Journal of Econometrics 201 (2017) 19–42

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

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

Neil Shephard [a,b], Dacheng Xiu [c,*]

[a] Department of Economics, Harvard University, United States
[b] Department of Statistics, Harvard University, United States
[c] Booth School of Business, University of Chicago, United States

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form 3 December 2016
Accepted 6 April 2017
Available online 18 August 2017

JEL classification:
C01
C14
C58
D53
D81

Keywords:
EM algorithm
Kalman filter
Market microstructure noise
Asynchronous data
Factor model
Portfolio allocation
Quasi-likelihood
Semimartingale

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 maximum 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 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. We also propose an alternative estimator using a factor model, which scales well with the number of assets. We derive the theoretical properties of these estimators and prove that they achieve the efficient rate of convergence. Our estimators are also analysed using Monte Carlo methods and applied to equity data with varying levels of liquidity.

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 definite and deals with both market microstructure effects such as bid/ask bounce and crucially asynchronous recording of data (the Epps (1979) effect). Positive definiteness allows us to define a coherent estimator of correlations and betas, objects of importance in financial economics. Our likelihood approach can accommodate additional assumptions of the data, such as a factor model, which scales well with the number of assets. We derive the theoretical properties of the QML estimators with or without using a factor model, and prove that both achieve the efficient rate of convergence. The estimators are also analysed using Monte Carlo methods and applied on equity data. Our results show our method delivers 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.

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 et al. (2001b) and Barndorff-Nielsen and Shephard (2002). There the likelihood is generated by assuming the log-price is a Brownian motion.

* Corresponding author.
E-mail addresses: shephard@fas.harvard.edu (N. Shephard), dacheng.xiu@chicagobooth.edu (D. Xiu).

http://dx.doi.org/10.1016/j.jeconom.2017.04.003
0304-4076/

Multivariate versions of these estimators were developed in Andersen et al. (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 asynchronous 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 realised kernel of Barndorff-Nielsen et al. (2008) but with a suboptimal bandwidth. Note the related Zhou (1996), Zhou (1998), Andersen et al. (2001a) and Hansen et al. (2008).

Our paper moves beyond this work, producing distinctive and empirically important results. It proposes two estimators and develops the asymptotic theory in a setting when covariances are stochastic. The simulation results also demonstrate the desired consistency of the estimators when applied to asynchronous data.

1.3. Alternative approaches

A number of authors have approached this sophisticated problem using other 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 et al. (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 asynchronous trading. Related is Hayashi and Yoshida (2005) who tried to overcome asynchronous 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 asynchronous trading become important issues. There are several existing methods which have been proposed. Only two deliver positive definite estimators, so allowing correlations and betas to be coherently computed. They are the multivariate realised kernel of Barndorff-Nielsen et al. (2011) and the non-bias-corrected preaveraging estimator of Christensen et al. (2010). Both use a synchronisation device called refresh time sampling. Neither converges at the optimal rate.

In addition, Zhang (2011) and Bibinger (2012) extend the two-scale and multi-scale volatility estimators to the covariance estimation, respectively. Christensen et al. (2013) and Koike (2015) combine the pre-averaging estimator of Christensen et al. (2010) with the Hayashi–Yoshida method for synchronisation. Park et al. (2016) develop a Fourier based estimator of covariances, extending Mancino and Sanfelici (2011) and Sanfelici and Mancino (2008). These estimators do not necessarily give positive definite estimates. None of them achieve the efficiency bound when covariance matrix is constant. Bibinger et al. (2014) and Altmeyer and Bibinger (2015) propose a local method of moments estimator that achieves the nonparametric efficiency bound, however the resulting estimator is not positive definite. To obtain positive definiteness for a flat-top kernel estimator, Varneskov (2016) proposes a finite sample correction based on projections of symmetric matrices. The aforementioned methods all require certain choice of tuning parameters, which could be cumbersome for users.

Fan et al. (2016) and Aït-Sahalia and Xiu (2016b) develop consistent estimators of a large covariance matrix based on factor models, using regression and principal component analysis, respectively, in the absence of microstructure noise. They establish the consistency of these estimators using a joint in-fill and increasing dimensionality asymptotics. In contrast, this paper develops the central limit theory of a factor-based covariance matrix in the presence of noise using in-fill asymptotics alone.

Finally, we note that related univariate work on ameliorating the effect of noise includes Zhou (1996), Zhou (1998), Hansen and Lunde (2006), Zhang et al. (2005), Barndorff-Nielsen et al. (2008), Jacod et al. (2009)), Bandi and Russell (2008), Kalnina and Linton (2008), Jacod et al. (2010), 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 et al. (2012) and Hansen and Horel (2009). Surveys include, for example, McAleer and Medeiros (2008), Park and Linton (2012), 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 motions 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, and how this approach can be adapted to allow a factor structure.

The above procedure can be implemented computationally efficiently using Kalman filtering, optimising the likelihood via an EM algorithm. The resulting estimator of the integrated covariance is positive definite. In practice it can be computed rapidly, even in significant dimensions when a factor model is imposed.

1.5. Some particularly noteworthy papers

There are a group of papers which are closest to our approach. Aït-Sahalia et al. (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 $\mathrm{Var}(x_1 + x_2)$ and $\mathrm{Var}(x_1 - x_2)$. The implied covariance matrix is not guaranteed to be positive definite.

Corsi et al. (2015) 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. We also provide a factor-model-based estimator to facilitate a potentially large dimension. However, the overarching theme is the same: using a missing value approach based on Brownian motion observed with error. Further, Peluso et al. (2012) extend 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). Liu and Tang (2014) also study a realised QML estimator of a multivariate synchronised dataset. They propose using refresh time type devices to achieve exact synchronicity. They do not obtain an explicit asymptotic distribution of the estimator either.

1.6. Structure of the paper

The structure of our paper is as follows. In Section 2 we define the models. In Section 3, we propose the quasi-likelihood estimator and discuss the implementation. In Section 4 we derive the estimator's asymptotic properties. In Section 5 we extend the core results in various important directions. In Section 6 we assess the finite sample performance of our estimators using some Monte Carlo experiments. In Section 7 we provide results from empirical studies, where the performance of the estimators is evaluated with a variety of equity prices. In Section 8 we draw our conclusions. The paper finishes with a lengthy appendix which contains the proofs of various theorems given in the paper and the detailed algorithm for implementation.

2. Models

2.1. Notation

We use $\otimes$ to denote the Kronecker product, and $^+$ to denote the Moore–Penrose inverse of a matrix. We use $\mathbb{I}_m$ to denote an m-dimensional identity matrix. For any $m \times n$ matrix A, $\text{vec}(A) = (A_{11}, A_{21}, \ldots, A_{m1}, A_{12}, A_{22}, \ldots, A_{m2}, \ldots, A_{1n}, A_{2n}, \ldots, A_{mn})^\intercal$. For any $d \times d$ symmetric matrix B, we use $\text{vech}(B)$ to denote the $d(d+1)/2 \times 1$ vector that is obtained from $\text{vec}(B)$ by eliminating all supradiagonal elements of B. The operator diag is defined as $\text{diag}(B) = (B_{11}, B_{22}, \ldots, B_{dd})^\intercal$. In addition, we define its inverse operator Diag, which maps a vector of length d to a $d \times d$ diagonal matrix.

We consider a d-dimensional log-price process x, observed irregularly and asynchronously over the interval $[0, T]$, where T is fixed and thought of as a single day or a single week. Prices will be referred to as trades, but our methods also apply to quotes. We write the union of all calendar times of trades as t_i, $1, 2, \ldots, n$, where we have ordered the distinct times so that $0 \le t_1 < \cdots < t_i < \cdots < t_n \le T$. We denote the observed log-price at t_i as x_i. 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 $1 \le d_i \le d$. Then Z_i is $d_i \times d$, full of zeros and ones where each row sums exactly to one. A unit element in column k of Z_i shows the kth asset traded at time t_i.

2.2. A nonparametric model

$\{x_j\}_{j=0}^n$ is assumed to be driven by $\{y_t\}_{0\le t\le T}$, the efficient log-price, abstracting from market microstructure effects. The efficient price is modelled as a *Brownian semimartingale* defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$,

$$y_t = \int_0^t \mu_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u, \quad \Sigma_t = \sigma_t\sigma_t^\intercal > 0, \tag{1}$$

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

$$[y, y]_T = \int_0^T \Sigma_u \mathrm{d}u, \quad \text{where } [y, y]_T = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^n \left\{y_{t_j} - y_{t_{j-1}}\right\}\left\{y_{t_j} - y_{t_{j-1}}\right\}^\intercal,$$

(e.g. Protter (2004, p. 66–77)) for any sequence of deterministic partitions $0 = t_0 < t_1 < \cdots < t_n = T$ with $\sup_j\{t_{j+1} - t_j\} \to 0$ for $n \to \infty$. Our interest is in estimating $[y, y]_T$ using $\{x_j\}_{j=0}^n$.

2.3. A semiparametric factor model

The arbitrage pricing theory by Ross (1976) provides an economic motivation for the presence of a factor structure in asset returns. When estimating a large covariance matrix, we can impose such a model:

$$y_t = \beta y_t^f + y_t^e, \tag{2}$$

where y_t^f denotes a $f \times 1$ vector of factors at time t, y_t^e, a $d \times 1$ vector, denotes the idiosyncratic component of each asset, and β is a constant $d \times f$ factor loading matrix. The constant β assumption is not restrictive, given that the time horizon we consider, $[0, T]$, is only a day or a week. y_t^f and y_t^e satisfy the same model as in (1), with different drift and diffusive components. Their instantaneous covariation matrices at time t are denoted by Σ_t^f and Σ_t^e, respectively.

There is no fundamental difference between factor models in continuous time and discrete time. Similar to the conventional factor model for low-frequency data, the relationship between y_t and its factor y_t^f is spelt out in a linear regression model (2). Fan et al. (2016) and Aït-Sahalia and Xiu (2016b) adopt this model for high frequency data and find evidence in support of such a structure.

