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

Neil Shephard
Department of Economics, Harvard University
Department of Statistics, Harvard University
shephard@fas.harvard.edu

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

First full draft: February 2012
This version: April 17, 2014

Abstract

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

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

1 Introduction

1.1 Core message

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

Our realised quasi maximum likelihood (QML) estimator of the covariance matrix of asset prices is positive semidefinite and deals with both market microstructure effects such as bid/ask bounce and crucially non-synchronous recording of data (the Epps (1979) effect). Positive semidefiniteness

allows us to define a coherent estimator of correlations and betas, objects of importance in financial economics. We derive the theoretical properties of our estimator and prove that it achieves the efficient rate of convergence. The estimator is also analysed using Monte Carlo methods and applied on equity data in a high dimensional case.

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

1.2 Quasi-likelihood context

Our approach naturally integrates three influential econometric estimators, completing a line of research and opening up many more areas of development and application.

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

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

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

Our paper moves beyond this work, producing distinctive and empirically important results. It proposes and analyses in detail the multivariate realised QML estimator which deals with irregularly spaced non-synchronous noisy data. Our methods are easily implemented and we develop the corresponding asymptotic theory under realistic assumptions.

1.3 Alternative approaches

A number of authors have approached this sophisticated problem using a variety of techniques. Here we place our work in this context. As we said above the first generation of multivariate estimators, realised covariances, were based upon moderately high frequency data. Introduced by Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2004), these

realised covariances use synchronised data sampled sufficiently sparsely that they could roughly ignore the effect of noise and non-synchronous trading. Related is Hayashi and Yoshida (2005) who tried to overcome non-synchronous trading but did not deal with any aspects of noise.

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 six existing methods which have been proposed. Only 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 a synchronisation device called refresh time sampling. Neither converges at the optimal rate.

Two other estimators have been suggested which rely on polarisation of quadratic variation. 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 efficiency bound even when covariance matrix is constant. Park and Linton (2012a) develop a Fourier based estimator of covariances, extending Mancino and Sanfelici (2011) and Sanfelici and Mancino (2008). Bibinger, Hautsch, Malec, and Reis (2014) proposes a local method of moments estimator that achieves the nonparametric efficiency bound, however the resulting estimator is not positive semi-definite.

Finally, we note that related univariate work on ameliorating the effect of noise includes 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)), Bandi and Russell (2008), Kalnina and Linton (2008), Li and Mykland (2007), Gloter and Jacod (2001a), Gloter and Jacod (2001b), Kunitomo and Sato (2013), Reiss (2011), Malliavin and Mancino (2002), Mancino and Sanfelici (2008), Malliavin and Mancino (2009), Aït-Sahalia, Jacod, and Li (2012) and Hansen and Horel (2009). Surveys include, for example, McAleer and Medeiros (2008), Park and Linton (2012b) and Aït-Sahalia and Xiu (2012).

1.4 More details on our paper

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

The above approach can be implemented computationally efficiently using Kalman filtering, optimising the likelihood via an EM algorithm. The resulting estimator of the integrated covariance is positive semidefinite. In practice it can be computed rapidly, even in significant dimensions.

1.5 Some particularly noteworthy papers

There are a group of papers which are closest to our approach. Aït-Sahalia, Fan, and Xiu (2010) apply the univariate estimator of Xiu (2010) to the multivariate case using polarisation. They estimate the covariance between x_1 and x_2, by applying univariate methods to estimate $\text{Var}(x_1+x_2)$ and $\text{Var}(x_1-x_2)$. The implied covariance matrix is not guaranteed to be positive semi-definite.

Corsi, Peluso, and Audrino (2014) 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: using a missing value approach based on Brownian motion observed with error. Further, Peluso, Corsi, and Mira (2012) extends this work using Bayesian techniques. They have no limiting theory for their quasi-likelihood based approach. Related to these papers is the earlier more informal univariate analysis of Owens and Steigerwald (2006) and the multivariate analysis of Cartea and Karyampas (2011).

More recently we also learnt of Liu and Tang (2014) who study a realised QML estimator of a multivariate exactly synchronised dataset. They propose using refresh time type devices to achieve exact synchronicity. Under exact synchronicity their theoretical development is significant and independently generates the results in one of the theorems in this paper. We will spell out the precise theoretical overlap with our paper later. To be explicit they do not deal with asynchronous trading, which is the major contribution of our paper.

1.6 Structure of the paper

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

2 Models

2.1 Notation

We consider a d-dimensional log-price process $x = (x_1, ..., x_d)'$. These prices are observed irregularly and non-synchronous over the interval $[0, T]$, where T is fixed and thought of as a single day. Prices

will be referred to as trades, but our methods also apply to quotes. We write the union of all times of trades as t_i, $1, 2, ..., n$, where we have ordered the distinct times so that $0 \leq t_1 < ... < t_i < ... < t_n \leq T$. Associated with each t_i is an asset selection matrix Z_i. Let the number of assets which trade at time t_i be d_i and so $1 \leq d_i \leq d$. Then Z_i is $d_i \times d$, full of zeros and ones where each row sums exactly to one. Unit elements in column k of Z_i shows the k-th asset traded at time t_i.

2.2 Efficient price

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

$$y(t) = \int_0^t \mu(u)\mathrm{d}u + \int_0^t \sigma(u)\mathrm{d}W(u), \quad \Sigma(t) = \sigma(t)\sigma(t)', \tag{1}$$

where μ is a vector of predictable locally bounded drifts, σ is a càdlàg volatility matrix process and W is a vector of independent Brownian motions. Then the ex-post covariation is

$$[y, y]_T = \int_0^T \Sigma(u)\mathrm{d}u, \quad \text{where} \quad [y, y]_T = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \{y(\tau_j) - y(\tau_{j-1})\} \{y(\tau_j) - y(\tau_{j-1})\}',$$

(e.g. Protter (2004, p. 66–77)) for any sequence of deterministic partitions $0 = \tau_0 < \tau_1 < ... < \tau_n = T$ with $\sup_j\{\tau_{j+1} - \tau_j\} \to 0$ for $n \to \infty$. Our interest is in estimating $[y, y]_T$ using x.

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

Throughout we see a blurred version of y, with our data being

$$x_i = Z_i y(t_i) + Z_i \varepsilon_i, \quad i = 1, 2, ..., n, \quad \mathrm{E}(\varepsilon_i) = 0, \quad \mathrm{Cov}(\varepsilon_i) = \Lambda,$$

where ε_i is a vector of potential market microstructure effects and Λ is diagonal[1].

2.3 A Gaussian quasi-likelihood

We proxy the Brownian semimartingale by Brownian motion, which is non-synchronously observed. This will be used to generate a quasi-likelihood. We model

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

[1]Corsi, Peluso, and Audrino (2014) use a slightly different approach. They update at time points Ti/n whether there is new data or not. They used 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 a price is entirely missing, the input for their Kalman innovations v_i is set to zero.

Then writing $\Sigma = \sigma\sigma'$, we have that $y(t_i) - y(t_{i-1}) \sim N(0, \Sigma(t_i - t_{i-1}))$, while all the non-overlapping innovations are independent. Throughout we will write

$$u_i = y(t_i) - y(t_{i-1}), \quad \Delta_i^n = t_i - t_{i-1} > 0.$$

At this point we assume, for a diagonal Λ, that $\varepsilon_i \overset{iid}{\sim} N(0, \Lambda)$. Then the time series of observations $x_{1:n} = (x_1, \ldots, x_n)'$ is a Gaussian state space model (e.g. Durbin and Koopman (2001)).

2.4 ML estimation via EM algorithm

We will develop a positive semidefinite estimator 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 (e.g. Durbin and Koopman (2001, Ch. 7.3.4)) to maximising the log-likelihood function is attractive.

Note that the complete log-likelihood is, writing $e_i = x_i - Z_i y(t_i)$, $u_i = y(t_i) - y(t_{i-1})$, $y_{1:n} = (y_1, \ldots, y_n)'$, and assuming $x_1 \sim N(\hat{y}_1, P_1)$ which is independent of (Σ, Λ),

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

Then the EM algorithm works with the

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

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

$$\mathrm{E}\left\{e_i'\left(Z_i \Lambda Z_i'\right)^{-1} e_i|x_{1:n}\right\} = tr\left\{\left(Z_i \Lambda Z_i'\right)^{-1}\mathrm{E}\left(e_i e_i'|x_{1:n}\right)\right\} = tr\left[\left(Z_i \Lambda Z_i'\right)^{-1}\left\{\widehat{e}_{i|n}\widehat{e}_{i|n}' + D_{i|n}\right\}\right],$$

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

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

Then the EM update is

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

Iterating these updates, the sequence of $(\widehat{\Sigma}, \widehat{\Lambda})$ converges to a maximum in the likelihood function. The Appendix details the efficient computation of $\widehat{e}_{i|n}$, $\widehat{u}_{i|n}$, $D_{i|n}$ and $N_{i|n}$ in $O(nd^3)$ calculations.

3 Econometric theory

Here we develop the asymptotic theory for the bivariate case. To get to the heart of the issues our analysis follows three steps. First we look at the bivariate ML estimator case where the volatility is fixed and there are equidistant data. Secondly we show how those results change when the noise is non-Gaussian and we have stochastic volatility effects, but still have equidistant observations. Thirdly and more realistically we discuss the impact of having unequally spaced data which is wrongly synchronised in the quasi-likelihood. Finally we calculate the impact of non-synchronised data on a fully non-synchronised quasi-likelihood.

3.1 Step one: benchmark bivariate MLE

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

Then we observe returns

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

with, we assume,

$$x_i = y(i/n) + \varepsilon_i, \quad i = 0, 1, 2, ..., n, \quad y(i/n) = y((i-1)/n) + \sqrt{T/n}u_i, \tag{2}$$

where

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \overset{i.i.d.}{\sim} N\begin{pmatrix} \Lambda & 0 \\ 0 & \Sigma \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}. \tag{3}$$

The observed returns are $r = (r_{1,1}, r_{1,2}, \ldots, r_{1,n}, r_{2,1}, \ldots, r_{2,n})'$, so the likelihood is

$$L = -n\log(2\pi) - \frac{1}{2}\log(\det\Omega) - \frac{1}{2}r'\Omega^{-1}r, \tag{4}$$

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

$$J_n = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{5}$$

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

Before we give the bivariate limit theory recall the univariate case (e.g. Stein (1987), Gloter and Jacod (2001a), Gloter and Jacod (2001b), Aït-Sahalia, Mykland, and Zhang (2005), Xiu (2010))

$$n^{\frac{1}{4}}\left(\widehat{\Sigma}_{11} - \Sigma_{11}\right) \xrightarrow{L} N\left(0, 8\Lambda_{11}^{1/2}\Sigma_{11}^{3/2}T^{-1/2}\right).$$

Thus $n^{1/4}$ is the optimal rate with noisy data. We now go onto the bivariate case.

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

$$n^{\frac{1}{4}} \begin{pmatrix} \widehat{\Sigma}_{11} - \Sigma_{11} \\ \widehat{\Sigma}_{12} - \Sigma_{12} \\ \widehat{\Sigma}_{22} - \Sigma_{22} \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Pi).$$

For a 3×1 *vector* $\Sigma_\theta = vech(\Sigma) = (\Sigma_{1,1}, \Sigma_{1,2}, \Sigma_{2,2})'$ *the* Π *matrix is such that* $\Pi^{-1} = \frac{\partial \Psi_\theta}{\partial \Sigma_{\theta'}}$ *with*

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

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

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

Proof. Given in the Appendix.

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

When $\Sigma_{12} = 0$, there is no "externality," i.e. the asymptotic variances for $\widehat{\Sigma}_{11}$ and $\widehat{\Sigma}_{22}$ in the bivariate case reproduce the univariate case. As the correlation increases from 0 to 1, $\widehat{\Sigma}_{11}$ becomes more efficient than the univariate one, as more information is collected via the correlation with the other series. This is illustrated in Figure 1. Finally, we note that Bibinger, Hautsch, Malec, and Reis (2014) have independently also developed the Fisher information matrix for this problem.

3.2 Step two: bivariate QMLE with equidistant observations

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

Assumption 1 *The underlying latent d-dimensional log-price process satisfies (1), where* μ *is predictable locally bounded, the* σ *is a locally bounded Itô semimartingale and* W *is a Brownian motion.*

Assumption 2 *The noise* ε_i *is i.i.d., and independent of* t_i, W, σ *and has fourth moments*[2].

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

[2]The i.i.d. assumption can be replaced by more general noise process, which is independent conditionally on Y. This allows some heteroskedasticity. This is the focus of, for example, Jacod, Podolskij, and Vetter (2010). We choose not to adopt it as the idea of the proof remains the same except for some technicalities.

