorem 2.

A.3 Proof of Corollary 1

Denote $\Delta_i = \bar{\Delta}(1 + \xi_i)$ and ξ_i is $\overset{i.i.d}{\sim} O_p(1)$. Note that

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

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

$$\begin{aligned} \Omega^{-1} &= (\bar{\Omega}(I + \bar{\Delta}\bar{\Omega}^{-1}\Sigma \otimes \Xi))^{-1} = (I + \bar{\Delta}\bar{\Omega}^{-1}\Sigma \otimes \Xi)^{-1}\bar{\Omega}^{-1} \\ &= \bar{\Omega}^{-1} + \sum_{k=1}^{\infty} (-1)^k \bar{\Delta}^k (\bar{\Omega}^{-1}\Sigma \otimes \Xi)^k \bar{\Omega}^{-1} \end{aligned}$$

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,

$$\begin{aligned} E\left(\frac{\partial \log(\det \Omega)}{\partial \theta_1}\right) &= E\left(\text{tr}\left(\Omega^{-1} \frac{\partial \Omega}{\partial \theta_1}\right)\right) \\ &= E\left(\text{tr}\left(\bar{\Omega}^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1}\right)\right) + \bar{\Delta} E\left(\text{tr}\left(\bar{\Omega}^{-1}\Sigma \otimes \Xi \bar{\Omega}^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} + \bar{\Omega}^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi\right)\right) \\ &\quad + \bar{\Delta}^2 E\left(\text{tr}\left((\bar{\Omega}^{-1}\Sigma \otimes \Xi)^2 \bar{\Omega}^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \bar{\Omega}^{-1}\Sigma \otimes \Xi \bar{\Omega}^{-1} \frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi\right)\right) + o(\bar{\Delta}^2) \end{aligned}$$

Because $E(\Xi) = 0$,

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

Also,

$$\begin{aligned} &E\left(\text{tr}\left((\bar{\Omega}^{-1}\Sigma \otimes \Xi)^2 \bar{\Omega}^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \bar{\Omega}^{-1}(\Sigma \otimes \Xi) \bar{\Omega}^{-1} \left(\frac{\partial \Sigma}{\partial \theta_1} \otimes \Xi\right)\right)\right) \\ &= \text{tr}\left(\bar{\Omega}^{-1}(\Sigma \otimes I) D(\Sigma \otimes I) \bar{\Omega}^{-1} \frac{\partial \bar{\Omega}}{\partial \theta_1} - \bar{\Omega}^{-1}(\Sigma \otimes I) D\left(\frac{\partial \Sigma}{\partial \theta_1} \otimes I\right)\right) \text{var}(\xi) \end{aligned}$$

where

$$D = \begin{pmatrix} \text{diag}(\bar{\Omega}_{11}^{-1}) & \text{diag}(\bar{\Omega}_{12}^{-1}) \\ \text{diag}(\bar{\Omega}_{12}^{-1}) & \text{diag}(\bar{\Omega}_{22}^{-1}) \end{pmatrix}$$

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

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

where

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

In fact, we can show that

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

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