The implication of a high frequency factor model is that, the covariation of y_t on $[0, T]$ can be decomposed as

$$\int_0^T \Sigma_u du = \beta\left(\int_0^T \Sigma_u^f du\right)\beta^\intercal + \int_0^T \Sigma_u^e du.$$

We are interested in estimating $\int_0^T \Sigma_u du$, as well as its components $\int_0^T \Sigma_u^f du$, $\int_0^T \Sigma_u^e du$, and β. For parsimony, we consider a strict factor model, that is, Σ_u^e is diagonal, as opposed to the approximate factor model introduced in Chamberlain and Rothschild (1983). The use of a strict factor model reduces the number of parameters in the covariance matrix from $\frac{1}{2}d \times (d+1)$ in the fully nonparametric model to $d \times f + \frac{1}{2}f \times (f+1) + d$. When d is much larger than f, this tremendously simplifies the estimation problem and alleviates the curse of dimensionality.

2.4. Relating efficient prices to observed data and times

We assume that $\{y_t\}$ 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 et al. (2014)). 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_t, with our data being

$$x_i = Z_i y_{t_i} + Z_i\varepsilon_i, \quad i = 1, 2, \ldots, n, \quad \mathrm{E}(\varepsilon_i) = 0, \quad \mathrm{Cov}(\varepsilon_i) = \Lambda, \tag{3}$$

where ε_i is a vector of potential market microstructure effects and Λ is diagonal and non-singular.

Fama and French (1993) introduce a widely used three-factor model, where they build portfolios to mimic the market, the value, and the size factors, which are shown to drive the dynamics of stock returns empirically. There is a large literature devoted to the search for additional factors, many of which are constructed using portfolios, see, e.g., Harvey et al. (2016). We summarise these portfolio proxies as y_t^f. In the high-frequency setting, however, we only observe a blurred version of these variables at t_i, denoted by x_i^f, due to the presence of microstructure noise ε^f. That is,

$$x_i^f = y_{t_i}^f + \varepsilon_i^f \tag{4}$$

where ε^f is i.i.d. with mean 0 and a $f \times f$ diagonal covariance matrix Λ^f.

Except for some observation error due to the microstructure noise, our factors are in fact "observable" as opposed to the "latent" factors discussed in, e.g., Stock and Watson (2002) and Bai and Ng (2002). Our approach is directly applicable to latent factors, provided sufficient identification restrictions, see, e.g., Anderson and Amemiya (1988). The asymptotic distribution of these estimators is typically non-tractable with the cross-sectional dimension being fixed. Therefore, we focus on observable factors in this paper. Recently, Bai and Li (2012) study likelihood based estimators of large factor models and obtain tractable asymptotic distributions using a joint large n and large d asymptotic design. Extensions of our likelihood approach to this increasing dimensionality setting are left for future work.

3. Estimators

We build the estimators using a quasi-likelihood approach, see, e.g., White (1982), based on a simplified and misspecified model in which all Brownian semimartingales are proxied by Brownian motions with constant covariance matrices. The estimators are therefore easy to construct and implement.

3.1. Gaussian state-space models

We first consider the nonparametric case. The simplified model assumes that $y_t = \sigma W_t$, and then writing $\Sigma = \sigma\sigma^\intercal$, we have that

$$x_i = Z_i y_{t_i} + Z_i \varepsilon_i, \quad y_{t_i} - y_{t_{i-1}} = u_i \sim \mathcal{N}(0, \Sigma(t_i - t_{i-1})). \tag{5}$$

At this point, we also assume that $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \Lambda)$. Therefore, the time series of observations $x_{1:n} = (x_1, \ldots, x_n)^\intercal$ is a Gaussian state space model (e.g. Durbin and Koopman, 2001). The parameters to be estimated include Σ and Λ.

In the semiparametric factor model, assuming Brownian motions with constant covariance matrices and normality of ε and ε^f, the simplified model remains a Gaussian state space model:

$$\begin{pmatrix} x_i \\ x_i^f \end{pmatrix} = \begin{pmatrix} Z_i & \\ & Z_i^f \end{pmatrix} \begin{pmatrix} y_{t_i} \\ y_{t_i}^f \end{pmatrix} + \begin{pmatrix} Z_i & \\ & Z_i^f \end{pmatrix} \begin{pmatrix} \varepsilon_i \\ \varepsilon_i^f \end{pmatrix}, \tag{6}$$

$$\begin{pmatrix} y_{t_i} \\ y_{t_i}^f \end{pmatrix} = \begin{pmatrix} y_{t_{i-1}} \\ y_{t_{i-1}}^f \end{pmatrix} + \begin{pmatrix} \mathbb{I}_d & \beta \\ 0 & \mathbb{I}_f \end{pmatrix} \begin{pmatrix} u_i^e \\ u_i^f \end{pmatrix}, \tag{7}$$

where $u_i^e = y_{t_i}^e - y_{t_{i-1}}^e \sim \mathcal{N}(0, \Sigma^e(t_i - t_{i-1}))$, and $u_i^f = y_{t_i}^f - y_{t_{i-1}}^f \sim \mathcal{N}(0, \Sigma^f(t_i - t_{i-1}))$. The parameters in this model are Σ^f, β, and Σ^e, hence $\Sigma = \beta\Sigma^f\beta^\intercal + \Sigma^e$.

3.2. ML estimation via EM algorithm

We will develop a positive definite estimator of Σ in both models. We aim for a method that works in high dimensions, so the EM approach to maximising the log-likelihood function is attractive. The EM algorithm is a well-known tool for iterative maximum likelihood estimation. It consists of an E-step (expectation) and M-step (maximisation). For the state space model, it has a particular neat form (see, e.g., Durbin and Koopman (2001, Ch. 7.3.4)).

We start with model (5). Note that the complete log-likelihood is, writing $e_i = x_i - Z_i y_{t_i}$, $u_i = y_{t_i} - y_{t_{i-1}}$, $y_{1:n} = \left(y_{t_1}, y_{t_2}, \ldots, y_{t_n}\right)^\intercal$, $\Delta_i = t_i - t_{i-1}$, and initialising $y_{t_1} \sim \mathcal{N}(x_1, \Lambda)$,[1]

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

Then the EM algorithm works with the

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

Writing $\widehat{e}_{i|n} = \mathrm{E}\left(e_i|x_{1:n}\right)$ and $D_{i|n} = \mathrm{Var}\left(e_i|x_{1:n}\right)$, we have

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

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

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

Therefore, the EM update is

$$\widehat{\Sigma} = \frac{1}{n-1}\sum_{i=2}^n \frac{1}{\Delta_i}\left\{\widehat{u}_{i|n}\widehat{u}_{i|n}^\intercal + N_{i|n}\right\},$$

$$\mathrm{diag}\left(\widehat{\Lambda}\right) = \left(\sum_{i=1}^n Z_i^\intercal Z_i\right)^{-1}\mathrm{diag}\left(\sum_{i=1}^n Z_i^\intercal\left\{\widehat{e}_{i|n}\widehat{e}_{i|n}^\intercal + D_{i|n}\right\}Z_i\right).$$

Iterating these updates, the sequence of $(\widehat{\Sigma}, \widehat{\Lambda})$ converges to a maximum in the likelihood function. Appendix B 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.

For the model specified in (6) and (7), the algorithm is similar. We write

$$\bar{x} = \begin{pmatrix} x \\ x^f \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} y \\ y^f \end{pmatrix}, \quad \bar{\varepsilon} = \begin{pmatrix} \varepsilon \\ \varepsilon^f \end{pmatrix}, \quad \bar{Z} = (Z, Z^f),$$

$$\bar{\Lambda} = \begin{pmatrix} \Lambda & \\ & \Lambda^f \end{pmatrix}, \quad \bar{u} = \begin{pmatrix} u \\ u^f \end{pmatrix},$$

$$\text{and} \quad \bar{\Sigma} := \begin{pmatrix} \bar{\Sigma}^{(1,1)} & \bar{\Sigma}^{(1,2)} \\ \bar{\Sigma}^{(2,1)} & \bar{\Sigma}^{(2,2)} \end{pmatrix} = \begin{pmatrix} \Sigma^e + \beta\Sigma^f\beta^\intercal & \beta\Sigma^f \\ \Sigma^f\beta^\intercal & \Sigma^f \end{pmatrix}. \tag{8}$$

Using Theorems 2.2, 3.9, 8.2 and 8.3, Equation (4) of Section 2.4, Equation (1) of Section 3.7 in Magnus and Neudecker (1999), and taking first order derivatives of

$$-\frac{1}{2}\sum_{i=2}^n \log|\bar{\Sigma}| - \frac{1}{2}\sum_{i=2}^n \frac{1}{\Delta_i} tr\left[\bar{\Sigma}^{-1}\left\{\widehat{\bar{u}}_{i|n}\widehat{\bar{u}}_{i|n}^\intercal + \bar{N}_{i|n}\right\}\right]$$

with respect to Σ^f, $\mathrm{diag}(\Sigma^e)$, and β, respectively, we have

$$\begin{aligned}
&\text{wrt } \mathrm{diag}(\Sigma^e): \quad \mathrm{diag}\left(\left(\Sigma^e\right)^{-1} - \left(\Sigma^e\right)^{-1}\check{\Sigma}^{(1,1)}\left(\Sigma^e\right)^{-1}\right. \\
&\qquad \left. + \left(\Sigma^e\right)^{-1}\check{\Sigma}^{(1,2)}\left(\check{\Sigma}^{(2,2)}\right)^{-1}\check{\Sigma}^{(2,1)}\left(\Sigma^e\right)^{-1}\right) = 0, \\
&\text{wrt } \Sigma^f: \quad \left(\beta^\intercal : \mathbb{I}_f\right)\left(\bar{\Sigma}^{-1} - \bar{\Sigma}^{-1}\check{\Sigma}\bar{\Sigma}^{-1}\right)\left(\beta^\intercal : \mathbb{I}_f\right)^\intercal = 0, \\
&\text{wrt } \beta: \quad \left(\mathbb{I}_d : 0_{d\times f}\right)\left(\bar{\Sigma}^{-1} - \bar{\Sigma}^{-1}\check{\Sigma}\bar{\Sigma}^{-1}\right)\left(\Sigma^f\beta^\intercal : \Sigma^f\right)^\intercal = 0,
\end{aligned}$$