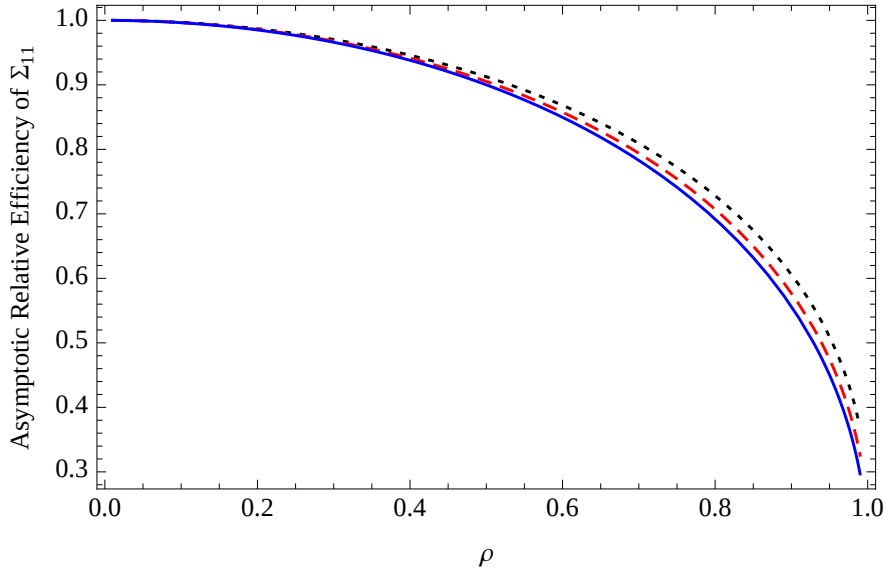

Figure 1: The figure plots the relative efficiency for bivariate MLE of Σ_{11} over the univariate alternative, against the correlation. $\Sigma_{11} = 0.25^2$. As the number falls below zero the gains from bivariate MLE become greater. The blue line, the red dashed line, and the black dotted line correspond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

Define $R_T = \left(\frac{1}{T}\int_0^T \sigma_t^4 \mathrm{d}t\right) / \left(\frac{1}{T}\int_0^T \sigma_t^2 \mathrm{d}t\right)^2 \geq 1$, by Jensen's inequality, and recall the result

$$n^{\frac{1}{4}}\left(\widehat{\Sigma}_{11} - \frac{1}{T}\int_0^T \sigma_t^2 \mathrm{d}t\right) \xrightarrow{\mathcal{L}s} MN\left(0, (5R_T + 3)\,\Lambda_{11}^{1/2}\left(\frac{1}{T}\int_0^T \sigma_t^2 \mathrm{d}t\right)^{3/2} T^{-1/2}\right),$$

by rewriting Xiu (2010). The asymptotic variance increases with R_T keeping $\frac{1}{T}\int_0^T \sigma_t^2 \mathrm{d}t$ fixed.

The asymptotic properties of the multivariate realised QML estimator $\widehat{\Sigma}$ is given below.

Theorem 2 (Bivariate QMLE) *Under Assumptions 1-3, we have*

$$n^{\frac{1}{4}}\begin{pmatrix} \widehat{\Sigma}_{11} - \frac{1}{T}\int_0^T \Sigma_{11,t}\mathrm{d}t \\ \widehat{\Sigma}_{12} - \frac{1}{T}\int_0^T \Sigma_{12,t}\mathrm{d}t \\ \widehat{\Sigma}_{22} - \frac{1}{T}\int_0^T \Sigma_{22,t}\mathrm{d}t \end{pmatrix} \xrightarrow{\mathcal{L}s} MN(0, \Pi_Q), \quad \textit{where} \tag{6}$$

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

Here all the derivatives are evaluated at $\Sigma = \frac{1}{T}\int_0^T \Sigma_t \mathrm{d}t$.

Proof. Given in the Appendix.

Liu and Tang (2014) have independent and concurrently established a similar result, although their derivation follows the steps in Xiu (2010), which is different from our approach.

3.3 Step 3: bivariate QMLE with irregularly space synchronised observations

We now build a quasi-likelihood upon irregularly spaced but synchronised times $\{t_1, t_2, \ldots, t_n\}$, so that $Z_i = I_2$, which imply a collection of increments $\{\Delta_i^n = t_i - t_{i-1}, 1 \leq i \leq n\}$.

Assumption 4 *Assume that* $\Delta_i^n = \bar{\Delta}(1 + \xi_i)$, $i = 1, 2, ..., n$, $\bar{\Delta} = \frac{T}{n}$, $\mathrm{E}(\xi_i) = 0$, *where* $\mathrm{Var}(\xi_i) < \infty$ *and* $\{\xi_i, 1 \leq i \leq n\}$ *are i.i.d.. Also assume* Y *and* $\{\xi_i\}$ *are independent.*

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

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

Proof. Given in the Appendix.

This makes it clear that in the synchronised case irregularly spacing of the data has no impact on the asymptotic distribution of the multivariate realised QML estimator. The reason for this is that the effect is less important than the presence of noise, a result which echoes the univariate work of Aït-Sahalia, Mykland, and Zhang (2005) and Aït-Sahalia and Mykland (2003), and recent multivariate result by Bibinger, Hautsch, Malec, and Reis (2014). This contrasts with the realised variance where Mykland and Zhang (2006) have shown that the quadratic variation of the sampling times impacts the asymptotic distribution in the absence of noise.

This synchronised case is important. Most researchers have analysed covariances by applying a synchronisation scheme to the non-synchronous data. 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 Aït-Sahalia, Fan, and Xiu (2010). Alternatively, Zhang (2011) discusses the Previous Tick approach, which always discards more data than refresh time.

3.4 Step 4: bivariate QMLE with non-synchronous observations

In this paper, synchronisation is not needed. Instead we write $x_{j,i} = y_j(t_{j,i}) + \varepsilon_{j,i}$, $j = 1, 2$, $i = 1, 2, \ldots, n_j$, where $t_{j,i}$ is the i-th observation on the j-th asset. The returns are $r_{j,i} = x_{j,i} - x_{j,i-1}$.

Assumption 5 *Define* $\Delta_i^{n_j} = t_{j,i} - t_{j,i-1}$, $i = 1, 2, \ldots, n_j$ *and* $\Delta^{n_j} = diag(\Delta_i^{n_j})$. *Assume that* $\Delta_i^{n_j} = \bar{\Delta}_j$, *for* $j = 1, 2$, *and* $\bar{\Delta}_2 = m\bar{\Delta}_1$, *where* $\bar{\Delta}_j = T/n_j$.

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

Writing $r = (r_{1,1}, r_{1,2}, \ldots, r_{1,n_1}, r_{2,1}, \ldots, r_{2,n_2})'$, then the likelihood is

$$L = c - \frac{1}{2}\log(\det\Omega) - \frac{1}{2}r'\Omega^{-1}r, \quad \text{where} \quad \Omega = \begin{pmatrix} \Sigma_{11}\Delta^{n_1} + \Lambda_{11}J_{n_1} & \Sigma_{12}\Delta^{n_1,n_2} \\ \Sigma_{12}\Delta^{n_2,n_1} & \Sigma_{22}\Delta^{n_2} + \Lambda_{22}J_{n_2} \end{pmatrix}, \quad (7)$$

with

$$\Delta_{i,j}^{n_1,n_2} = \begin{cases} t_{2,j} - t_{1,i-1}, & \text{if} \quad t_{2,j-1} \leq t_{1,i-1} < t_{2,j} \leq t_{1,i}; \\ t_{1,i} - t_{2,j-1}, & \text{if} \quad t_{1,i-1} \leq t_{2,j-1} < t_{1,i} \leq t_{2,j}; \\ t_{1,i} - t_{2,i-1}, & \text{if} \quad t_{2,j-1} \leq t_{1,i-1} < t_{1,i} \leq t_{2,j}; \\ t_{1,j} - t_{2,j-1}, & \text{if} \quad t_{2,i-1} \leq t_{1,j-1} < t_{1,j} \leq t_{2,i}; \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Again J_n appears in (5). Note this likelihood function does not require Assumption 5.

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

$$\begin{pmatrix} n_1^{1/4}(\widehat{\Sigma}_{11} - \Sigma_{11}) \\ n_2^{1/4}(\widehat{\Sigma}_{12} - \Sigma_{12}) \\ n_2^{1/4}(\widehat{\Sigma}_{22} - \Sigma_{22}) \end{pmatrix} \xrightarrow{L} N(0, \Pi_A), \quad \textit{where} \quad \Pi_A = \begin{pmatrix} \frac{\partial\Psi_{\Sigma_{1,1}}}{\partial\Sigma_{1,1}} & 0 & 0 \\ 0 & \frac{\partial\Psi_{\Sigma_{1,2}}}{\partial\Sigma_{1,2}} & \frac{\partial\Psi_{\Sigma_{2,2}}}{\partial\Sigma_{1,2}} \\ 0 & \frac{\partial\Psi_{\Sigma_{2,2}}}{\partial\Sigma_{1,2}} & \frac{\partial\Psi_{\Sigma_{2,2}}}{\partial\Sigma_{2,2}} \end{pmatrix}^{-1},$$

such that

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

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

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

Proof. Given in the Appendix.

This Theorem suggests that including the extremely illiquid assets into estimation should not greatly affect variance and covariance estimates of the liquid ones. This property is distinct from any approach in the literature which uses a synchronization method.

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

3.4.1 Extension

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

$$\begin{pmatrix} n_1^{1/4}(\widehat{\Sigma}_{11} - \Sigma_{11}) \\ n_2^{1/4}(\widehat{\Sigma}_{12} - \Sigma_{12}) \\ n_2^{1/4}(\widehat{\Sigma}_{22} - \Sigma_{22}) \end{pmatrix} \xrightarrow{L} N(0, \Pi_{A,m}),$$

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

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

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

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

$$\begin{pmatrix} n_1^{1/4}(\widehat{\Sigma}_{11} - \Sigma_{11}) \\ n_2^{1/4}(\widehat{\Sigma}_{12} - \Sigma_{12}) \\ n_2^{1/4}(\widehat{\Sigma}_{22} - \Sigma_{22}) \end{pmatrix} \xrightarrow{\mathcal{L}s} MN(0, \Pi_{Q,m}), \tag{9}$$

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

4 Additional developments

4.1 Realised QML correlation and regression estimator

An immediate corollary for the realised QML estimator is the correlation version

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

Also of importance is the corresponding "realised QML beta"

$$\widehat{\beta}_{1|2} = \frac{\widehat{\Sigma}_{12}}{\widehat{\Sigma}_{22}}, \quad \text{which estimates} \quad \beta_{1|2} = \frac{\frac{1}{T}\int_0^T \Sigma_{12,t}\mathrm{d}t}{\frac{1}{T}\int_0^T \Sigma_{22,t}\mathrm{d}t}.$$

The limit theory follows by the delta method: $\mathrm{Avar}(\widehat{\rho}_{12}) = \nu_\rho V_Q \nu_\rho'$, and $\mathrm{Avar}(\widehat{\beta}_{1|2}) = \nu_\beta V_Q \nu_\beta'$,

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

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 Multistep realised QML estimator

There are robustness advantages in estimating the integrated variances using univariate QML methods $\widehat{\Sigma}_{11}$, $\widehat{\Sigma}_{22}$. These two estimates can then be combined with the QML correlation estimator $\widehat{\rho}_{12}$, obtained by maximising the likelihood with respect to ρ_{12} keeping Σ_{11}, Σ_{22} fixed at $\widehat{\Sigma}_{11}$, $\widehat{\Sigma}_{22}$. This is called the "multistep covariance estimator". An advantage of this is that model specification for one asset price will not impact the estimator of the integrated variance for the other asset.

4.3 Sparse and subsampled realised QML

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

To overcome this dependence we define a "sparse realised QML" estimator, which corresponds to the sparse sampling realised variance. The approach we have explored is as follows.

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

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

Once we have defined a sparse realised QML, it is obvious that we could also simply subsample this approach, which means constructing k sets of subsampled datasets and for each computing the corresponding quasi-likelihood. 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 employs all of the data while being less sensitive to the i.i.d. assumption.

5 Monte Carlo experiments

5.1 Monte Carlo design

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

$$\mathrm{d}y_{it} = \alpha_{it}\mathrm{d}t + \sigma_{it}\mathrm{d}W_{it}, \quad \mathrm{d}\sigma_{it}^2 = \kappa_i\left(\overline{\sigma}_i^2 - \sigma_{it}^2\right)\mathrm{d}t + s_i\sigma_{it}\mathrm{d}B_{it} + \sigma_{it-}J_{it}^V\,\mathrm{d}N_{it},$$

where $\mathrm{E}(\mathrm{d}W_{it}\mathrm{d}B_{jt}) = \delta_{ij}\rho_i\mathrm{d}t$ and $\mathrm{E}(\mathrm{d}W_{1t}\mathrm{d}W_{2t}|\rho^*) = \rho^*\mathrm{d}t$. Here $\kappa_i > 0$. Ignoring the jumps, the variance process $\sigma_{it}^2 \sim \Gamma(2\kappa_i\overline{\sigma}_i^2/s_i^2, s_i^2/2\kappa_i)$. Throughout when jumps happen the log-jumps $\log J_{it}^V \overset{iid}{\sim} N(\theta_i, \mu_i)$, while N_{it} is a Poisson process with intensity λ_i. Likewise $\varepsilon_{it} \overset{iid}{\sim} N(0, a_i^2)$.