where

$$\check{\Sigma} = \frac{1}{n-1}\sum_{i=2}^n \frac{1}{\Delta_i}\left\{\widehat{\bar{u}}_{i|n}\widehat{\bar{u}}_{i|n}^\intercal + \bar{N}_{i|n}\right\},$$

and $\widehat{\bar{u}}_{i|n}$, $\bar{D}_{i|n}$ and $\bar{N}_{i|n}$ are defined similarly as $\widehat{u}_{i|n}$, $D_{i|n}$ and $N_{i|n}$ but for the stacked variables. Solving the first order conditions leads to the EM iteration:

$$\widetilde{\beta} = \check{\Sigma}^{(1,2)}\left(\check{\Sigma}^{(2,2)}\right)^{-1}, \quad \widetilde{\Sigma}^f = \check{\Sigma}^{(2,2)},$$

$$\widetilde{\Sigma}^e = \mathrm{diag}\left(\check{\Sigma}^{(1,1)} - \check{\Sigma}^{(1,2)}\left(\check{\Sigma}^{(2,2)}\right)^{-1}\check{\Sigma}^{(2,1)}\right),$$

$$\widehat{\bar{\Sigma}} = \begin{pmatrix} \widetilde{\Sigma}^e + \widetilde{\beta}\widetilde{\Sigma}^f\widetilde{\beta}^\intercal & \widetilde{\beta}\widetilde{\Sigma}^f \\ \widetilde{\Sigma}^f\widetilde{\beta}^\intercal & \widetilde{\Sigma}^f \end{pmatrix},$$

$$\mathrm{diag}\left(\widehat{\bar{\Lambda}}\right) = \left(\sum_{i=1}^n \bar{Z}_i^\intercal\bar{Z}_i\right)^{-1}\mathrm{diag}\left(\sum_{i=1}^n \bar{Z}_i^\intercal\left\{\widehat{\bar{e}}_{i|n}\widehat{\bar{e}}_{i|n}^\intercal + \bar{D}_{i|n}\right\}\bar{Z}_i\right).$$

---

[1] We assume the opening prices are available for all assets. This initialisation amounts to imposing a diffuse prior on y_{t_1}, see, e.g., Durbin and Koopman (2001).

Most researchers have analysed covariances by applying a synchronisation scheme to the asynchronous data. The most well known such scheme is the refresh time method analysed by Barndorff-Nielsen et al. (2011) and subsequently employed by, for example, Christensen et al. (2010) and Aït-Sahalia et al. (2010). Alternatively, Zhang (2011) discusses the Previous Tick approach, which always discards more data than the refresh time approach.

Typically there is a very large drop in the sample size due to synchronisation. In contrast, we treat the problem of asynchronicity as a missing data problem by the use of the selection matrix Z, hence the entire dataset is used for estimation. Ansley and Kohn (1983) suggest the use of Kalman filter to construct the exact likelihood of vector-autoregressive-moving average process with missing data. See also Wu (1983) for a more general discussion on the convergence of the EM.

4. Econometric theory

Here we develop the asymptotic theory for the quasi-likelihood estimators in the previous section. We consider the usual in-fill asymptotic design with n increasing to ∞ and T being fixed. We also fix d and f. We will establish the desired consistency and central limit theory under correctly specified models.

Throughout this section, we assume that the trades are synchronised at times $t_i = Ti/n$, $Z_i = \mathbb{I}_n$, and $\Delta = T/n$. This is an unfortunate but necessary simplification to derive some useful theoretical insight. Under certain assumptions, e.g., Assumption 3.1 of Bibinger et al. (2014), the asynchronicity does not affect the asymptotic distribution of their local method of moment estimator. See also Altmeyer and Bibinger (2015) and a similar result for the pre-averaging estimator in Koike (2015). The intuition for this is that the presence of noise requires smoothing within a local window of size shrinking to 0, which helps smooth out the asynchronicity bias.

We would expect the same lack of asynchronicity bias to hold for our case, but the derivation of the asymptotic distribution in a general asynchronous setting is rather cumbersome for likelihood based estimators due to the complex form of the likelihood function. As a result, there are (unfortunately) no formal results about the irrelevance of asynchronicity for QML-type estimators. That said, we compare the performance of these estimators in a simulation setting where asynchronicity is present. Our estimator does outperform the alternatives that use refresh time sampling.

4.1. Benchmark MLE

We start with the case where the true model has a constant covariance matrix. We also assume the market microstructure effects are i.i.d. normal. In this case, our QML estimator becomes the MLE. Bibinger et al. (2014) develop the efficiency bound in this setting using asymptotic equivalence in Le Cam's sense. Our MLE achieves this bound.

More specifically, we observe returns

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

with,

$$x_i = y_{\tau_i} + \varepsilon_i, \quad i = 0, 1, 2, \dots, n, \quad y_{\tau_i} = y_{\tau_{i-1}} + u_i, \tag{9}$$

where, writing $\Delta = T/n$,

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\begin{pmatrix} \Lambda & 0 \\ 0 & \Delta\Sigma \end{pmatrix}. \tag{10}$$

Throughout we consider only the case where Σ and Λ are non-degenerate, i.e., $\Sigma > 0$ and $\Lambda > 0$.

The observed returns are $r = (r_{1,1}, r_{1,2}, \dots, r_{1,n}, r_{2,1}, \dots, r_{2,n}, \dots, r_{d,1}, \dots, r_{d,n})^\intercal$, so the likelihood is

$$L_n = -n\log(2\pi) - \frac{1}{2}\log(\det\Omega) - \frac{1}{2}r^\intercal\Omega^{-1}r, \tag{11}$$

where $\Omega = \Delta\Sigma \otimes \mathbb{I}_n + \Lambda \otimes \mathbb{J}_n$. Here $\mathbb{J}_n$ is a $n \times n$ matrix

$$\mathbb{J}_n = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{12}$$

The likelihood (11) is tractable as we know the eigenvalues and eigenvectors of $\mathbb{J}_n$ and so Ω.

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

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

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

Theorem 1. *Assume the model (9)–(10) is true. Then the ML estimator $\widehat{\Sigma}$ satisfies the central limit theorem as $n \to \infty$,*

$$n^{1/4}\left(\mathrm{vech}\left(\widehat{\Sigma}\right) - \mathrm{vech}\left(\Sigma\right)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Pi),$$

where, writing $\mathbb{D}_d$ as the $d^2 \times d(d+1)/2$ duplication matrix, see, e.g., Magnus and Neudecker (1999),

$$\Pi = 4T^{-1/2}\mathbb{D}_d^+\left\{\Sigma \otimes \left(\Lambda^{1/2}\left(\Lambda^{-1/2}\Sigma\Lambda^{-1/2}\right)^{1/2}\Lambda^{1/2}\right) + \left(\Lambda^{1/2}\left(\Lambda^{-1/2}\Sigma\Lambda^{-1/2}\right)^{1/2}\Lambda^{1/2}\right) \otimes \Sigma\right\}\mathbb{D}_d^{+\intercal}.$$

In particular, for any $1 \le p, q \le d$, we have

$$n^{1/4}\left(\widehat{\Sigma}_{pq} - \Sigma_{pq}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4T^{-1/2}\left(\Lambda_{qq}^{1/2}\Sigma_{pp}(\Sigma^{1/2})_{qq} + \Lambda_{pp}^{1/2}\Sigma_{qq}(\Sigma^{1/2})_{pp}\right)\right).$$

Proof. Given in Appendix A.1.

In the bivariate case, suppose the correlation between two asset returns is ρ. The asymptotic variance of $\widehat{\Sigma}_{11}$ is given by $4T^{-1/2}\Lambda_{11}^{1/2}\Sigma_{11}^{3/2}(\sqrt{1-\rho} + \sqrt{1+\rho})$, a decreasing function of $|\rho|$. When $\rho = 0$, $(\Sigma^{1/2})_{pp} = \Sigma_{pp}^{1/2}$, for $p = 1, 2$, then there is no "externality", i.e., the asymptotic variance for $\widehat{\Sigma}_{11}$ in the bivariate case reproduces the univariate case. As ρ increases from 0 to 1, $\widehat{\Sigma}_{11}$ becomes more efficient than the univariate alternative, as more information is collected via the correlation with the other asset. Now suppose that the two assets have the same variance, i.e., $\Sigma_{11} = \Sigma_{22}$. As $\rho \to 1$, the asymptotic variance of $\widehat{\Sigma}_{11}$ in the bivariate case approaches to $1/\sqrt{2}$ times of that in the univariate case, because the two assets are identical and hence the effective sample size doubles.

4.2. Nonparametric multivariate QML estimator

We now move to more realistic case, in which stochastic volatility is present.

Assumption 1. The underlying latent d-dimensional log-price process satisfies (1), where μ_t is predictable locally bounded, σ_t is a locally bounded Itô semimartingale and W_t is a Brownian motion.

Assumption 2. The noise ε_i in (3) is i.i.d., and independent of the stock price, and has fourth moments.[2] The cross-sectional covariance matrix of ε is diagonal.

Define $R_T = \left(\frac{1}{T}\int_0^T \sigma_u^4 \mathrm{d}u\right)/\left(\frac{1}{T}\int_0^T \sigma_u^2 \mathrm{d}u\right)^2 \geq 1$, by Cauchy–Schwarz inequality, and recall the result

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

by rewriting (Xiu, 2010). $\mathcal{MN}$ denotes a mixture of normal, as the asymptotic variance is random. The asymptotic distribution is defined on an extension of the original probability space, which is conditionally normal with respect to the filtration $\mathcal{F}$ generated by y_t and σ_t. The asymptotic variance increases with R_T keeping $\frac{1}{T}\int_0^T \sigma_u^2 \mathrm{d}u$ fixed.

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

Theorem 2. *Under Assumptions 1–2, we have*

$$n^{1/4}\left(\mathrm{vech}\left(\widehat{\Sigma}\right) - \mathrm{vech}\left(\frac{1}{T}\int_0^T \Sigma_u du\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^Q\right),$$

where the explicit expression of Π^Q is given by (A.19) in the appendix.