We now depart slightly from their setup. For each day we draw $\sigma_{i0}^2 \overset{iid}{\sim} \Gamma(2\kappa_i\overline{\sigma}_i^2/s_i^2, s_i^2/2\kappa_i)$ over $i = 1, 2$, so each replication will be independent. For each separate day we simulate independently $\rho^* \sim \rho_0 Beta(\rho_1^*, \rho_2^*)$, where $\rho_0 = \sqrt{(1-\rho_1^2)(1-\rho_2^2)}$, guaranteeing the positive-definiteness of the covariance matrix of (W_1, W_2, B_1, B_2). This means $\mathrm{E}(\rho^*) = \rho_0\rho_1^*/\left(\rho_1^* + \rho_2^*\right)$ and $\mathrm{sd}(\rho^*) = \rho_0\sqrt{\rho_1^*\rho_2^*}/\left\{(\rho_1^* + \rho_2^*)\sqrt{\rho_1^* + \rho_2^* + 1}\right\}$. The values of a_i, α_i, ρ_i, κ_i, θ_i, μ_i, λ_i, s_i, $\overline{\sigma}_i^2$, ρ_1^* and ρ_2^* are given in Table 1. To check our limit theory calculations, Figure 2 plots the histograms

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

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

of the standardized pivotal statistics (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. The histograms show the limiting result provides a reasonable guide to the finite sample behavior in these cases.

In our main Monte Carlo we take $n \in \{117, 1170, 11700\}$ and all results are based on $1,000$ stochastically independent replications. Having fixed the overall sample size n we randomly and

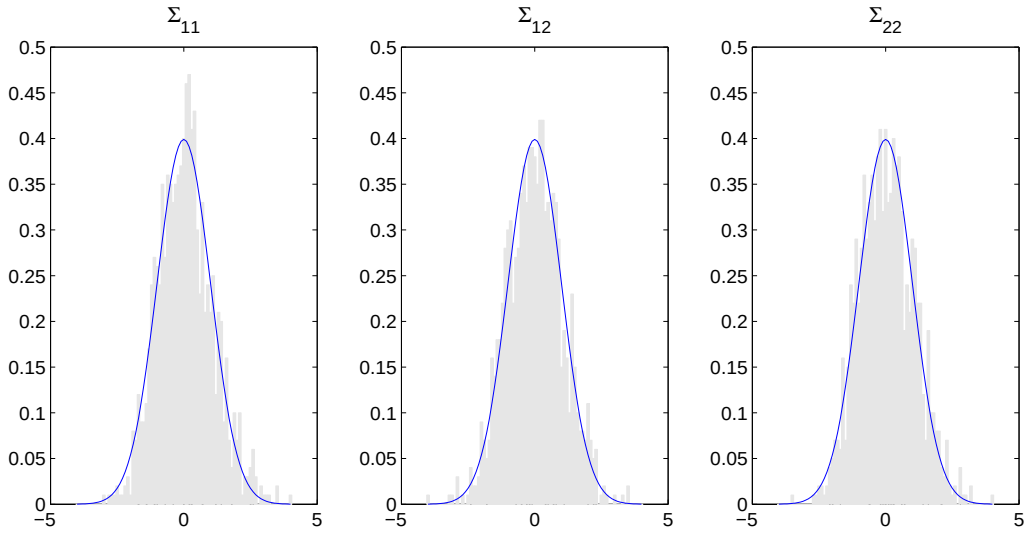

Figure 2: The figure plots the histograms of the standardized pivotal statistics to verify the asymptotic theory developed in Theorem 2. The standardisation is carried out using the infeasible true random asymptotic variance for each replication.

uniform scatter these points over the time interval $t \in [0, T]$, recalling $T = 1/252$. For asset 1 we will scatter exactly $n\digamma$ points and for asset 2 there will be exactly $n(1-\digamma)$ points. This kind of stratified scatter corresponds to a sample from a Poisson bridge process with intensity $n\digamma/T$ and $n(1-\digamma)/T$, respectively. We call $\digamma$ the mixture rate and take $\digamma \in \{0.1, 0.5, 0.9\}$.

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

5.2 Our suite of estimators

We will compute five estimators of Σ_{11}, Σ_{22}, Σ_{12}, $\beta_{1|2}$, $\beta_{2|1}$ and $\rho_{1,2}$. The five are: (i) realised QML, (ii) multistep realised QML estimator, (iii) realised QML but using the reduced data synchronised by refresh time, (iv) the realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) which uses refresh time, (v) Aït-Sahalia, Fan, and Xiu (2010) which uses polarisation and refresh time. We write this generically as $\widehat{\theta}_L$, with $L \in \{QML, Step, RT, Kern, Pol\}$.

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

5.3 Results

Table 2 reports simulation based estimates of the 0.9 quantiles of $\left|n^{1/4}\left(\widehat{\theta}_L - \theta_L\right)\right|$ for various values of n, L and $\digamma$. The results indicate that all five 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 realised QML being the best by a small margin. QML almost always slightly outperformed Step. The quantiles for Kern seem to mildly increase with n, which is expected due to their slower rate of convergence. For Σ_{12} there are signs that the realised QML type estimators "QML"and "Step" perform better than "RT". All seem to perform more strongly than the "Kern" and "Pol" estimators. The differences are less important in the case where $F = 0.5$.

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

Before we progress to the regression case it is helpful to calibrate how accurately we have estimated $\rho_{1,2}$ in the realised QML case. When $F = 0.1$ the quantile is 2.61 with $n = 117$, so the corresponding quantile for $\left|\widehat{\theta}_{QML} - \theta_{QML}\right|$ is 0.794. When $n = 1,170$ it is 0.388. When $n = 11,700$ it is 0.191. In the balanced case $F = 0.5$ the corresponding results are 0.556, 0.277 and 0.137. Hence balanced data helps, but not by very much as long as n is moderately large and the realised QML method is used. Balancing is much more important for RT, Kern and Pol. We think this makes the realised QML approach distinctly promising. A final point is worth noting. Even though $n = 11,700$ the quantiles in the balanced case of 0.137 are not close to zero. Hence although we can non-parametrically estimate the correlation between assets, the estimation in practice is not without important error. This is important econometrically when we come to using these objects for forecasting or decision making.

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

6 Empirical implementation

6.1 Our database

We use data from the cleaned trade database of Lunde, Shephard, and Sheppard (2012). It is taken from the TAQ database accessed through the Wharton Research Data Services (WRDS) system.

			$\digamma = 0.9$					$\digamma = 0.5$					$\digamma = 0.1$				
	n	$n^{1/4}$	QML	Step	RT	Kern	Pol	QML	Step	RT	Kern	Pol	QML	Step	RT	Kern	Pol
$\widehat{\Sigma}_{11}$	117	3.29	**0.40**	0.43	0.75	0.78	0.79	**0.49**	0.52	**0.49**	0.55	0.54	0.72	0.72	0.74	**0.69**	0.72
	1,170	5.84	**0.38**	0.39	0.70	0.84	0.79	**0.44**	0.45	0.48	0.74	0.47	**0.69**	0.73	0.70	0.83	0.73
	11,700	10.3	0.36	**0.35**	0.63	0.95	0.64	**0.42**	0.43	0.43	0.95	0.43	**0.63**	0.67	0.65	0.96	0.67
$\widehat{\Sigma}_{12}$	117	3.29	0.23	**0.22**	0.27	0.32	0.31	0.19	**0.18**	**0.18**	0.22	0.22	0.27	**0.24**	0.28	0.33	0.32
	1,170	5.84	**0.19**	**0.19**	0.25	0.35	0.31	**0.17**	**0.17**	**0.17**	0.22	0.20	0.25	**0.24**	0.25	0.35	0.32
	11,700	10.3	**0.18**	**0.18**	0.23	0.36	0.27	**0.17**	**0.17**	**0.17**	0.22	0.18	**0.22**	**0.22**	0.24	0.34	0.27
$\widehat{\Sigma}_{22}$	117	3.29	**0.29**	0.35	**0.29**	0.39	0.35	**0.16**	0.20	0.19	0.24	0.19	**0.13**	0.16	0.29	0.40	0.35
	1,170	5.84	**0.23**	0.29	**0.23**	0.46	0.29	**0.14**	0.16	0.15	0.26	0.16	**0.11**	0.14	0.22	0.46	0.28
	11,700	10.3	**0.21**	0.22	**0.21**	0.42	0.22	**0.14**	**0.14**	**0.14**	0.25	0.15	**0.11**	**0.11**	0.21	0.45	0.24
$\widehat{\rho}_{1,2}$	117	3.29	**2.35**	**2.35**	3.05	2.96	8.14	1.83	1.83	2.02	**1.58**	2.46	**2.61**	**2.61**	3.21	2.81	7.06
	1,170	5.84	**1.60**	**1.60**	2.53	2.54	3.10	**1.62**	**1.62**	1.66	1.72	1.86	**2.27**	**2.27**	2.50	2.52	3.04
	11,700	10.3	**1.48**	**1.48**	2.17	2.61	2.47	**1.43**	**1.43**	1.54	2.07	1.67	**1.99**	**1.99**	2.11	2.57	2.47
$\widehat{\beta}_{1\|2}$	117	3.29	**6.36**	11.70	7.10	7.69	14.09	**3.20**	3.28	3.36	3.41	4.05	5.12	**4.97**	7.76	7.91	13.98
	1,170	5.84	**3.18**	3.24	4.09	4.75	4.82	2.62	**2.61**	2.82	3.00	3.03	**3.56**	3.65	4.12	5.06	4.93
	11,700	10.3	**2.82**	2.86	4.10	5.50	4.36	**2.62**	2.64	2.82	3.44	3.04	**3.52**	3.54	3.78	5.06	4.65
$\widehat{\beta}_{2\|1}$	117	3.29	2.97	**2.55**	7.62	3.22	23.27	3.28	3.38	3.85	**1.75**	4.17	5.60	11.05	7.18	**3.42**	15.16
	1,170	5.84	2.05	**2.02**	3.91	3.17	4.44	**2.14**	2.23	2.25	2.60	2.51	3.31	3.79	3.89	**3.29**	4.72
	11,700	10.3	1.68	**1.67**	3.04	3.26	3.20	**1.75**	1.76	1.91	3.61	1.97	2.66	**2.65**	2.83	3.50	3.24

Table 2: Monte Carlo results for the volatility, covariance, correlation and beta estimation. We report the 0.9 quantiles of $\left|n^{1/4}\left(\widehat{\theta}_L - \theta_L\right)\right|$ over 1,000 independent replications. $\digamma$ denotes the % of the data corresponding to trades in asset 1. "QML"is our multivariate QMLE. "Step"is our multistep QMLE. "RT"is our multivariate QML using the Refresh Time. "Kern"is the existing multivariate realised kernel. "Pol"is the existing polarisation and Refresh Time estimator. The numbers in bold indicates the minimum of quantiles in comparison.

The exchanges open at 9.30 and close at 16.00 local time.

An important feature of this TAQ data is that times are recorded to a second, so we take the median of multiple trades which occur in the same second. This median is thus a form of miniature preaveraging. Prices are recorded in seconds, so the maximum daily sample size is $23,400$.

The data range from 1st January 2006 until 31st December 2009. We have selected 13 stocks from the S&P 500 with the aim of having 2 infrequently traded and 11 highly traded assets.

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

6.2 Summaries

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 and the corresponding result for all the 13 assets.