Proof. Given in Appendix A.2.

If the covariance matrix Σ_t is constant, the asymptotic covariance matrix satisfies $\Pi^Q = \Pi$. In general, the asymptotic covariance matrix is complicated due to the complexity of the problem, hence is left in the appendix, yet it is easily implementable (Mathematica will solve the integrals in closed forms). The asymptotic covariance matrix involves integrated quarticity and noise variances, which can be estimated using the estimators in Aït-Sahalia and Xiu (2016a).

4.3. Semiparametric multivariate QML estimator

We now provide the asymptotic theory for the semiparametric estimator. Because this estimator is built on a factor model, we refer to it as the FQML estimator below. We assume

Assumption 3. The underlying latent d-dimensional log-price process satisfies (2), where y^f and y^e are continuous Itô semimartingales similar to (1).

Assumption 4. The noise ε_i^f in (4) is i.i.d., and independent of the stock price, and has fourth moments. Its covariance matrix Λ^f is diagonal.

Using the notation in (8), the likelihood function can be written as:

$$\bar{L}_n = -n\log(2\pi) - \frac{1}{2}\log(\det\bar{\Omega}) - \frac{1}{2}\bar{r}^{\intercal}\bar{\Omega}^{-1}\bar{r}, \tag{13}$$

where $\bar{r} = (\bar{r}_{1,1}, \bar{r}_{1,2}, \ldots, \bar{r}_{1,n}, \bar{r}_{2,1}, \ldots, \bar{r}_{2,n}, \ldots, \bar{r}_{d+f,1}, \ldots, \bar{r}_{d+f,n})^{\intercal}$, $\bar{r}_{j,i} = \bar{x}_{j,i} - \bar{x}_{j,i-1}$, for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, d+f$, and $\bar{\Omega} = \Delta\bar{\Sigma}\otimes\mathbb{I}_n + \bar{\Lambda}\otimes\mathbb{J}_n$.

We can establish the following asymptotic result:

Theorem 3. *Under Assumptions 3–4, we have*

$$n^{1/4}\left(\mathrm{vec}\left(\widetilde{\beta}\right) - \mathrm{vec}\left(\beta\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^{\beta}\right),$$

$$n^{1/4}\left(\mathrm{vech}\left(\widetilde{\Sigma}^f\right) - \mathrm{vech}\left(\frac{1}{T}\int_0^T \Sigma_u^f du\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^f\right),$$

$$n^{1/4}\left(\mathrm{diag}\left(\widetilde{\Sigma}^e\right) - \mathrm{diag}\left(\frac{1}{T}\int_0^T \Sigma_u^e du\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^e\right),$$

where the explicit expressions of Π^{β}, Π^f, and Π^e are given by (A.26) in the appendix. Therefore, the FQML estimator $\widetilde{\Sigma} = \widetilde{\beta}\widetilde{\Sigma}^f\widetilde{\beta}^{\intercal} + \widetilde{\Sigma}^e$ satisfies:

$$n^{1/4}\left(\mathrm{vech}\left(\widetilde{\Sigma}\right) - \mathrm{vech}\left(\frac{1}{T}\int_0^T \Sigma_u du\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^F\right),$$

where Π^F is given by (A.27) in the appendix.

Proof. Given in Appendix A.3.

The semiparametric estimator $\widetilde{\Sigma}$ achieves a smaller standard error than the nonparametric estimator $\widehat{\Sigma}$ when Assumption 3 holds, and is more reliable in finite sample as shown in simulations of Section 6.

To illustrate the intuition of the efficiency gain from a factor model, we consider the case with $d = f = 1$, in which a two-dimensional nonparametric estimator is equivalent to a factor model based estimator, since they both have exactly three independent parameters. Therefore, comparing the efficiency between a univariate nonparametric volatility estimator and a factor model based estimator is amount to comparing the asymptotic variances between the univariate estimator and the $(1, 1)$ entry of a two-dimensional nonparametric covariance matrix estimator. We have learned in Section 4.1 that the latter estimator is more efficient as long as the correlation between the factor and this asset is nonzero. Moreover, as the magnitude of this correlation increases, the efficiency gain from a factor model based estimator becomes larger.

5. Additional developments

5.1. Realised QML correlation and regression estimator

An immediate extension for the nonparametric QML estimator is the correlation version

$$\widehat{\rho}_{ij} = \frac{\widehat{\Sigma}_{ij}}{\sqrt{\widehat{\Sigma}_{ii}\widehat{\Sigma}_{jj}}} \in [-1, 1],$$

$$\text{of}\quad \rho_{ij}^{\star} = \frac{\frac{1}{T}\int_0^T \Sigma_{ji,u}\mathrm{d}u}{\sqrt{\left(\frac{1}{T}\int_0^T \Sigma_{ii,u}\mathrm{d}u\right)\left(\frac{1}{T}\int_0^T \Sigma_{jj,u}\mathrm{d}u\right)}} \in [-1, 1].$$

Also of importance is the corresponding realised market beta. In a CAPM model, with the second asset being the market portfolio, the market beta for the first asset can be defined as

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

The limit theory follows immediately by the delta method. These are noise and asynchronous trading robust versions of the realised quantities studied by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004).

In a factor model, Theorem 3 provides an estimator for factor betas, in which the market beta (the coefficient corresponding to a market portfolio proxy in the regression) differs from the realised

[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 et al. (2010). We choose not to adopt it as the notation would be rather complex and the idea of the proof remains the same except for some technicalities.

market beta above, in that the former uses a regression model to control other factors. We can construct another multivariate beta estimator using the nonparametric QML, or any alternative nonparametric estimators of the covariance matrix. We stack returns of assets and factors together as in Section 3.2, and divide any nonparametric estimator $\widehat{\Sigma}$ into 2×2 blocks:

$$\widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}^{(1,1)} & \widehat{\Sigma}^{(1,2)} \\ \widehat{\Sigma}^{(2,1)} & \widehat{\Sigma}^{(2,2)} \end{pmatrix}.$$

The factor beta can be estimated as:

$$\widehat{\beta} = \widehat{\Sigma}^{(1,2)}\left(\widehat{\Sigma}^{(2,2)}\right)^{-1}.$$

Such an estimator is not as efficient as the semiparametric estimator $\widetilde{\beta}$ we propose in Section 4.3, when a factor model (2) is assumed. The central limit theorem of $\widehat{\beta}$ can be derived using the delta method and Theorem 2:

$$n^{1/4}(\mathrm{vec}(\widehat{\beta}) - \mathrm{vec}(\beta)) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \tilde{\Pi}^{\beta}\right), \quad \text{where}$$

$$\begin{aligned}
\tilde{\Pi}^{\beta} &= \left\{(\Sigma^{\star(2,2)})^{-1}[0_{f\times d} : \mathbb{I}_f]\right\} \\
&\quad \otimes \left\{[\mathbb{I}_d : 0_{d\times f}] - \Sigma^{\star(1,2)}\left(\Sigma^{\star(2,2)}\right)^{-1}[0_{f\times d} : \mathbb{I}_f]\right\} \times \mathbb{D}_d \Pi^{Q} \mathbb{D}_d^{\intercal} \\
&\quad \times \left\{(\Sigma^{\star(2,2)})^{-1}[0_{f\times d} : \mathbb{I}_f]\right\}^{\intercal} \\
&\quad \otimes \left\{[\mathbb{I}_d : 0_{d\times f}] - \Sigma^{\star(1,2)}\left(\Sigma^{\star(2,2)}\right)^{-1}[0_{f\times d} : \mathbb{I}_f]\right\}^{\intercal}, \quad \text{and} \\
\Sigma^{\star} &:= \begin{pmatrix} \Sigma^{\star(1,1)} & \Sigma^{\star(1,2)} \\ \Sigma^{\star(2,1)} & \Sigma^{\star(2,2)} \end{pmatrix} \\
&= \frac{1}{T}\int_0^T \begin{pmatrix} \beta \Sigma_u^f \beta^{\intercal} + \Sigma_u^e & \beta \Sigma_u^f \\ \Sigma_u^f \beta & \Sigma_u^f \end{pmatrix} du.
\end{aligned}$$

5.2. Multistep realised QML estimator

There are robustness advantages in estimating the integrated variances using univariate QML methods $\widehat{\Sigma}_{ii}$, $1 \le i \le d$. These estimates can then be combined with the QML correlation estimator $\widehat{\rho}_{ij}$, obtained by maximising the likelihood with respect to ρ_{ij} keeping Σ_{ii}, Σ_{jj} fixed at $\widehat{\Sigma}_{ii}$, $\widehat{\Sigma}_{jj}$. 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.

5.3. Sparse and subsampled realised QML estimators

A virtue of the realised QML is that it is applied to all of the high frequency data. That said, this estimator may have challenges if the noise associated with prices sampled at the highest frequency 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 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 with sparsity k, 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 kth 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.

In practice it makes sense to amend this approach so that for each i, $n_i^* \ge n_{\min}$ 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.

6. Monte Carlo experiments

6.1. Monte Carlo design I: a bivariate nonparametric model

We first compare our nonparametric estimator in a bivariate model with alternative estimators. This setting is practical for beta estimation. 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 adopt the log-volatility model in Li and Xiu (2016) for each asset $i = 1, 2$:

$$\begin{aligned}
dy_{it} &= \gamma_i dt + \sigma_{it} dW_{it}, \quad \sigma_{it}^2 = \exp(\alpha_i + \beta_i F_t), \\
dF_{it} &= \kappa_i F_{it} dt + s_i dB_{it} + J_{it} dN_{it} - \lambda_i \mu_i dt,
\end{aligned}$$

where $\mathrm{E}(dW_{it} dB_{jt}) = \delta_{ij}\rho_i dt$ and $\mathrm{E}(dW_{1t} dW_{2t}|\rho^*) = \rho^* dt$. Here $\kappa_i < 0$, so that volatility is mean-reverting. Throughout when jumps happen, the jump size follows $J_{it} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$, while N_{it} is a Poisson process with intensity λ_i. This setup allows more persistent volatility paths and downward volatility jumps. The microstructure noise follows $\varepsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, a_i^2)$.