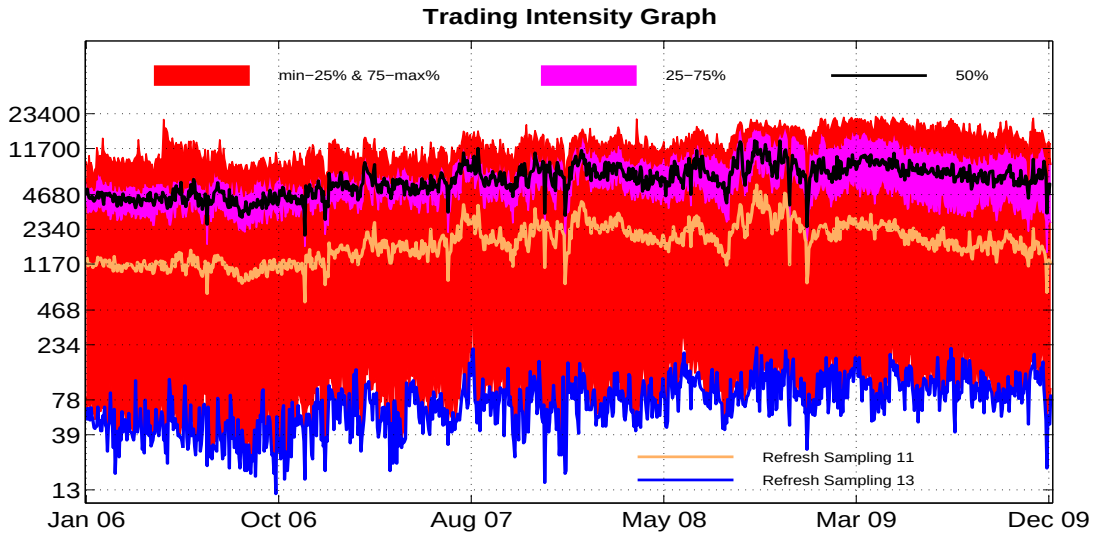

Figure 3: The trading intensity graph for our 13 assets. The figure plots the min, max, 25%, 50%, and 75% quantiles of the number of observations for the cleaned dataset. The number of refresh sampling for the 11 asset and 13 asset databases are also plotted.

The median intensity for the 13 assets is around $5,000$ a day, slightly increasing through time. The maximum daily sample sizes are around $15,000$. 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 graph demonstrates that the refresh time approach is limited, for in large unbalanced systems it leads to a significant reduction in data. This damages the effectiveness of the realised kernel or preaveraging estimators.

6.3 Volatilities

We start by looking at the univariate volatilities. We will compare realised QML, kernels and volatilities. The comparison will be made using estimators which use the one and 13 dimensional datasets. The question is whether the use of the high dimensional series will disrupt the behavior of the volatility estimators, due to the use of refresh time[3]. We will focus on a single representative series, Exxon Mobile Corporation common stock (XOM), and some cross sectional summaries. It is important not to overreact to the specific features of a single series, we will make remarks only on characteristics which work out in the cross section. Our web Appendix details various summary statistics for the volatility statistics, including volatility signature plots and correlograms.

A way of seeing the relative importance of these realised measures is through prediction. Here we use GARCHX models, supplementing the usual GARCH models of returns with X variables which are lagged realised type quantities. In particular we fit $\sigma_t^2 = \text{Var}(y_t|\mathcal{F}_{t-1}^{y,x})$ where $\sigma_t^2 = \omega+\alpha y_{t-1}^2+\beta\sigma_{t-1}^2+\gamma x_{t-1}$. Here y_t is the t-th open to close return. These kind of extended GARCH models are now common in the literature, examples include Engle and Gallo (2006), Brownlees and Gallo (2010), Shephard and Sheppard (2010) and Hansen, Huang, and Shek (2011).

The model is fitted using a Gaussian quasi-likelihood with $-\frac{1}{2}\sum_{t=12}^{n}\left(\log\sigma_t^2 + y_t^2/\sigma_t^2\right)$, taking $\sigma_1^2 = \frac{1}{11}\sum_{t=1}^{11} y_t^2$. We will report the estimated α, β, γ and the non-negative change the log likelihood in comparison with the simpler GARCH model (when $\gamma = 0$). If the presence of the realised quantity moves the likelihood up we will think this is evidence for its statistical usefulness. The results for all 13 assets are in our Web Appendix.

First focus on the XOM case. Table 3 shows the results in the univariate and 13 dimensional

[3]To be explicit, when we compute the realised kernel we take the data and approximately synchronise it using Refresh Time. This synchronised dataset is then used in all the realised kernel calculations. In the case where the analysis is carried out using the univariate databases, Refresh Time has no impact. In the 13 dimensional case it dramatically reduces the sample size due to the inclusion of the slow trading markets. When we sparsely sample, we first sparsely sample and then compute the Refresh Time coordination. This has less impact than one might expect at first sight, as sparsely sampling has little impact on the slow trading stocks and these largely determine the Refresh Times. Hence the realised kernel will not be very impacted by sparse sampling in the multivariate case. There is an argument that with the realised kernel we should only report results for sparsity being one, as that is the way it was introduced. For completeness though we have recomputed it for all the different levels of sparsity.

			GARCHX											
			QML				RK				RV			
Mkt	Dimen	Spar	α	β	γ	logL	α	β	γ	logL	α	β	γ	logL
XOM	1		0.10	0.86										
XOM	1	1	0.03	0.58	0.24	30.8	0.01	0.62	0.26	31.7	0.01	0.51	0.34	31.4
XOM	1	2	0.01	0.62	0.26	33.3	0.01	0.64	0.26	31.8	0.01	0.55	0.30	31.8
XOM	1	3	0.01	0.63	0.27	32.8	0.01	0.64	0.25	31.6	0.01	0.60	0.27	31.9
XOM	1	5	0.01	0.63	0.27	32.7	0.01	0.66	0.24	31.4	0.01	0.63	0.25	31.7
XOM	1	10	0.01	0.67	0.23	31.3	0.01	0.69	0.22	31.1	0.01	0.64	0.26	32.5
XOM	1	15	0.01	0.70	0.21	30.3	0.01	0.70	0.22	30.2	0.01	0.64	0.25	31.7
XOM	1	20	0.01	0.73	0.19	29.6	0.01	0.71	0.21	29.7	0.01	0.65	0.24	31.7
XOM	1	30	0.01	0.74	0.18	27.8	0.01	0.72	0.20	28.8	0.01	0.67	0.23	31.3
XOM	1	60	0.01	0.75	0.18	27.3	0.00	0.72	0.22	28.0	0.01	0.71	0.20	30.0
XOM	1	120	0.01	0.74	0.19	27.3	0.00	0.74	0.21	26.2	0.01	0.74	0.19	28.1
XOM	1	180	0.00	0.75	0.19	27.3	0.01	0.75	0.20	27.1	0.01	0.75	0.17	26.9
XOM	1	300	0.00	0.77	0.17	25.7	0.01	0.76	0.19	26.3	0.01	0.75	0.18	25.8
XOM	13	1	0.02	0.63	0.27	31.9	0.00	0.72	0.25	29.9	0.01	0.51	0.34	31.4
XOM	13	2	0.01	0.63	0.29	33.0	0.00	0.71	0.26	30.3	0.01	0.54	0.30	31.7
XOM	13	3	0.01	0.63	0.29	32.7	0.00	0.74	0.23	29.3	0.02	0.56	0.29	31.6
XOM	13	5	0.01	0.63	0.27	32.1	0.00	0.77	0.23	30.0	0.01	0.60	0.27	31.8
XOM	13	10	0.01	0.65	0.24	31.2	0.00	0.76	0.24	27.5	0.01	0.62	0.26	32.1
XOM	13	15	0.01	0.67	0.23	30.7	0.00	0.77	0.23	27.2	0.01	0.63	0.26	32.5
XOM	13	20	0.01	0.69	0.21	29.4	0.00	0.76	0.24	28.7	0.01	0.62	0.26	32.5
XOM	13	30	0.01	0.70	0.20	29.0	0.00	0.77	0.24	29.2	0.01	0.65	0.24	32.4
XOM	13	60	0.02	0.71	0.20	28.3	0.00	0.76	0.25	27.8	0.01	0.69	0.22	31.4
XOM	13	120	0.01	0.75	0.18	27.5	0.00	0.78	0.23	26.8	0.01	0.70	0.21	30.4
XOM	13	180	0.01	0.76	0.16	27.0	0.00	0.80	0.22	26.5	0.01	0.72	0.19	28.9
XOM	13	300	0.01	0.76	0.16	25.7	0.00	0.81	0.20	21.3	0.01	0.74	0.18	26.7

Table 3: Forecasting exercises. GARCHX models. LogL denotes increase in the log-likelihood compared to the GARCH model. $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma x_{t-1}$, where x_t is a realised quantity.

cases. The results show across the board important improvements when using the realised quantities and α is basically forced to zero. This is the common feature of these models in the literature, once the realised quantities are there there is no need for the square return (Shephard and Sheppard (2010)). Further β falls dramatically and meaningfully. It means that the average lookback of the forecast has reduced considerably, making it also more robust to structural breaks.

For some series the realised volatility adds little when the sparsity is 1 (due to the impact of the market microstructure), but this is not the case for XOM. There is some evidence that the QML estimator does a little better when sparsity is a tiny amount above 1. All the estimators trail off as the sparsity gets large. When we move to the 13 dimensional case the QML results hardly change and obviously the RV case does not change at all. The realised kernel results are reasonably consistently damaged in this case, although the damage is not exceptional.

We can now average the likelihood changes using the cross-section of thickly traded stocks. The results are given in Figure 4. This reports the likelihood for QML minus the likelihood for RK. On the left hand side we deal with the univariate case. On the right the 13 dimensional case is

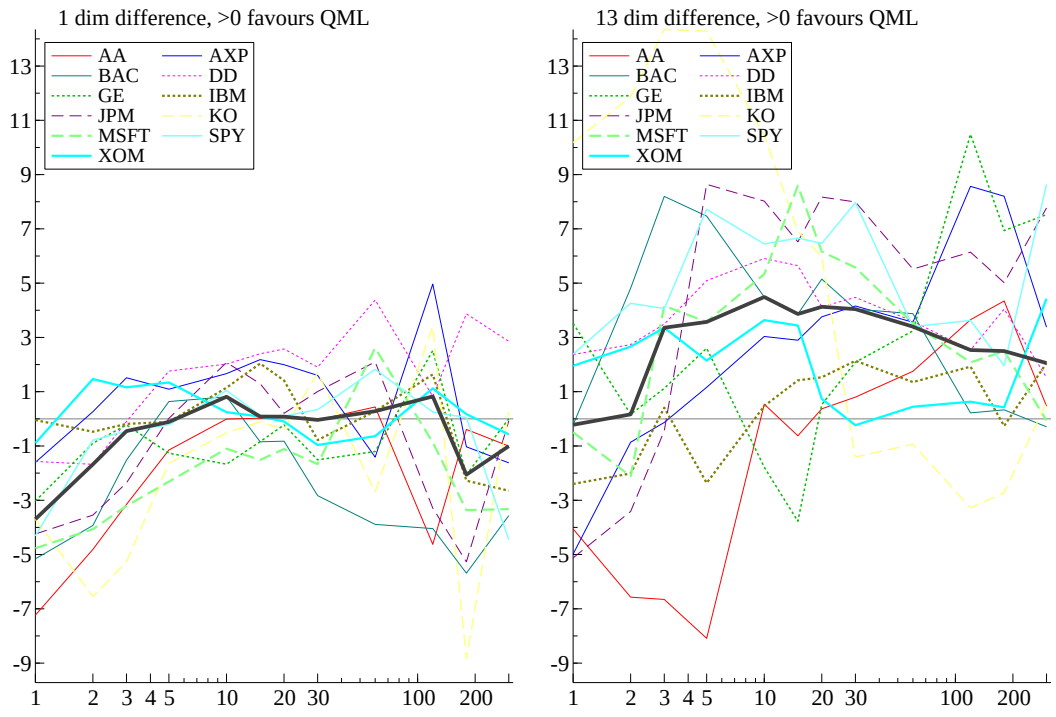

Figure 4: logL improvement of realised QML minus logL of RK, with numbers greater than 0 being supportive of QML. X-axis is the level of sparsity. 1 and 13 dimensional cases are on the left and right hand sides respectively.

the focus. So negative numbers prefer RK. This shows in the univariate case at sparsity of 1 RK is better, but this preference is removed by the time we reach sparsity of 3. After that they are basically the same. When we look at the 13 dimensional case, except for sparsity of 1, QML is better. This is consistent across many different levels of sparsity.

6.4 Dependence

6.4.1 Summary statistics

We now turn to looking at covariation amongst the assets. We will focus on QML, realised kernel and realised covariance[4] estimators, which are computed each day. In all cases they will be based on the 13 unbalanced database, which means we compute a 13×13 positive semidefinite estimator of the covariance matrix. To look inside the covariance matrix we will focus on pairs of assets. To be concrete our focus will be on Bank of America (BAC) and SPY, the other 77 pairs are discussed in our web appendix. Again we will only flag up issues which hold up in the cross section.

The top left of Figure 5 shows the time series evolution of the conditional volatilities for these series based upon the past QMLs. It shows the typically lower level in the SPY series. What is

[4]For the realised covariance we use the last price update available at the times the prices are sampled. This means that for a d-dimensional dataset, at each price update $d-1$ of the prices will be stale. For sparsity of 1 this will lead to significant bias in the covariance, the so-called Epps effect. Larger sparsity delivers smaller bias.

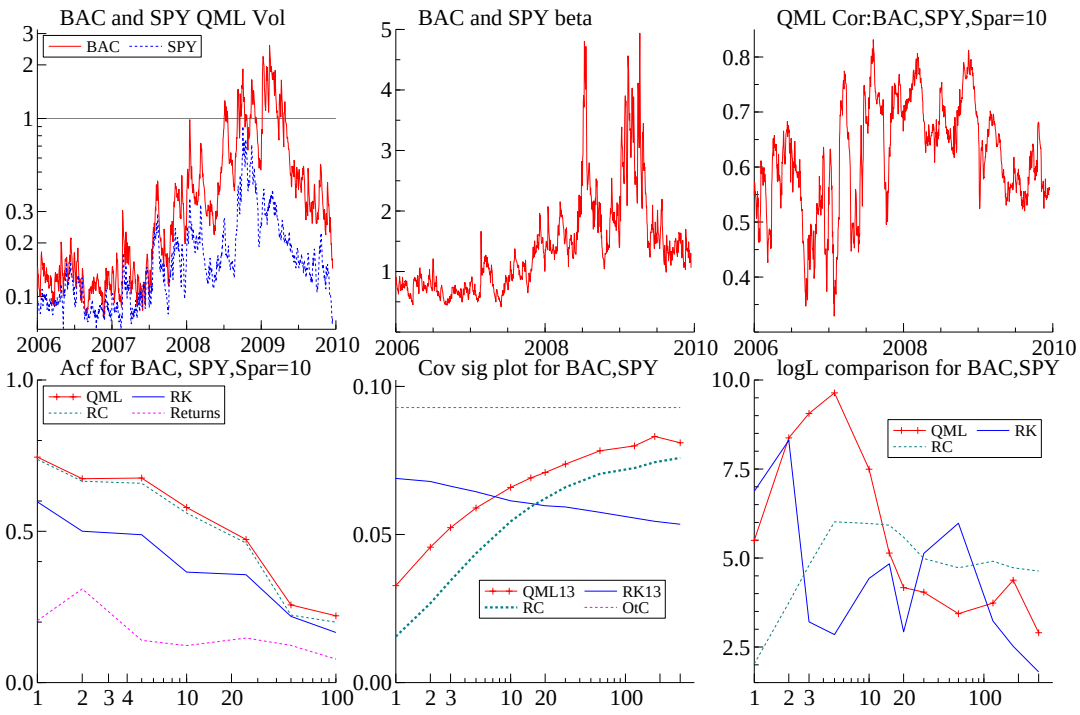

Figure 5: Summary picture for of the BAC and SPY pair. All realised quantities are computed using all 13 series. Top left are left are the conditional volatilities based upon the realised QMLs. Top middle, conditional beta for BAC on SPY. Top right is the daily realised QML covariance. Bottom right is the acf of the realised QML, realised kernel and realised covariance estimators. Also shown is the acf of the cross-product of the daily returns. Bottom middle is the covariance signature plot, graphed against the level of sparsity. Bottom right is the log likelihood improvement in model fit by including the realised quantity, graphed against the level of sparsity.

key is that the wedge between these two series dramatically opens from 2008 onwards which means the ratio of the standard deviation of BAC and SPY has increased a great deal. If the correlation between the two series is stable this would deliver a massive increase in the "beta" of BAC. This is what actually happened, as can be seen in the middle top graph.

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

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

Bottom left shows the autocorrelations of the QML, realised kernel, realised covariance and

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

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

Table 4: BAC and SPY pair summaries for the covariances. OtC denotes the open to close (annualised) daily log returns and the figure which follows it is the sample covariance of the returns. The acf of the OtC is the acf of the cross product of the returns, here reported at lags 1 and 100. QML, RK and RV are, respectively, the realised QML, the realised realised kernel and the realised covariance. All estimate the daily covariance of the series and are computed using the 13 dimensional database. Spar denotes sparsity. Mean is the temporal average of the time series.

Table 4 is more informative. It first shows the average open to close covariance during this sample and the average level of the daily covariance estimator, using the QML, kernel and realised covariance. This is printed out for different levels of sparsity. The focus here is thus on the different biases in the estimators[5]. Estimating covariances is hard. The Epps effect can be seen through RV, which underestimates the covariance by an order of magnitude when using every trade. It takes sparsity of 10 to reproduce half of the correlation.

The same impression can be gleaned from looking at the lower part of Figure 5, which is a covariance signature plot — showing the temporal average of the daily covariance estimators as a function of the level of sparsity. For low levels of sparsity the realised kernel is the least biased and the realised covariance the most. For higher levels of sparsity the realised kernel gets a little worse and both the QML and realised covariance improves and overtakes the realised kernel. Throughout

[5]Recall the realised kernel uses Refresh Time which is a kind of sparse sampling. Hence it having less bias at small levels of sparse sampling is not a surprise.

QML is better than the realised covariance by a considerable margin. The Table also shows the autocorrelation of the individual time series at lags 1 and 100. These results are striking, with more dependence for the QML series than the kernel. This matches results we see for the volatilities.

6.4.2 Dependence forecasting

We now move on to forecasting. The focus will be on the conditional covariance matrix $\Sigma_t = \text{Cov}(y_t|\mathcal{F}_{t-1}^{y,x})$, where y_t is the d-dimensional open to close daily return vector. We will write $\Sigma_t = D_t R_t D_t$ where D_t is a diagonal matrix with conditional standard deviations on the diagonal where the conditional variances are $\sigma_{it}^2 = \text{Var}(y_{it}|\mathcal{F}_{t-1}^{y,x})$ where $\sigma_{i,t}^2 = \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 + \gamma_i x_{i,t-1}$, $\omega_i, \alpha_i, \beta_i, \gamma_i \geq 0$. This is the same conditional volatility model as we used in the previous subsection. We use these volatilities to construct the i-th devolatilised series

$$e_{i,t} = \frac{y_{i,t}}{\sigma_{i,t}}. \tag{10}$$

Here the ij-th element of R_t is $\text{Cor}(y_{i,t}, y_{j,t}|\mathcal{F}_{t-1}^{y,x}) = \text{Cor}(e_{i,t}, e_{j,t}|\mathcal{F}_{t-1}^{y,x})$. It is our focus here and we will assume it follows, with $\omega + \alpha + \beta + \gamma = 1$,

$$R_t = \omega\Pi + \alpha C_{t-1} + \beta R_{t-1} + \gamma X_{t-1}, \quad \omega, \alpha, \beta, \gamma \geq 0. \tag{11}$$

X is a realised type correlation matrix and Π is a correlation matrix. The ij-th element of C_t is a moving block correlation of the devolatilised series

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

In the case where $M = d$ and there is no X variables, then (11) is the Tse and Tsui (2002) model[6]. Throughout we set $M = 66$, representing about 3 months of past data. Here X will represent the realised QML, realised kernel and realised covariance matrices, while $R_t = C_M$ for all $t \leq M$.

Note (11) is a weighted sum of four correlation matrices, where the weights sum to one. Hence R_t is always a correlation matrix. If X is biased then Π can partially compensate by not being the unconditional correlations of the innovations (10). This happens in practice.

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

$$\log L \;\; = \;\; -\frac{1}{2}\log|\Sigma_t| - \frac{1}{2} y_t' \Sigma_t^{-1} y_t = \log L_M + \log L_C, \quad \text{where}$$

[6] An alternative would be to use the DCC model (e.g. Engle (2009)) which would have the form of

$$Q_t = \omega\Pi + \alpha e_{t-1} e_{t-1}' + \beta Q_{t-1} + \gamma X_{t-1}, \quad R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2}.$$

Unfortunately the impact of the non-linear transform for R_t could be rather gruesome on the realised correlation matrix X_{t-1} as the rescaling by the diagonal elements of Q_t destroys all of its attractive properties.

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

$$\log L_{M_i} = -\frac{1}{2}\sum_{t=M+1}^{n}\left(\log\sigma_{it}^2 + \frac{y_{it}^2}{\sigma_{it}^2}\right), \quad \text{and} \quad \log L_C = -\frac{1}{2}\sum_{t=M+1}^{n}\left(\log|R_t| + \frac{1}{2}e_t' R_t^{-1} e_t - e_t' e_t\right).$$

Here $e_t = D_t^{-1} y_t$, the vector of "devolatilised returns". The $\log L_C$ term is a copula type likelihood.

The model is estimated using a two-step procedure (e.g. Newey and McFadden (1994)). First we estimate the univariate models, and fix the volatility dynamic parameters at those estimated values. We then estimate the dependence model by optimising $\log L_C$.

	QML					RK					RV				
Spar	ρ	α	β	γ	logL	ρ	α	β	γ	logL	ρ	α	β	γ	logL
	0.69	0.59	0.00	0.00		0.69	0.59	0.00	0.00		0.69	0.59	0.00	0.00	
1	0.99	0.31	0.13	0.25	5.5	0.94	0.06	0.73	0.14	6.8	0.99	0.43	0.00	0.17	2.0
2	1.00	0.12	0.41	0.25	8.3	0.99	0.09	0.69	0.17	8.3	0.99	0.31	0.00	0.23	3.7
3	1.00	0.08	0.53	0.23	9.0	0.80	0.09	0.71	0.11	3.2	0.99	0.13	0.34	0.18	4.8
5	0.99	0.05	0.61	0.21	9.6	0.78	0.09	0.70	0.10	2.8	0.99	0.07	0.44	0.18	6.0
10	0.99	0.06	0.64	0.20	7.5	0.81	0.05	0.77	0.10	4.4	0.99	0.07	0.50	0.19	5.9
15	0.99	0.07	0.72	0.15	5.1	0.80	0.07	0.71	0.12	4.8	0.99	0.07	0.55	0.18	5.9
20	0.99	0.07	0.74	0.14	4.1	0.80	0.07	0.74	0.10	2.9	0.99	0.08	0.55	0.19	5.5
30	0.99	0.06	0.78	0.11	4.0	0.84	0.05	0.74	0.12	5.1	0.99	0.10	0.57	0.18	4.9
60	0.96	0.07	0.79	0.11	3.4	0.87	0.04	0.76	0.12	5.9	0.99	0.10	0.64	0.16	4.7
120	0.95	0.08	0.76	0.11	3.7	0.82	0.08	0.71	0.09	3.2	0.99	0.14	0.56	0.20	4.9
180	0.89	0.09	0.71	0.14	4.3	0.82	0.08	0.72	0.08	2.5	0.99	0.13	0.57	0.21	4.7
300	0.85	0.09	0.70	0.12	2.9	0.82	0.08	0.72	0.08	1.8	0.99	0.10	0.69	0.15	4.6

Table 5: Forecasting exercises for crosss-sectional dependence between BAC (Bank of America) and SPY (S&P 500 exchange traded fund). The estimated model is (11). Here ρ is the non-unit element of Π. LogL is the improvement in the log likelihood compared to $\gamma = 0$.

The results from this forecasting exercise are given in Table 5. These are based upon the innovations from the univariate volatility models for BAC and SPY conditioning on lagged realised QML statistics. The results for the dependence model when we do not condition on any additional realised quantities, that is $\gamma = 0$, are given above the line in the Table. As $\beta = 0$ it means $\omega = 1 - \alpha$ and so is roughly 0.41. Here ρ is the non-unit element of Π. Hence the estimated model for the conditional correlation is $0.41 \times 0.69 + 0.59 C_{1,2,t-1}$ where $C_{1,2,t-1}$ is the block correlation amongst the BAC and SPY innovations. This can be thought of as simply a shrunk block correlation.

When we condition on lagged realised quantities the likelihood will typically rise. The improvement is recorded as logL in the Table. The results are reported for the QML, realised kernel and realised covariance. Obviously the results vary with the level of sparsity. For low levels of sparsity, RK does best. It drives α down to near zero, reminding us of the results we saw in the univariate cases discussed in the previous subsection. However, the improvement in the log likelihood is relatively modest, certainly less than from the univariate cases.

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

6.4.3 Cross section

Here we just focus on the cross section involving the 12 SPY based pairs. The top left of Figure 6 shows the log-likelihood improvement in $\log L_C$ by including the realised QML information, i.e. allowing γ to be greater than zero. The improvement is shown for each level of sparsity and is plotted separately for each of the 12 pairs. The median improvement is shown by the dark line. Almost throughout the improvement is modest, for a sole series the improvement is quite large. The realised QML performs better as the level of sparsity increases, but once again it tails off at the very end with very large sparsity as the realised estimator is noisy.