For each 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 $\mathrm{sd}(\rho^*) = \rho_0\sqrt{\rho_1^*\rho_2^*}/\left\{\left(\rho_1^* + \rho_2^*\right)\sqrt{\rho_1^* + \rho_2^* + 1}\right\}$. The values of μ_i , α_i, ρ_i, κ_i, β_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, Fig. 1 plots the histograms of the standardised pivotal statistics with 1000 Monte Carlo repetitions sampled regularly in time at frequency of every 10 s, that is $n = 2340$. This corresponds to a 6.5 h trading day, which is the case for the NYSE and NASDAQ. The histograms show the limiting result in Theorem 2 provides a reasonable guide to the finite sample behaviour in these cases.

In our main Monte Carlo we take $n \in \{117, 1170, 11700\}$ and all results are based on 1000 stochastically independent replications. Having fixed the overall sample size n we randomly and uniformly scatter these points over the time interval $t \in [0, T]$. 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 $\Sigma_{11}^{\star} = \frac{1}{T}\int_0^T \Sigma_{11,u} du$, $\Sigma_{22}^{\star} = \frac{1}{T}\int_0^T \Sigma_{22,u} du$, $\Sigma_{12}^{\star} = \frac{1}{T}\int_0^T \Sigma_{12,u} du$, $\rho_{1,2}^{\star} = \Sigma_{12}^{\star}/\sqrt{\Sigma_{11}^{\star}\Sigma_{22}^{\star}}$, $\beta_{1|2}^{\star} = \Sigma_{12}^{\star}/\Sigma_{22}^{\star}$, $\beta_{2|1}^{\star} = \Sigma_{12}^{\star}/\Sigma_{11}^{\star}$.

We will compute six estimators of $\Sigma_{11}^{\star}$, $\Sigma_{22}^{\star}$, $\Sigma_{12}^{\star}$, $\beta_{1|2}^{\star}$, $\beta_{2|1}^{\star}$ and $\rho_{1,2}^{\star}$. The six are: (i) realised QML, (ii) multistep realised QML estimator, (iii) realised QML but using the reduced data synchronised by refresh time, (iv) Aït-Sahalia et al. (2010) which uses polarisation and refresh time, (v) the realised kernel of Barndorff-Nielsen et al. (2011) using refresh time, and (vi) the modulated

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

	a_i	γ_i	α_i	β_i	κ_i	s_i	λ_i	μ_i	σ_i^2	ρ_i	ρ_i^*	
$i = 1$	0.002	0.05	−2.8	6	−4	0.8	25	0.02	0.02	−0.8	2	$\rho_0 = 0.529$
$i = 2$	0.001	0.04	−2.8	4	−2	0.6	30	0.01	0.02	−0.6	1	$\mathrm{E}(\rho^*) = 0.176$, $sd(\rho^*) = 0.125$

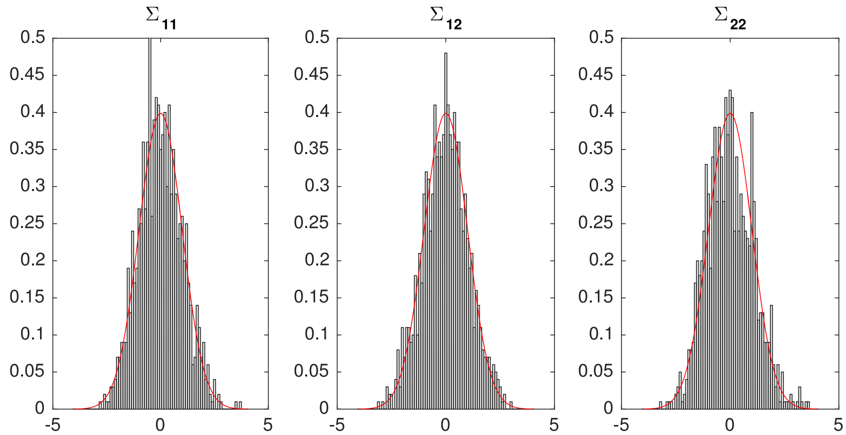

Fig. 1. The figure plots the histograms of the standardised pivotal statistics to verify the asymptotic theory developed in Theorem 2. The red solid line is the density of the standard normal. The standardisation is carried out using the infeasible true random asymptotic variance for each replication.

realised covariance of Christensen et al. (2010) using refresh time. We write this generically as $\widehat{\theta}_L$, with $L \in$ {QML, Step, RT, Pol, Kern, MRC}.[3]

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

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, θ_L and $\digamma$. The results indicate that all six estimators for $\Sigma_{11}^{\star}$, $\Sigma_{12}^{\star}$, and $\Sigma_{22}^{\star}$ perform roughly similarly when $\digamma = 0.5$, with a small degree of underperformance for Kern and MRC. When the data was more unbalanced, with $\digamma = 0.1$ or 0.9, then RT, Pol, Kern and MRC were considerably worse while realised QML being the best by a small margin. QML slightly outperformed Step, yet Step is slightly better in $\Sigma_{12}^{\star}$, which regards correlations.

When we move onto $\rho_{1,2}^{\star}$, the differences become more significant, although recalling that realised QML and Step are identical in this case. When $\digamma = 0.9$ then the Pol estimator struggles with the quantiles being considerably worse than 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 RT, Kern and MRC sit between Pol and QML, with MRC being a bit better. This 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 $\digamma = 0.1$. Overall the new methods seem to deliver an order of magnitude improvement in the accuracy of the estimator. In the balanced sampling case of $\digamma = 0.5$ the differences are more moderate but similar.

Before we progress to the regression case it is helpful to calibrate how accurately we have estimated $\rho_{1,2}^{\star}$ in the realised QML case. When $\digamma = 0.1$ the quantile is 2.15 with $n = 117$, so the corresponding quantile for $\left|\widehat{\rho}_{1,2} - \rho_{1,2}^{\star}\right|$ is 0.654. When $n = 1170$ it is 0.280. When $n = 11700$ it is 0.137. In the balanced case $\digamma = 0.5$ the corresponding results are 0.401, 0.197 and 0.105. 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. Even though $n = 11700$ the quantiles in the balanced case are not close to zero. Hence although we can nonparametrically 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 cases with $\beta_{1|2}^{\star}$ and $\beta_{2|1}^{\star}$ deliver the same type of results to the correlation, with again the QML and Step performing around an order of magnitude better than RT, Pol, and Kern, and with MRC sitting in between.

Finally, in Table 3 we investigate the impact of the dependent noise on the estimates. We keep the same parameters in the model except that the noises are simulated from an AR(1) model and a MA(1) model for each asset, respectively. The autoregressive and moving-average coefficients are fixed at 0.25 for both models. The standard deviations of the innovations are chosen as 0.002 and 0.001. We fix $\digamma = 0.5$. We consider Spar = 0 (the original estimator), 1, and 3, respectively. For Σ_{11}, Σ_{12}, and Σ_{22}, Kern and MRC perform better than QML when Spar = 0. This is not surprising because these two estimators are more robust to dependent noises at the expense of a lower rate. RT is also competitive, because of the use of refresh time, which effectively sparsifies the data. When Spar = 1 or 3, QML improves substantially, compared to the case with Spar = 0. In contrast, Kern and MRC perform similarly over different choices of Spar. For estimation of $\widehat{\rho}_{1,2}$, $\widehat{\beta}_{1|2}$, and $\widehat{\beta}_{2|1}$, however, the gain in efficiency dominates the bias due to dependent noise, so that QML performs better even when Spar = 0.

6.2. Monte Carlo design II: a semiparametric factor model

In this section, we focus on the performance of the factor-based realised QML estimator (FQML), and compare it with that of the

[3] To implement Kern, we use Parzen kernel with bandwidth $H = \max([0.5\bar{n}^{0.61}], 2)$, where $\bar{n}$ is the sample size after synchronisation. We have also used the average of the optimal bandwidth H_i for each asset i given by Barndorff-Nielsen et al. (2011). The results are fairly similar. To implement MRC, we choose the non-bias-corrected modulate realised covariance estimator that guarantees the positive definiteness, so we fix $k_n = \max([0.5\bar{n}^{0.61}], 2)$ and $g(x) = \min(x, 1 - x)$. We omit the comparison with its bias-corrected version of MRC, and the Hayashi–Yoshida version introduced in Christensen et al. (2013), because when $\digamma = 0.1$ or 0.9, these estimates tend to violate the positive definite constraint, leading to ill-defined correlations. The Hayashi–Yoshida based pre-averaging estimator even gives negative volatilities.

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 1000 independent replications. $\mathcal{F}$ denotes the % of the data corresponding to trades in asset 1. "QML" is our multivariate QML. "Step" is our multistep QML. "RT" is our multivariate QML using the Refresh Time. "Pol" is the existing polarisation and Refresh Time estimator. "Kern" is the existing multivariate realised kernel. "MRC" is the existing modulated realised covariance (pre-averaging) estimator. The numbers in bold indicate the minimum of quantiles in comparison.