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

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

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

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

7 Conclusion

This paper 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

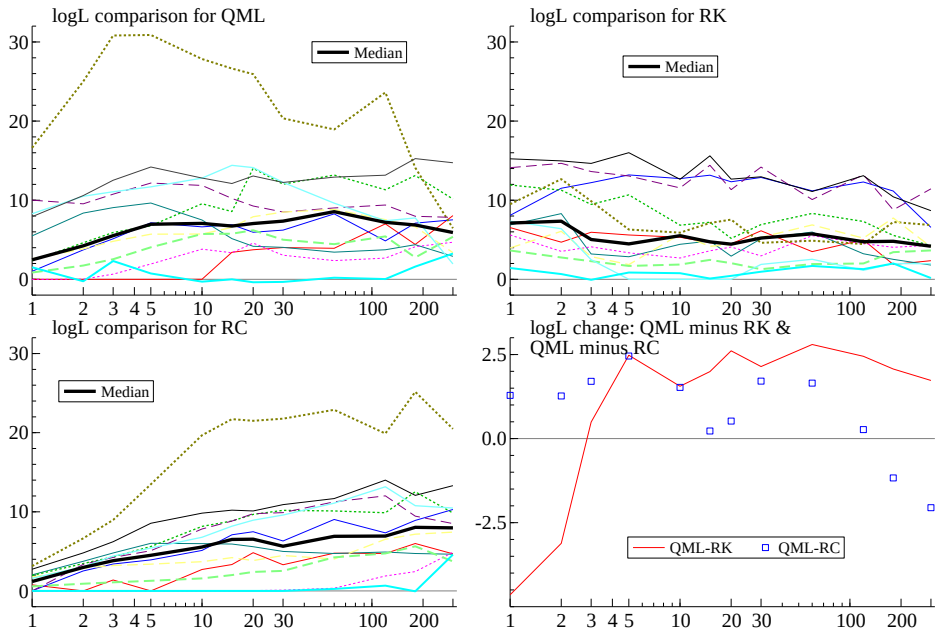

Figure 6: Gains in logL for pairs involving SPY by including the realised quantities. On the x-axis is the level of sparsity, the heavy line denotes the median. All realised quantities are computed using 13 series. Bottom right is the logL for the model including realised QML minus the corresponding figure for realised kernel. Also drawn is the result for the realised QML against realised covariance.

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 show how to numerically optimise the likelihood 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 of betas and correlations. Particularly important is our theory for asynchronous data. Our Monte Carlo experiments are extensive, comparing the estimator to various alternatives. The realised QML performs well in these comparisons, in particular in unbalanced cases.

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

Our recommendation for empirical work is to use realised QML or realised kernel estimators

inside conditional volatility models. When modelling dependence amongst the devolatilised returns the decision to use extra realised information is more balanced — it does increase the performance of the model but not by a great deal. Our analysis shows why this is to be expected.

8 Acknowledgements

We are particularly grateful for help we have received from Francesco Audrino, Markus Bibinger, Fulvio Corsi, Siem Jan Koopman, Cheng Liu, Stefano Peluso, Kevin Sheppard and Cheng Yong Tang. We also thank seminar participants at CEMFI, Harvard University, Boston University, University of Chicago, University of Illinois at Chicago, University of Montreal, especially Stéphane Bonhomme, Enrique Sentana, and David Veredas. This research was supported in part by the FMC Faculty Scholar Fund at the University of Chicago Booth School of Business.

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., J. Jacod, and J. Li (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics 168*, 207 – 222.

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.

Aït-Sahalia, Y. and D. Xiu (2012). Likelihood-based volatility estimators in the presence of market microstructure noise: A review. In L. Bauwens, C. M. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and their Applications*, Chapter 14, pp. 347–361. Wiley.

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 (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 (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.

Bibinger, M., N. Hautsch, P. Malec, and M. Reis (2014). Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. *Annals of Statistics*. Forthcoming.

Brownlees, C. T. and G. M. Gallo (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics 8*, 29–56.

Cartea, A. and D. Karyampas (2011). Volatility and covariation of financial assets: A high-frequency analysis. *Journal of Banking and Finance 35*, 3319–3334.

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 (2014). Missing asynchronicity: a Kalman-EM approach to multivariate realized covariance estimation. *Journal of Applied Econometrics*. Forthcoming.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Engle, R. F. (2009). *Anticipating Correlations*. Princeton University Press.

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

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.

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., Z. Huang, and H. H. Shek (2011). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*, 877–906.

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.

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., M. Podolskij, and M. Vetter (2010). Limit theorems for moving averages of discretized processes plus noise. *Annals of Statistics 38*, 1478 – 1545.

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 (2013). Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise. *North American Journal of Economics and Finance 26*, 282–309.

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 endogeniety of time matters. Working Paper, Department of Statistics, University of Chicago.

Liu, C. and C. Y. Tang (2014). A quasi-maximum likelihood approach to covariance matrix with high frequency data. *Journal of Econometrics*. Forthcoming.

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.

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

Malliavin, P. and M. E. Mancino (2009). A fourier transform method for nonparametric estimation of multivariate volatility. *Annals of Statistics 37*, 1983–2010.

Mancino, M. E. and S. Sanfelici (2008). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics and Data Analysis 52*, 2966–2989.

Mancino, M. E. and S. Sanfelici (2011). Covariance estimation and dynamic asset allocation under microstructure effects via fourier methodology. *Journal of Financial Econometrics 9*, 367–408.

McAleer, M. and M. C. Medeiros (2008). Realized volatility: a review. *Econometric Reviews 27*, 10–45.

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

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.

Owens, J. P. and D. G. Steigerwald (2006). Noise reduced realized volatility: A Kalman filter approach. In T. Fomby and D. Terrell (Eds.), *Advances in Econometrics*, Volume 20, pp. 211–227. Elsevier.

Park, S. and O. B. Linton (2012a). Estimating the quadratic covariation matrix for an asynchronously observed continuous time signal masked by additive noise. Unpublished paper: Financial Markets Group, London School of Economics.

Park, S. and O. B. Linton (2012b). Realized volatility: Theory and application. In L. Bauwens, C. Hafner, and L. Sebastien (Eds.), *Volatility Models And Their Applications*, pp. 293–312. London: Wiley.

Peluso, S., F. Corsi, and A. Mira (2012). A Bayesian high-frequency estimator of the multivariate covariance of noisy and asynchronous returns. Unpublished paper: University of Lugano, Switzerland.

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.

Sanfelici, S. and M. E. Mancino (2008). Covariance estimation via Fourier method in the presence of asynchronous trading and microstructure noise. Economics Department Working Papers 2008-ME01, Department of Economics, Parma University (Italy).

Shephard, N. and K. K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*, 197–231.

Stein, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association 82*, 765–772.

Tse, Y. K. and K. C. Tsui (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economic Statistics 20*, 351–362.

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. Aït-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 Collection of proofs

A.1 Proof of Theorem 1

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

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

where

$$\begin{aligned} &\Omega_{12} = \Sigma_{12}\Delta \otimes I, \quad \Omega_{ii} = \Sigma_{ii}\Delta \otimes I + \Lambda_{ii} \otimes J, \quad i = 1, 2, \\ &u_{ij} = \sqrt{\frac{2}{n+1}} \sin\Big(\frac{i \cdot j}{n+1}\pi\Big), \quad i, j = 1, \ldots, n, \\ &\mu_{ij} = \Sigma_{ii}\Delta + 2\Lambda_{ii}\Big(1 - \cos\Big(\frac{j}{n+1}\pi\Big)\Big), \quad i = 1, 2, \text{and } j = 1, \ldots, n. \end{aligned}$$

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

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

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

$$\begin{aligned} tr\Big(\Omega^{-1}\tfrac{\partial \Omega}{\partial \Sigma_{11}}\Big) &= tr\Big(V^{-1}\tfrac{\partial V}{\partial \Sigma_{11}}\Big) = \textstyle\sum_{i=1}^n \frac{\mu_{2i}\Delta}{\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2} \\ tr\Big(\Omega^{-1}\tfrac{\partial \Omega}{\partial \Sigma_{12}}\Big) &= tr\Big(V^{-1}\tfrac{\partial V}{\partial \Sigma_{12}}\Big) = \textstyle\sum_{i=1}^n \frac{-2\Sigma_{12}\Delta^2}{\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2}, \\ tr(\Omega^{-1}\tfrac{\partial \Omega}{\partial \Sigma_{11}}\Omega^{-1}rr') &= tr\left(\begin{pmatrix} \mathrm{diag}\Big(\frac{\mu_{2i}^2\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) & \mathrm{diag}\Big(\frac{-\Sigma_{12}\mu_{2i}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) \\ \mathrm{diag}\Big(\frac{-\Sigma_{12}\mu_{2i}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) & \mathrm{diag}\Big(\frac{\Sigma_{12}^2\Delta^3}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) \end{pmatrix} \begin{pmatrix} U & \\ & U \end{pmatrix} rr' \begin{pmatrix} U' & \\ & U' \end{pmatrix}\right) \\ tr(\Omega^{-1}\tfrac{\partial \Omega}{\partial \Sigma_{12}}\Omega^{-1}rr') &= tr\left(\begin{pmatrix} \mathrm{diag}\Big(\frac{-2\mu_{2i}\Sigma_{12}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) & \mathrm{diag}\Big(\frac{\Sigma_{12}^2\Delta^3 + \mu_{1i}\mu_{2i}\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) \\ \mathrm{diag}\Big(\frac{\Sigma_{12}^2\Delta^3 + \mu_{1i}\mu_{2i}\Delta}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) & \mathrm{diag}\Big(\frac{-2\mu_{1i}\Sigma_{12}\Delta^2}{(\mu_{1i}\mu_{2i} - \Sigma_{12}^2\Delta^2)^2}\Big) \end{pmatrix} \begin{pmatrix} U & \\ & U \end{pmatrix} rr' \begin{pmatrix} U' & \\ & U' \end{pmatrix}\right) \end{aligned}$$

Therefore, by direct calculations and using symmetry, we have

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

By symmetry, we can obtain I_{23}^{Σ}, I_{33}^{Σ}, and I_{22}^{Σ} by simply switching the indices 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

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

Similarly derivations on $1/n$-scaled likelihood can show that the asymptotic variance for $(\widehat{\Lambda}_{11},\widehat{\Lambda}_{22})$ 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. 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 written in integral forms. The pseudo true parameters are marked with a superscript such as $\bar{\Sigma}$ and $\bar{\Lambda}$, and the QMLEs are marked as $\widehat{\Sigma}$ and $\widehat{\Lambda}$. 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 (4), we have

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

Now we consider the following function:

$$\bar{L} = -n\log(2\pi) - \frac{1}{2}\log(\det\Omega) - \frac{1}{2}tr(\Omega^{-1}\Omega_0), \quad \text{where} \quad \Omega_0 = \begin{pmatrix} \Omega_0^{11} & \Omega_0^{12} \\ \Omega_0^{21} & \Omega_0^{22} \end{pmatrix},$$

where the subscript 0 denotes the true value with $\Omega^{ll}_{0,ii} = \int_{t_{i-1}}^{t_i}\Sigma_{ll,t}dt + 2\Lambda_{0,ll}$, $\Omega^{ll}_{0,i,i+1} = \Omega^{ll}_{0,i,i-1} = \Lambda_{0,ll}$, and $\Omega^{12}_{0,i,i} = \Omega^{21}_{0,i,i} = \int_{t_{i-1}}^{t_i}\Sigma_{12,t}dt$. Therefore, the difference between L and $\bar{L}$ is given by:

$$\begin{aligned}
L - \bar{L} =& \frac{1}{2}tr\Big(\Omega^{-1}(rr' - \Omega_0)\Big) = \frac{1}{2}tr\left(\begin{pmatrix} U' & \\ & U' \end{pmatrix} V^{-1} \begin{pmatrix} U & \\ & U \end{pmatrix}(rr' - \Omega_0)\right) \\
=& \frac{1}{2}\sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\omega_{ii}^{ls}\Big((\Delta_i y_l)(\Delta_i y_s) - \int_{t_{i-1}}^{t_i}\Sigma_{ls,t}dt\Big) + \sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j<i}\omega_{ij}^{ls}\Delta_i^n y_l \Delta_j^n y_s \\
&+ \sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}^{ls}\Delta_i^n\varepsilon_l\Delta_j^n y_s + \frac{1}{2}\sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}^{ls}\Big(\Delta_i^n\varepsilon_l\Delta_j^n\varepsilon_s - E(\Delta_i^n\varepsilon_l\Delta_j^n\varepsilon_s)\Big),
\end{aligned}$$

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.