		$\mathcal{F}=0.9$						$\mathcal{F}=0.5$						$\mathcal{F}=0.1$					
	n	QML	Step	RT	Pol	Kern	MRC	QML	Step	RT	Pol	Kern	MRC	QML	Step	RT	Pol	Kern	MRC
$\widehat{\Sigma}_{11}$	117	**0.13**	**0.13**	0.26	0.21	0.23	0.40	**0.15**	0.16	**0.15**	0.17	0.16	0.17	0.30	**0.22**	0.28	**0.22**	0.25	0.42
	1,170	**0.11**	**0.11**	0.20	0.23	0.23	0.27	**0.12**	0.13	0.13	0.14	0.20	0.20	**0.20**	0.22	**0.20**	0.22	0.24	0.27
	11,700	**0.11**	**0.11**	0.19	0.20	0.29	0.29	**0.13**	**0.13**	**0.13**	**0.13**	0.22	0.21	**0.18**	0.19	**0.18**	0.19	0.29	0.29
$\widehat{\Sigma}_{12}$	117	0.14	**0.11**	0.15	0.15	0.20	0.19	**0.09**	**0.09**	**0.09**	0.11	0.12	**0.09**	0.15	**0.13**	0.15	0.16	0.20	0.19
	1,170	0.10	**0.09**	0.11	0.15	0.15	0.14	**0.08**	**0.08**	**0.08**	0.09	0.10	0.10	**0.11**	**0.11**	**0.11**	0.14	0.15	0.14
	11,700	**0.09**	**0.09**	0.12	0.13	0.16	0.16	**0.08**	**0.08**	**0.08**	**0.08**	0.11	0.12	**0.10**	**0.10**	0.11	0.13	0.15	0.16
$\widehat{\Sigma}_{22}$	117	**0.19**	**0.19**	**0.19**	**0.19**	0.21	0.28	**0.12**	0.13	0.13	0.13	0.15	**0.12**	**0.09**	0.10	0.19	0.19	0.20	0.28
	1,170	**0.15**	0.17	**0.15**	0.17	0.19	0.17	**0.09**	0.10	**0.09**	0.10	0.13	0.13	**0.08**	**0.08**	0.15	0.17	0.19	0.18
	11,700	**0.13**	0.14	**0.13**	0.14	0.21	0.21	**0.09**	**0.09**	**0.09**	**0.09**	0.15	0.16	**0.07**	**0.07**	0.14	0.14	0.21	0.21
$\widehat{\rho}_{1,2}$	117	**1.88**	**1.88**	2.28	6.64	3.27	1.90	1.32	1.32	1.42	1.71	1.71	**1.29**	2.15	2.15	2.31	9.72	3.32	**2.08**
	1,170	**1.36**	**1.36**	1.72	2.21	1.93	1.91	1.15	1.15	**1.14**	1.26	1.38	1.43	**1.64**	**1.64**	1.74	2.17	1.96	1.87
	11,700	**1.30**	**1.30**	1.75	1.86	2.09	2.11	**1.09**	**1.09**	1.12	1.14	1.55	1.54	**1.42**	**1.42**	1.62	1.83	2.07	2.05
$\widehat{\beta}_{1\vert 2}$	117	**2.08**	4.24	2.76	5.48	4.93	2.46	**1.48**	1.54	1.58	1.97	2.12	1.50	2.58	**2.14**	3.17	6.20	5.46	2.71
	1,170	**1.53**	1.58	1.82	2.53	2.31	2.13	**1.21**	1.23	**1.21**	1.38	1.47	1.51	**1.69**	1.71	1.88	2.37	2.21	2.04
	11,700	**1.43**	1.44	1.88	1.97	2.23	2.26	**1.14**	1.15	1.17	1.24	1.60	1.61	**1.52**	1.55	1.73	1.92	2.22	2.26
$\widehat{\beta}_{2\vert 1}$	117	2.27	**1.87**	3.51	9.25	4.53	1.98	1.61	1.61	1.75	2.15	1.95	**1.34**	2.82	6.11	3.41	9.19	5.34	**2.11**
	1,170	1.48	**1.46**	2.06	2.46	2.06	1.95	1.28	**1.27**	1.28	1.38	1.50	1.51	**1.96**	2.07	2.02	2.59	2.10	**1.96**
	11,700	**1.36**	1.38	1.92	1.98	2.27	2.27	**1.19**	1.20	1.25	1.29	1.67	1.68	**1.61**	**1.61**	1.81	1.99	2.22	2.22

Table 3
Monte Carlo results for the volatility, covariance, correlation and beta estimation in the presence of dependent noise. We report the 0.9 quantiles of $\left|n^{1/4}\left(\widehat{\theta}_L - \theta_L\right)\right|$ over 1000 independent replications. $\mathcal{F}$ is fixed at 0.5. Spar denotes the sparsity level we select, which is introduced in Section 5.3. "QML" is our multivariate QML. "Step" is our multistep QML. "RT" is our multivariate QML using the Refresh Time. "Pol" is the existing polarisation and Refresh Time estimator. "Kern" is the existing multivariate realised kernel. "MRC" is the existing modulated realised covariance (pre-averaging) estimator. The numbers in bold indicate the minimum of quantiles in comparison.

		Spar = 0						Spar = 1						Spar = 3					
	n	QML	Step	RT	Pol	Kern	MRC	QML	Step	RT	Pol	Kern	MRC	QML	Step	RT	Pol	Kern	MRC
$\widehat{\Sigma}_{11}$	117	0.20	0.19	0.18	0.18	**0.17**	0.19	0.23	0.18	0.27	0.18	0.17	**0.16**	0.19	0.18	0.19	0.17	**0.16**	0.23
	1,170	0.25	0.25	**0.19**	0.22	0.21	0.21	**0.17**	0.18	**0.17**	0.18	0.21	0.23	0.21	0.20	**0.17**	0.18	0.20	0.22
	11,700	0.45	0.46	0.33	0.34	0.27	**0.23**	**0.17**	0.18	**0.17**	0.18	0.27	0.25	0.33	0.32	0.25	0.26	0.27	**0.24**
$\widehat{\Sigma}_{12}$	117	**0.09**	**0.09**	**0.09**	0.11	0.10	**0.09**	0.12	**0.10**	0.12	0.12	0.13	**0.10**	**0.10**	**0.10**	**0.10**	0.11	0.11	**0.10**
	1,170	0.10	0.10	**0.09**	0.11	0.10	0.10	**0.10**	**0.10**	**0.10**	0.12	0.13	0.13	0.11	0.10	**0.09**	0.11	0.12	0.11
	11,700	0.15	0.15	0.12	0.13	**0.11**	0.12	**0.09**	**0.09**	0.10	0.10	0.13	0.13	0.12	0.12	**0.10**	0.12	0.12	0.12
$\widehat{\Sigma}_{22}$	117	0.14	0.13	0.13	0.13	0.14	**0.12**	0.15	0.15	0.18	0.15	0.16	**0.13**	0.15	**0.14**	0.16	**0.14**	0.15	0.15
	1,170	0.15	0.15	**0.12**	0.14	0.15	0.15	**0.12**	**0.12**	**0.12**	0.13	0.17	0.15	0.14	0.13	**0.11**	0.13	0.14	0.14
	11,700	0.26	0.25	0.19	0.19	**0.15**	**0.15**	**0.11**	**0.11**	**0.11**	0.12	0.19	0.18	0.21	0.19	**0.16**	0.17	0.17	0.17
$\widehat{\rho}_{1,2}$	117	1.28	1.28	1.38	1.70	1.40	**1.27**	1.78	1.78	2.03	2.68	1.97	**1.50**	1.54	1.54	1.64	1.94	1.67	**1.40**
	1,170	**1.20**	**1.20**	1.21	1.34	1.32	1.27	**1.41**	**1.41**	1.45	1.63	1.81	1.66	**1.28**	**1.28**	1.35	1.50	1.58	1.51
	11,700	**1.28**	**1.28**	1.31	1.35	1.59	1.58	**1.38**	**1.38**	1.43	1.49	1.72	1.73	1.30	1.30	**1.29**	1.39	1.57	1.61
$\widehat{\beta}_{1\vert 2}$	117	**1.44**	1.52	1.53	1.91	1.77	1.51	2.02	2.26	2.31	2.85	2.42	**1.83**	1.73	1.81	1.83	2.22	1.96	**1.55**
	1,170	**1.33**	1.40	1.34	1.52	1.52	1.50	**1.56**	1.57	1.64	1.86	2.02	1.85	**1.45**	1.46	1.48	1.67	1.71	1.71
	11,700	**1.35**	1.36	1.38	1.50	1.58	1.68	**1.52**	1.55	1.57	1.63	1.86	1.83	1.47	1.46	**1.43**	1.57	1.71	1.72
$\widehat{\beta}_{2\vert 1}$	117	1.48	1.53	1.56	2.01	1.49	**1.30**	2.46	2.47	2.78	3.38	2.32	**1.49**	1.80	1.79	2.05	2.30	1.76	**1.38**
	1,170	1.28	**1.27**	1.33	1.40	1.37	1.35	**1.58**	1.59	1.63	1.79	1.84	1.69	**1.39**	**1.39**	1.45	1.65	1.59	1.50
	11,700	**1.41**	1.42	1.44	1.44	1.79	1.76	1.47	**1.46**	1.55	1.57	1.86	1.79	**1.38**	1.40	1.42	1.46	1.83	1.75

previous realised QML estimator (QML). We adopt the same high-frequency factor model from Fan et al. (2016).

We sample 1000 paths from a continuous-time f-factor model of d assets specified as:

$$dy_{it} = \sum_{j=1}^{f} \beta_{ij} dF_{jt} + dZ_{it}, \quad dF_{jt} = b_j dt + \sigma_{jt} dW_{jt}, \quad dZ_{it} = \gamma_{it}^{\intercal} dB_{it},$$

where W_j is a standard Brownian motion and B_i is a d-dimensional Brownian motion, for $i = 1, 2, \ldots, d$, and $j = 1, 2, \ldots, f$. They are mutually independent. F_j is the jth observable factor. One of the Fs is deemed the market factor, so that its associated βs are positive. The spot covariance matrix of Z is a block-diagonal matrix, denoted by Σ_t^e. Within each block, the idiosyncratic correlation is fixed to be ρ. On the diagonal, $\Sigma_{ii,t}^e = \sigma_{0t}^2$ for each i. We allow for time-varying σ_{jt}, which evolves according to the following system of equations:

$$d\sigma_{jt}^2 = \kappa_j(\theta_j - \sigma_{jt}^2)dt + \eta_j \sigma_{jt} d\widetilde{W}_{jt}, \quad j = 0, 1, 2, \ldots, f,$$

where $\widetilde{W}_j$ is a standard Brownian motion with $\mathbb{E}[dW_{jt} d\widetilde{W}_{jt}] = \rho_j dt$.

We first simulate the $d = 4$ and $f = 1$ case to verify the central limit theorem in Theorem 3, where data are sampled synchronously every 10 s. In addition, $\kappa_0 = 4$, $\kappa_1 = 3$, $\theta_0 = 0.06$, $\theta_1 = 0.09$, $\eta_0 = \eta_1 = 0.3$, $\rho_1 = -0.6$, and $b_1 = 0.05$. $\beta_1 \sim \mathcal{U}[0.25, 1.75]$. Σ^e is a diagonal matrix, i.e., $\rho = 0$. The variances of microstructure noises in all processes are fixed as 0.002. Fig. 2 plots the histograms of the standardised pivotal statistics with 1000 Monte Carlo repetitions sampled regularly in time at frequency of every 10 s, for all parameters of interest. Our asymptotic prediction matches the finite sample distribution rather well.

To mimic the empirical study and illustrate the advantage of a factor model, we simulate a larger factor model with $d = 12$ and $f = 1$. We first simulate all stocks at the 20-second frequency, then contaminate and censor the data to simulate the effects of microstructure noise and asynchronous trading. We add Gaussian noise to the simulated log prices before censoring. To censor the

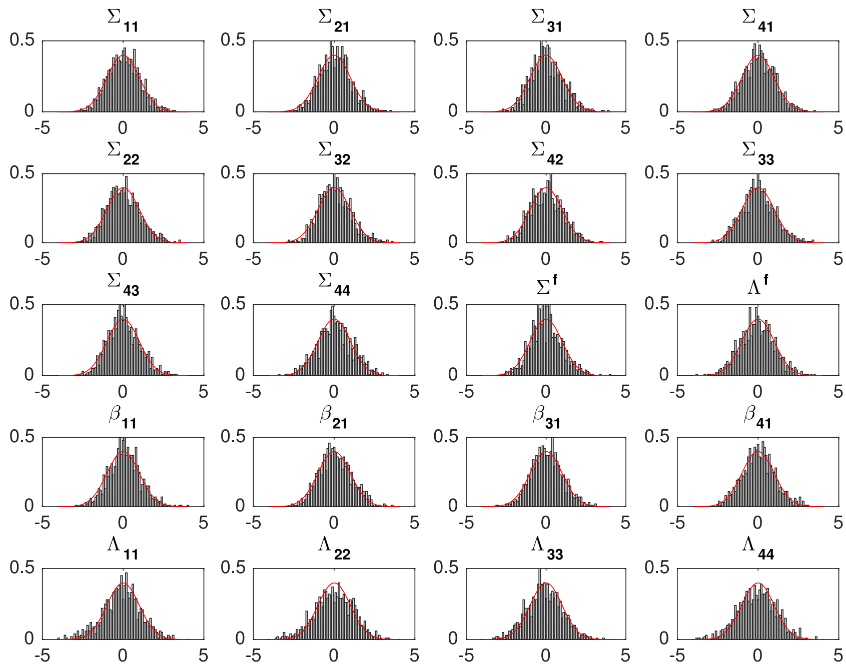

Fig. 2. The figure plots the histograms of the standardised pivotal statistics to verify the asymptotic theory developed in Theorem 3. The red solid line is the density of the standard normal. The standardisation is carried out using the infeasible true random asymptotic variance for each replication.

data, we generate the number of observed prices $n\digamma_i$ for each of the 12 assets, where $\digamma_i \sim \mathcal{U}[0.1, 0.9]$, and then distribute the observations uniformly within the day. We will report on the accuracy of the daily estimation of $\frac{1}{T}\int_0^T \Sigma_u \mathrm{d}u$, $\frac{1}{T}\int_0^T \Sigma_u^f \mathrm{d}u$, $\frac{1}{T}\int_0^T \mathrm{diag}(\Sigma_u^e)\mathrm{d}u$, and β under a variety of loss functions. We first study the case of $\rho = 0$, so that FQML is correctly specified.

Fig. 3 compares the accuracy, in terms of the 0.9 percent quantiles of the estimation error, for $\frac{1}{T}\int_0^T \Sigma_u \mathrm{d}u$, $\frac{1}{T}\int_0^T \mathrm{diag}(\Sigma_u^e)\mathrm{d}u$, and β, respectively. To estimate β using QML, Kern, and MRC, we use the multivariate realised factor beta estimator in Section 5.1. There are 78 parameters in $\frac{1}{T}\int_0^T \Sigma_u \mathrm{d}u$, and 12 parameters for $\frac{1}{T}\int_0^T \mathrm{diag}(\Sigma_u^e)\mathrm{d}u$ and β. We find that both QML and FQML estimators dominate the MRC and Kern by a wide margin, because the latter two estimators use a refresh time device and hence remove a large number of observations. FQML further dominates QML, because it uses a correctly specified factor model.

We further report in Table 4 the estimates of $\frac{1}{T}\int_0^T \Sigma_u \mathrm{d}u$ and its inverse under a variety of loss functions, including the $\mathbb{L}_2$ norm $\|\cdot\|$ and the $\mathbb{L}_\infty$ norm of the matrix on the vector space $\|\cdot\|_{\mathrm{MAX}}$. We also examine the cases where $\rho = 0.25$, 0.50, and 0.75, so that FQML is inconsistent, because of the misspecification of Σ_u^e. In the case of $\rho = 0$, the results suggest that QML and FQML are more accurate than Kern and MRC for covariance matrix estimation. Nonetheless, QML is the worst among all estimators in terms of the estimation of the inverse of the covariance matrix. This is not surprising because both Kern and MRC are modified sample covariance matrix estimator, whereas QML involves a more complex optimisation step that makes it difficult to control the matrix-wise properties. However, imposing a factor model solves this issue. In fact, FQML dominates QML, Kern, and MRC across all loss functions even when the factor model FQML is based on is (slightly) misspecified, i.e., $\rho = 0.25$ and 0.50. The difference is quite substantial for the inverse, even with only 12 dimensions. This result agrees with the theoretical prediction based on the large deviation theory by Fan et al. (2016) of the impact of dimensionality on covariance matrix estimation – the gain on the convergence rate by imposing a factor structure is substantial for the inverse of the covariance matrix estimation, whereas there is no gain on the convergence rate for estimating the

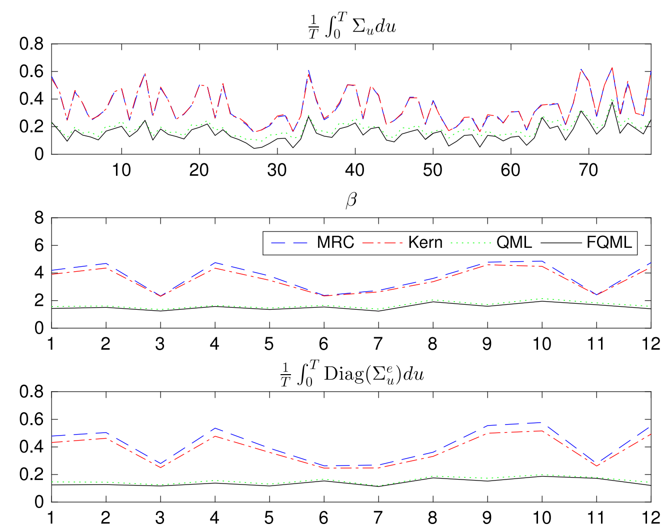

Fig. 3. The figure compares the 0.9 percent quantiles of $\left|n^{1/4}\left(\widehat{\theta}_L - \theta_L\right)\right|$, using $L =$ MRC (blue dashed), Kern (red dot-dashed), QML (green dotted), and FQML (black solid) estimators, respectively, where $\theta_L = \frac{1}{T}\int_0^T \Sigma_u du$ (upper panel), β (middle panel), and $\frac{1}{T}\int_0^T \mathrm{diag}(\Sigma_u^e)du$ (lower panel). We fix $\rho = 0$ so that FQML is correctly specified.

covariance matrix itself. The fact that the FQML is inconsistent does not jeopardise its relative performance simply because the model misspecification error is dominated by the estimation error. When $\rho = 0.75$, the factor model upon which FQML is based is strongly misspecified, and Kern and MRC dominate FQML in terms of the inverse covariance matrix estimation, and QML dominates FQML for covariance matrix estimation.

To illustrate the computational expense of both estimators, Fig. 4 provides results of a speed contest between QML and FQML for various choices of dimension (i.e., $d = 4, 8, 16, 32, 64$, and 128) and two sampling frequencies (i.e., $\Delta = 5$ and 20 s). For each choice of d and Δ, we report the average running time over 10 Monte Carlo trials. The data generating process is the same as

Table 4
This table reports the medians of the values of a variety of loss functions, based on estimates of $L =$ QML, FQML, Kern, and MRC, respectively. The idiosyncratic covariance matrix Σ_u^e is block-diagonal with 4 blocks, with the idiosyncratic correlations of each block ρ being equal to 0, 0.25, 0.5, and 0.75, respectively. For the case of $\rho = 0$, Σ_u^e matrix is diagonal, and FQML is correctly specified. The numbers in bold indicate the minimum of medians in comparison.