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

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

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

$$=\frac{1}{2(n+1)}\sum_{k=0}^{n}\frac{\sin(\frac{(k+1)l}{n+1}\pi)}{\sin(\frac{l}{n+1}\pi)}\frac{\mu_{2,k+2}\mu_{2,k}(\mu_{1,k+2}-\mu_{1,k})+(\mu_{2,k+2}-\mu_{2,k})\Sigma_{12}^2\Delta^2}{(\mu_{1k}\mu_{2k}-\Sigma_{12}^2\Delta^2)(\mu_{1,k+2}\mu_{2,k+2}-\Sigma_{12}^2\Delta^2)}$$
$$=\frac{2}{n+1}\sum_{k=0}^{n}\frac{\sin(\frac{(k+1)l}{n+1}\pi)\sin(\frac{\pi}{n+1})\sin(\frac{k+1}{n+1}\pi)}{\sin(\frac{l}{n+1}\pi)}\frac{(\mu_{2,k+2}\mu_{2,k}\Lambda_{11}+\Lambda_{22}\Sigma_{12}^2\Delta^2)}{(\mu_{1k}\mu_{2k}-\Sigma_{12}^2\Delta^2)(\mu_{1,k+2}\mu_{2,k+2}-\Sigma_{12}^2\Delta^2)}$$

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

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

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

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

Similarly, we can derive

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

To simply our notation, let

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

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

$$\frac{\partial}{\partial\Sigma}=\left(\frac{\partial}{\partial\Sigma_{11}},\frac{\partial}{\partial\Sigma_{12}},\frac{\partial}{\partial\Sigma_{22}}\right)',\text{ and }\frac{\partial}{\partial\Lambda}=\left(\frac{\partial}{\partial\Lambda_{11}},\frac{\partial}{\partial\Lambda_{22}}\right)'.$$

Then we have

$$\Psi_\Sigma-\bar\Psi_\Sigma=\frac{1}{2\sqrt{n}}\Big(M_1^{(\Sigma)}+2M_2^{(\Sigma)}+2M_3^{(\Sigma)}+M_4^{(\Sigma)}\Big),$$

where

$$M_1^{(\Sigma)}=\sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\frac{\partial\omega_{ii}^{ls}}{\partial\Sigma}\Big((\Delta_iy_l)(\Delta_iy_s)-\int_{\tau_{i-1}}^{\tau_i}\Sigma_{ls,t}dt\Big)$$
$$M_2^{(\Sigma)}=\sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j<i}\frac{\partial\omega_{ij}^{ls}}{\partial\Sigma}\Delta_i^ny_l\Delta_j^ny_s,\quad M_3^{(\Sigma)}=\sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\omega_{ij}^{ls}}{\partial\Sigma}\Delta_i^n\varepsilon_l\Delta_j^ny_s$$

$$M_4^{(\Sigma)} = \sum_{l=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\omega_{ij}^{ls}}{\partial\Sigma}\Big(\Delta_i^n\varepsilon_l\Delta_j^n\varepsilon_s - E(\Delta_i^n\varepsilon_l\Delta_j^n\varepsilon_s)\Big).$$

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

$$\begin{aligned}
& n^{-\frac{1}{4}}(M_1^{(\Sigma)} + 2M_2^{(\Sigma)}) \xrightarrow{\mathcal{L}_X} MN(0, \mathrm{Avar}^{(2)}(\Sigma)), \quad \text{where} \\
\mathrm{Avar}^{(2)}(\Sigma) &= \lim_{n\to\infty} 4n^{-\frac{1}{2}} \sum_{l,s,u,v=1}^{2}\sum_{i=1}^{n}\sum_{j<i}\frac{\partial\omega_{i,j}^{v,u}}{\partial\Sigma}\frac{\partial\omega_{i,j}^{l,s}}{\partial\Sigma'}\Sigma_{sv,t_i}\Sigma_{ul,t_j}\Delta^2 \\
&= 2T\sum_{l,s,u,v=1}^{2}\int_0^\infty \frac{\partial\omega^{vu}(\Sigma,\Lambda,x)}{\partial\Sigma}\frac{\partial\omega^{ls}(\Sigma,\Lambda,x)}{\partial\Sigma'}dx\int_0^T\Sigma_{sv,t}\Sigma_{ul,t}dt.
\end{aligned}$$

All the elements of the covariance matrix have closed-forms. Also, we have

$$\begin{aligned}
& n^{-\frac{1}{4}}2M_3^{(\Sigma)} \xrightarrow{\mathcal{L}_X} MN(0, \mathrm{Avar}^{(3)}(\Sigma)), \quad \text{where} \\
\mathrm{Avar}^{(3)}(\Sigma) &= \lim_{n\to\infty} 4\sum_{j=1}^{n} n^{-\frac{1}{2}} \sum_{l,s,v=1}^{2}\sum_{i=1}^{n}\Lambda_{0,ll}\frac{\partial\omega_{ij}^{ls}}{\partial\Sigma}\Big(2\frac{\partial\omega_{ij}^{lv}}{\partial\Sigma'} - \frac{\partial\omega_{i,j-1}^{lv}}{\partial\Sigma'} - \frac{\partial\omega_{i,j+1}^{lv}}{\partial\Sigma'}\Big)\Sigma_{sv,t_j}\Delta \qquad \text{(A.1)} \\
&= 4\sum_{l,s,v=1}^{2}\Lambda_{0,ll}\int_0^\infty\frac{\partial\omega^{ls}(\Sigma,\Lambda,x)}{\partial\Sigma}\frac{\partial\omega^{lv}(\Sigma,\Lambda,x)}{\partial\Sigma'}\pi^2x^2dx\int_0^T\Sigma_{sv,t}dt. \qquad \text{(A.2)}
\end{aligned}$$

Finally, we have

$$\begin{aligned}
& n^{-\frac{1}{4}}M_4^{(\Sigma)} \xrightarrow{\mathcal{L}} N(0, \mathrm{Avar}^{(4)}(\Sigma)), \quad \text{where} \\
\mathrm{Avar}^{(4)}(\Sigma) &= \lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{i,j,k,l=1}^{n}\left(\frac{\partial\omega_{i,j}^{11}}{\partial\Sigma}\frac{\partial\omega_{k,l}^{11}}{\partial\Sigma'}K_{11}^{ij,kl} + 4\frac{\partial\omega_{i,j}^{12}}{\partial\Sigma}\frac{\partial\omega_{k,l}^{12}}{\partial\Sigma'}K_{11}^{i,k}K_{22}^{j,l} + \frac{\partial\omega_{i,j}^{22}}{\partial\Sigma}\frac{\partial\omega_{k,l}^{22}}{\partial\Sigma'}K_{22}^{ij,kl}\right) \\
&= \lim_{n\to\infty} n^{-\frac{1}{2}}\Bigg(V_1\Big(\frac{\partial\omega^{11}}{\partial\Sigma},\frac{\partial\omega^{11}}{\partial\Sigma}\Big) + V_2\Big(\frac{\partial\omega^{11}}{\partial\Sigma},\frac{\partial\omega^{11}}{\partial\Sigma}\Big) + 2V_2\Big(\frac{\partial\omega^{12}}{\partial\Sigma},\frac{\partial\omega^{12}}{\partial\Sigma}\Big) \\
&\quad + V_1\Big(\frac{\partial\omega^{22}}{\partial\Sigma},\frac{\partial\omega^{22}}{\partial\Sigma}\Big) + V_2\Big(\frac{\partial\omega^{22}}{\partial\Sigma},\frac{\partial\omega^{22}}{\partial\Sigma}\Big)\Bigg),
\end{aligned}$$

and

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

$$+\frac{\partial\omega^{12}_{j+1,i+1}}{\partial\Sigma'}-2\frac{\partial\omega^{12}_{j+1,i}}{\partial\Sigma'}-2\Big(\frac{\partial\omega^{12}_{j,i-1}}{\partial\Sigma'}+\frac{\partial\omega^{12}_{j,i+1}}{\partial\Sigma'}-2\frac{\partial\omega^{12}_{j,i}}{\partial\Sigma'}\Big)\Big)\Big)$$
$$\sim 2(\Lambda_{0,11})(\Lambda_{0,22})\Big(\int_0^\infty \frac{\partial\omega^{12}}{\partial\Sigma}\frac{\partial\omega^{12}}{\partial\Sigma'}\pi^4x^4dx\Big)n^{\frac{1}{2}}.$$

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 $\mathrm{cum}_4[\varepsilon_1]$ and $\mathrm{cum}_4[\varepsilon_2]$ are the fourth cumulants of ε_1 and ε_2.

Therefore, we have

$$\mathrm{Avar}^{(4)}(\Sigma)=2\sum_{l,s=1}^{2}\Lambda_{0,ll}\Lambda_{0,ss}\int_0^\infty \frac{\partial\omega^{ls}(\Sigma,\Lambda,x)}{\partial\Sigma}\frac{\partial\omega^{ls}(\Sigma,\Lambda,x)}{\partial\Sigma'}\pi^4x^4dx. \tag{A.3}$$

In summary, we have

$$n^{-\frac{1}{4}}(\Psi_\Sigma-\bar{\Psi}_\Sigma)\xrightarrow{\mathcal{L}}N\Big(0,\frac{1}{4}\Big(\mathrm{Avar}^{(2)}(\Sigma)+\mathrm{Avar}^{(3)}(\Sigma)+\mathrm{Avar}^{(4)}(\Sigma)\Big)\Big).$$

Similarly, we can obtain

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

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

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

where Λ_0 is the noise's true covariance matrix, Γ is block diagonal, with $\Gamma^{ls}_{ii}=\int_{\tau_{i-1}}^{\tau_i}\Sigma_{ls,t}dt-\Sigma_{ls}\Delta$, and J is $n\times n$ tridiagonal with diagonal elements being 2 and off-diagonal elements being -1.

Similarly, for $\Lambda_{uu}\in\{\Lambda_{11},\Lambda_{22}\}$, we have

$$\bar{\Psi}_{\Lambda_{uu}}=\frac{1}{2n}\Big\{tr\Big(\Omega^{-1}\frac{\partial\Omega}{\partial\Lambda_{uu}}\Big)+\frac{\partial tr(\Omega^{-1}\Omega_0)}{\partial\Lambda_{uu}}\Big\}$$

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

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

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

Further, applying Xiu (2010, Thm 2), $\hat{\Sigma}_{ls}-\bar{\Sigma}_{ls}=o_p(1)$ and $\hat{\Lambda}_{ll}-\bar{\Lambda}_{ll}=o_p(1)$, so consistency follows.

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

$$
\begin{aligned}
\frac{\partial\bar{\Psi}_\Sigma}{\partial\Sigma} &= \begin{pmatrix} \frac{\partial\bar{\Psi}_{\Sigma_{11}}}{\partial\Sigma_{11}} & \frac{\partial\bar{\Psi}_{\Sigma_{11}}}{\partial\Sigma_{12}} & \frac{\partial\bar{\Psi}_{\Sigma_{11}}}{\partial\Sigma_{22}} \\ \frac{\partial\bar{\Psi}_{\Sigma_{12}}}{\partial\Sigma_{11}} & \frac{\partial\bar{\Psi}_{\Sigma_{12}}}{\partial\Sigma_{12}} & \frac{\partial\bar{\Psi}_{\Sigma_{12}}}{\partial\Sigma_{22}} \\ \frac{\partial\bar{\Psi}_{\Sigma_{22}}}{\partial\Sigma_{11}} & \frac{\partial\bar{\Psi}_{\Sigma_{22}}}{\partial\Sigma_{12}} & \frac{\partial\bar{\Psi}_{\Sigma_{22}}}{\partial\Sigma_{22}} \end{pmatrix} \xrightarrow{\mathcal{P}} \frac{\partial\Psi_{\Sigma_0}}{\partial\Sigma},\quad \text{where}\\
\frac{\partial\bar{\Psi}_{\Sigma_{0,uv}}}{\partial\Sigma_{ij}} &= -\frac{T}{2}\Big(\int_0^\infty\frac{\partial\omega^{ij}(\Sigma_0,\Lambda_0,x)}{\partial\Sigma_{uv}}dx\Big)-\frac{T}{2}\Big(\int_0^\infty\frac{\partial\omega^{ij}(\Sigma_0,\Lambda_0,x)}{\partial\Sigma_{uv}}dx\Big)1_{\{u\neq v\}},
\end{aligned}
$$

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

$$
\begin{aligned}
n^{\frac{1}{4}}(\widehat{\Sigma}-\Sigma_0) &= n^{\frac{1}{4}}\begin{pmatrix} \widehat{\Sigma}_{11}-\frac{1}{T}\int_0^T\Sigma_{11,t}dt \\ \widehat{\Sigma}_{12}-\frac{1}{T}\int_0^T\Sigma_{12,t}dt \\ \widehat{\Sigma}_{22}-\frac{1}{T}\int_0^T\Sigma_{22,t}dt \end{pmatrix}\xrightarrow{\mathcal{L}_X} MN(0,V_Q),\quad \text{where}\\
\Pi_Q &= \frac{1}{4}\Big(\frac{\partial\bar{\Psi}_{\Sigma_0}}{\partial\Sigma}\Big)^{-1}\Big(\text{Avar}^{(2)}(\Sigma_0)+\text{Avar}^{(3)}(\Sigma_0)+\text{Avar}^{(4)}(\Sigma_0)\Big)\Big(\Big(\frac{\partial\bar{\Psi}_{\Sigma_0}}{\partial\Sigma}\Big)^{-1}\Big)'.
\end{aligned}
$$

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 $\widehat{\Lambda}$ follows immediately from (A.4). 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 $\Xi = \mathrm{diag}(\xi_1,\ldots,\xi_i,\ldots,\xi_n)$, and $\bar{\Omega}$ is the covariance matrix in the equidistant case with Δ replaced by $\bar{\Delta}$. It turns out that

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

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,

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

Because $E(\Xi)=0$,

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

Also,

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

where

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

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

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

where

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

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 $\bar{\Delta}\to 0$, which concludes the proof.

A.4 Proof of Theorem 3

Since $n_1 >> n_2$, we have:

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

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

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

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

$$\begin{aligned} u^{n_k}_{ij} &= \sqrt{\frac{2}{n_k+1}} \sin\Big(\frac{i \cdot j}{n_k+1}\pi\Big), \quad i,j = 1,\ldots,n_k, k = 1,2, \\ V_{ii} &= \mathrm{diag}(\mu_{ij}) = \mathrm{diag}\Big(\Sigma_{ii}\bar{\Delta}_i + 2\Lambda_{ii}\Big(1 - \cos\Big(\frac{j}{n_i+1}\pi\Big)\Big)\Big), \quad i = 1,2, \text{and } j = 1,\ldots,n_i, \\ V_{12} &= (v^{12}_{i,j}) = \Sigma_{12} U^{n_1} \Delta^{n_1,n_2} (U^{n_2})'. \end{aligned}$$

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

$$\begin{aligned} v^{12}_{i,j} =& \Sigma_{12}\bar{\Delta}_1 \sum_{k=1}^{n_1} u^{n_1}_{i,k} u^{n_2}_{j,[(k-1)/m]+1} = \Sigma_{12}\bar{\Delta}_1 \sum_{l=1}^{n_2} u^{n_2}_{j,l} \sum_{k=m(l-1)+1}^{ml} u^{n_1}_{i,k} \\ =& \sqrt{\frac{1}{(n_1+1)(n_2+1)}} \frac{\sin\frac{im\pi}{2(1+n_1)}}{\sin\frac{i\pi}{2(1+n_1)}} \Sigma_{12}\bar{\Delta}_1 \\ & \Big(\frac{\sin\frac{n_2(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)} \cos\frac{(i-j)\pi}{2}}{\sin\frac{(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)}} - \frac{\sin\frac{n_2(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)} \cos\frac{(i+j)\pi}{2}}{\sin\frac{(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)}}\Big). \end{aligned}$$

Moreover,

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

where by the Woodbury formula:

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

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

$$\frac{\partial \log(\det\Omega)}{\partial\theta_1} = tr\Big(\Omega^{-1}\frac{\partial\Omega}{\partial\theta_1}\Big) = tr\Big(V^{-1}\frac{\partial V}{\partial\theta_1}\Big), \quad \text{so}$$

$$E\Big(-\frac{\partial^2 L}{\partial\theta_1\partial\theta_2}\Big) \;=\; -\frac{1}{2}\frac{\partial^2\log(\det\Omega)}{\partial\theta_1\partial\theta_2} = -\frac{1}{2}\frac{\partial}{\partial\theta_2}tr\Big(V^{-1}\frac{\partial V}{\partial\theta_1}\Big).$$

Hence, our goal now is to calculate the following expressions, as $n_1, n_2 \longrightarrow \infty$.

$$tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{11}}\Big),\quad tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{12}}\Big),\quad \text{and}\quad tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{22}}\Big).$$

Since $n_1 >> n_2$, we have

$$\begin{aligned} tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{22}}\Big) =& \bar{\Delta}_2 tr\Big(\Big(V_{22}-V_{12}'V_{11}^{-1}V_{12}\Big)^{-1}\Big) = \bar{\Delta}_2 tr\Big(\Big(diag\Big(V_{22}-V_{12}'V_{11}^{-1}V_{12}\Big)\Big)^{-1}\Big)(1+o(1)) \\ =& \bar{\Delta}_2\sum_{j=1}^{n_2}\Big(\mu_{2j}-\sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-1}\Big)^{-1}(1+o(1)). \end{aligned}$$

Notice that

$$\begin{aligned} \sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-1} =& \frac{\Sigma_{12}^2\bar{\Delta}_1^2}{(n_1+1)(n_2+1)}\sum_{i=1}^{n_1}\left(\frac{\sin\frac{im\pi}{2(1+n_1)}}{\sin\frac{i\pi}{2(1+n_1)}}\right)^2 \\ & \left(\frac{\sin\frac{n_2(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)}\cos\frac{(i-j)\pi}{2}}{\sin\frac{(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)}} - \frac{\sin\frac{n_2(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)}\cos\frac{(i+j)\pi}{2}}{\sin\frac{(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)}}\right)^2\mu_{1i}^{-1} \end{aligned}$$

The dominant terms in the second brackets of the last summation come from is such that either

$$\sin\frac{(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)} \quad\text{or}\quad \sin\frac{(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)}$$

is close to 0. As $n_1 >> n_2$, for each $1\le j\le n_2^{1/2+\delta}$, there exists m different is $\in[1,n_1]$ such that

$$\sin\frac{(j+jn_1-im-in_1)\pi}{2(1+n_1)(1+n_2)}\approx 0,\quad \text{or}\quad \sin\frac{(j+jn_1+im+in_1)\pi}{2(1+n_1)(1+n_2)}\approx 0.$$

On the other hand,

$$\left(\frac{\sin\frac{im\pi}{2(1+n_1)}}{\sin\frac{i\pi}{2(1+n_1)}}\right)^2 \approx m^2,\ \text{when } i\in[1,n_2^{1/2+\delta}]$$

and decrease rapidly once $i > n_2$. Hence, the dominant term when $j\in[1,n_2^{1/2+\delta}]$ is the one with $i\approx j$, in which case

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

Therefore,

$$\sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-1} = n_1^{-1}n_2^{-1}\mu_{1j}^{-1}\Sigma_{12}^2T^2(1+o(1))$$

so as $n_1\to\infty$, $n_2\to\infty$ and $m\to\infty$,

$$\frac{1}{\sqrt{n_2}}tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{22}}\Big) = \frac{\bar{\Delta}_2}{\sqrt{n_2}}\sum_{j=1}^{n_2}\Big(\mu_{2j}-\sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-1}\Big)^{-1} \longrightarrow \int_0^\infty \frac{\Sigma_{11}T^2}{\Sigma_{11}T(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T^2}dx$$

Similarly, we can derive

$$\begin{aligned}
tr\Big(V^{-1}\frac{\partial V}{\partial \Sigma_{11}}\Big) =& \bar{\Delta}_1\left(tr\Big(V_{11}^{-1}\Big) + tr\Big(V_{11}^{-1}V_{12}\Big(V_{22}-V_{12}'V_{11}^{-1}V_{12}\Big)^{-1}V_{12}'V_{11}^{-1}\Big)\right)\\
=& \bar{\Delta}_1\left(\sum_{i=1}^{n_1}\mu_{1i}^{-1} + \sum_{j=1}^{n_2}\left(\Big(\mu_{2j}-\sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-1}\Big)^{-1}\sum_{i=1}^{n_1}(v_{i,j}^{12})^2\mu_{1i}^{-2}\right)\right)(1+o(1))\\
=& \left(\sqrt{n_2}\int_0^\infty \frac{\Sigma_{12}^2 T}{\Sigma_{11}(\Sigma_{11}(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T)}dx\right.\\
&\left. + \sqrt{n_1}\int_0^\infty \frac{T}{\Sigma_{11}T+\Lambda_{11}\pi^2x^2}dx\right)(1+o(1)), \quad \text{and}\\
tr\Big(V^{-1}\frac{\partial V}{\partial \Sigma_{12}}\Big) =& \frac{2}{\Sigma_{12}}tr\Big(-V_{11}^{-1}V_{12}(V_{22}-V_{12}'V_{11}^{-1}V_{12})^{-1}V_{12}'\Big)\\
=& \left(\int_0^\infty \frac{2\Sigma_{12}T}{\Sigma_{11}(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T}dx\right)\sqrt{n_2}(1+o(1)).
\end{aligned}$$

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

The Fisher information matrix can then be constructed as:

$$\begin{aligned}
I_{11}^\Sigma &= \lim_{n_1\to\infty,n_2\to\infty} -\frac{1}{2\sqrt{n_1}}\frac{\partial}{\partial\Sigma_{11}}tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{11}}\Big) = \int_0^\infty \frac{T^2}{2(\Sigma_{11}T+\Lambda_{11}\pi^2x^2)^2}dx,\\
I_{22}^\Sigma &= \lim_{n_1\to\infty,n_2\to\infty} -\frac{1}{2\sqrt{n_2}}\frac{\partial}{\partial\Sigma_{12}}tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{12}}\Big) = \int_0^\infty \frac{T(\Sigma_{11}(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)+\Sigma_{12}^2T)}{(\Sigma_{11}(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T)^2}dx,\\
I_{23}^\Sigma &= \lim_{n_1\to\infty,n_2\to\infty} -\frac{1}{2\sqrt{n_2}}\frac{\partial}{\partial\Sigma_{22}}tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{12}}\Big) = \int_0^\infty \frac{\Sigma_{11}\Sigma_{12}T^2}{(\Sigma_{11}(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T)^2}dx,\\
I_{33}^\Sigma &= \lim_{n_1\to\infty,n_2\to\infty} -\frac{1}{2\sqrt{n_2}}\frac{\partial}{\partial\Sigma_{22}}tr\Big(V^{-1}\frac{\partial V}{\partial\Sigma_{22}}\Big) = \int_0^\infty \frac{\Sigma_{11}^2T^4}{2(\Sigma_{11}T(\Sigma_{22}T+\Lambda_{22}\pi^2x^2)-\Sigma_{12}^2T^2)^2}dx.
\end{aligned}$$

Finally, $I_{12}^\Sigma = I_{13}^\Sigma = 0$. Hence, concluding the proof with

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

B Algorithms

Computing $\widehat{e}_{i|n}$, $\widehat{u}_{i|n}$, $D_{i|n}$ and $N_{i|n}$ is routine and rapid using the "disturbance smoother." This starts by running with the Kalman filter (e.g. Durbin and Koopman (2001, p. 67)), which is run forward in time $i = 1, 2, \ldots, n$ through the data. In our case it takes on the form $v_i = x_i - Z_i\widehat{y}_i$, $F_i = Z_i\left(P_i+\Lambda\right)Z_i'$, $K_i = P_iZ_i'F_i^{-1}$, $L_i = I - K_iZ_i$ then $\widehat{y}_{i+1} = \widehat{y}_i + K_iv_i$, $P_{i+1} = P_iL_i' + \Delta_{i+1}^n\Sigma$. Here $\widehat{y}_i = \mathrm{E}(y_i|x_{i:i-1})$ and $F_i = \mathrm{Cov}(x_i|x_{i:i-1})$. These recursions need some initial conditions $\widehat{y}_1$ and P_1. Throughout we will assume their choice does not depend upon Σ or Λ. A typical selection for $\widehat{y}_1$ is the opening auction price, whereas an alternative is to use a diffuse prior. Here v_i is $d_i \times 1$,

F_i is $d_i \times d_i$, K_i is $d \times d_i$, $\widehat{y}_{i+1|i}$ 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. The log-likelihood is $\log f(x_{1:n}; \Lambda, \Sigma) = c - \frac{1}{2}\sum_{i=1}^{n} \log |F_i| - \frac{1}{2}\sum_{i=1}^{n} v_i' F_i^{-1} v_i$.

The disturbance smoother (e.g. Durbin and Koopman (2001, p. 76)) is run backwards $i = n, n-1, ..., 1$ through the data. Writing $H_i = Z_i \Lambda Z_i'$, a $d_i \times d_i$ matrix, $\widehat{e}_{i|n} = H_i(F_i^{-1} v_i - K_i' r_i)$, $D_{i|n} = H_i - H_i(F_i^{-1} + K_i' M_i K_i) H_i$, $\widehat{u}_{i|n} = \Delta_i^n \Sigma r_{i-1}$, $N_{i|n} = \Delta_i^n \Sigma - (\Delta_i^n)^2 \Sigma M_{i-1} \Sigma$, where we recursively compute $r_{i-1} = Z_i' F_i^{-1} v_i + L_i' r_i$, $M_{i-1} = Z_i' F_i^{-1} Z_i + L_i' M_i L_i$, starting out with $r_n = 0$, $M_n = 0$. Here $\widehat{e}_{i|n}$ 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.