	QML	FQML	Kern	MRC	QML	FQML	Kern	MRC
Loss function	$\rho = 0$				$\rho = 0.25$			
$\\|\widehat{\Sigma}_L - \frac{1}{T}\int_0^T \Sigma_u du\\|$	0.165	**0.154**	0.326	0.319	0.139	**0.135**	0.289	0.289
$\\|\widehat{\Sigma}_L - \frac{1}{T}\int_0^T \Sigma_u du\\|_{\text{MAX}}$	0.054	**0.050**	0.097	0.096	0.048	**0.045**	0.085	0.090
$\left\\|(\widehat{\Sigma}_L)^{-1} - (\frac{1}{T}\int_0^T \Sigma_u du)^{-1}\right\\|$	145.32	**9.908**	56.188	46.443	157.284	**11.903**	56.748	43.127
$\left\\|(\widehat{\Sigma}_L)^{-1} - (\frac{1}{T}\int_0^T \Sigma_u du)^{-1}\right\\|_{\text{MAX}}$	55.986	**8.835**	22.026	17.603	59.096	**7.981**	21.855	16.140
Loss function	$\rho = 0.50$				$\rho = 0.75$			
$\\|\widehat{\Sigma}_L - \frac{1}{T}\int_0^T \Sigma_u du\\|$	**0.163**	**0.163**	0.335	0.329	**0.159**	0.174	0.337	0.337
$\\|\widehat{\Sigma}_L - \frac{1}{T}\int_0^T \Sigma_u du\\|_{\text{MAX}}$	**0.052**	0.056	0.094	0.094	**0.049**	0.065	0.093	0.093
$\left\\|(\widehat{\Sigma}_L)^{-1} - (\frac{1}{T}\int_0^T \Sigma_u du)^{-1}\right\\|$	1593	**19.836**	62.105	49.304	55539	52.51	59.56	**51.64**
$\left\\|(\widehat{\Sigma}_L)^{-1} - (\frac{1}{T}\int_0^T \Sigma_u du)^{-1}\right\\|_{\text{MAX}}$	524.2	**12.131**	22.658	17.474	17036	32.00	25.03	**24.91**

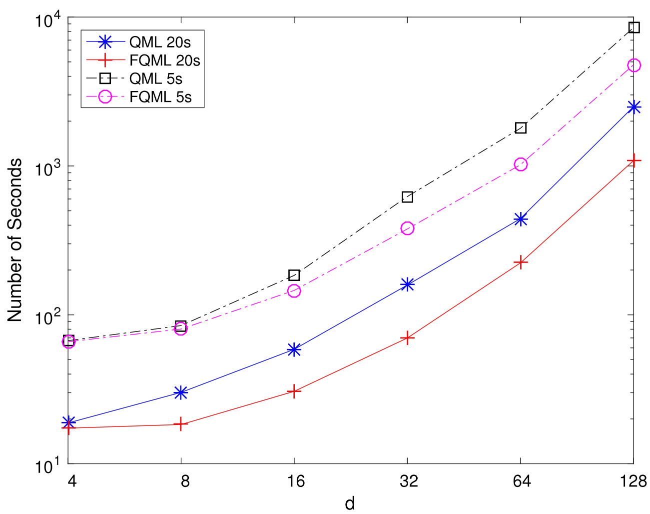

Fig. 4. The figure compares the running time between QML and FQML for various dimensions and sampling frequencies. We implement both algorithms in Matlab 2016a and run them on a desktop with a 4 GHz Core i7 processor and 32 GB memory. The running time is calculated based on the average time of 10 Monte Carlo samples for each choice of d and Δ.

above with diagonal residual covariance matrix and with f fixed at 1. Not surprisingly, we see a more significant speed-up for FQML as d increases, because the number of parameters is dramatically reduced with a factor structure imposed. Compared to alternative estimators, the computational cost for QML-type estimators could become a severe concern when the dimension is in the hundreds.

7. Empirical implementation

To demonstrate the empirical relevance of the proposed estimators, we design two empirical exercises. We first investigate the value of realised volatility and covariance estimates in a forecasting exercise based on GARCH-X type of models. The performance hinges on the point-wise accuracy of the covariance matrix estimation, so we only report the results based on QML and Kern estimates for ease of presentation. The results using FQML and MRC are similar to those of QML and Kern, respectively, as is confirmed from simulations. In the comparison, we also include the realised covariance estimator as a benchmark. To examine and compare the matrix-wise properties of the covariance estimates, we conduct a horse race based on minimum variance portfolios constructed using estimates of FQML, QML, Kern, and MRC.

7.1. Our database

We use data from the cleaned trade database of Lunde et al. (2012). It is taken from the TAQ database accessed through the Wharton Research Data Services (WRDS) system. 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 assets 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".

7.2. Summaries

Fig. 5 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.

The median intensity for the 13 assets is around 5000 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 1000 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 Kern or MRC based on refresh times.

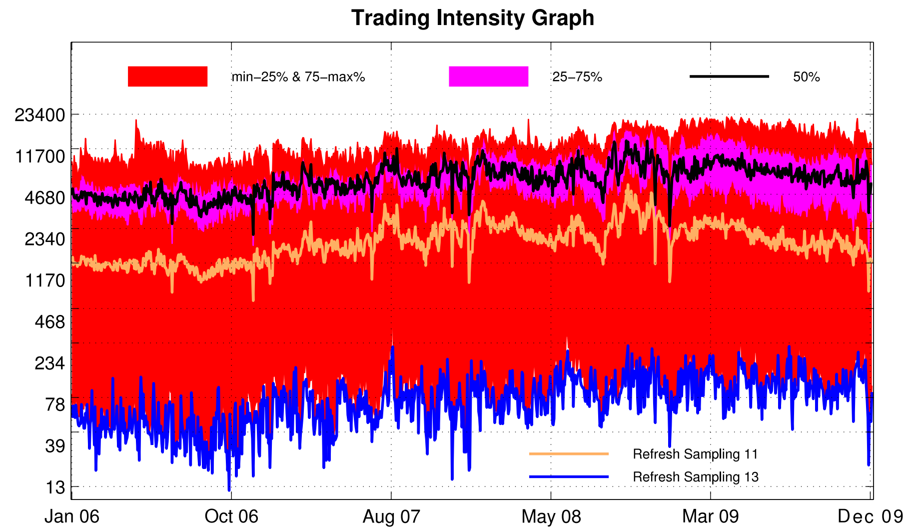

Fig. 5. The trading intensity graph for our 13 assets. The figure plots the min, max, 25%, 50%, and 75% quantiles of the number of observations for the cleaned dataset. The number of refresh sampling for the 11 assets and 13 assets databases is also plotted.

7.3. 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 behaviour of the volatility estimators, due to the use of refresh time.[4] 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.[5]

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 tth 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 et al. (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 of 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 evident for its statistical usefulness.

First focus on the XOM case. Table 5 shows the results in the univariate and 13 dimensional 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 squared return (Shephard and Sheppard, 2010). Further β falls dramatically and meaningfully, compared to the case with $\gamma = 0$. 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 Fig. 6. This reports the likelihood for QML minus the likelihood for Kern. On the left hand side we deal with the univariate case. On the right the 13 dimensional case is the focus. So negative numbers prefer Kern. This shows in the univariate case at sparsity of 1 Kern 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.

7.4. Dependence

We now turn to looking at covariation amongst the assets. We will focus on QML, realised kernel and realised covariance estimators, which are computed each day.[6] In all cases they will be based on the 13 unbalanced database, which means we compute a 13×13 positive definite 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. Again we will only flag up issues which hold up in the cross section.

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

[5] We had a web appendix to the previous draft, which details various summary statistics for the volatility statistics, including volatility signature plots and correlograms. It is available upon request.

[6] 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.

Table 5
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. Spar indicates the sparsity associated with the sparse realised QML defined in Section 5.3.

			GARCHX											
			QML				Kern				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

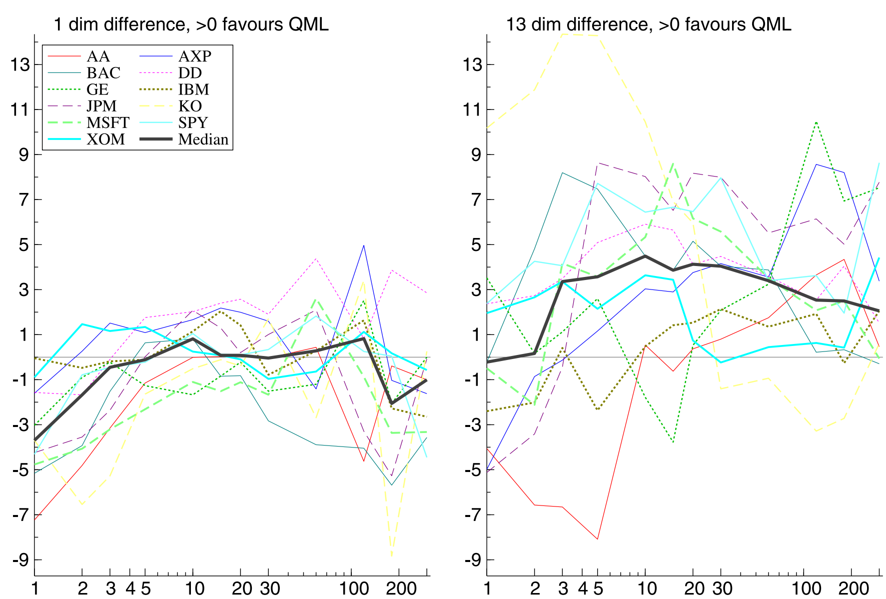

Fig. 6. logL improvement of realised QML minus logL of Kern, with numbers greater than 0 being supportive of QML. X-axis is the level of sparsity defined in Section 5.3. 1 and 13 dimensional cases are on the left and right hand sides respectively.

The focus here will be on the forecasting of 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 ith devolatilised series

$$e_{i,t} = \frac{y_{i,t}}{\sigma_{i,t}}. \tag{14}$$

Here the ijth 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 as follows, with $\omega + \alpha + \beta + \gamma = 1$,

$$R_t = \omega \Pi + \alpha C_{t-1} + \beta R_{t-1} + \gamma X_{t-1}, \quad \omega, \alpha, \beta, \gamma \geq 0. \tag{15}$$

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

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

Table 6
Forecasting exercises for cross-sectional dependence between BAC (Bank of America) and SPY (S&P 500 exchange traded fund). The estimated model is (15). Here ρ is the non-unit element of Π. LogL is the improvement in the log likelihood compared to $\gamma = 0$. Spar indicates the sparsity associated with the sparse realised QML defined in Section 5.3.

	QML					Kern					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

In the case where $M = d$ and there are no X variables, then (15) is the Tse and Tsui (2002) model.[7] 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 $t \le M$.

Note (15) 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 (14). This happens in practice.

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

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

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

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

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

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 $\log L_C$.

The results from this forecasting exercise are given in Table 6. These are based upon the innovations from the univariate volatility models for BAC and SPY conditioning on lagged realised QML statistics. The results for the dependence model when we do not condition on any additional realised quantities, that is $\gamma = 0$, are given above the line in the table. As $\beta = 0$ it means $\omega = 1 - \alpha$ and so is roughly 0.41. Here ρ is the non-unit element of Π of (15). Hence the estimated model for the conditional correlation is $0.41 \times 0.69 + 0.59C_{1,2,t-1}$ 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, Kern 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 Kern, but the difference is not particularly large.

Now we just focus on the cross section involving the 12 SPY based pairs. The top left of Fig. 7 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. When sparsity level is high, the realised covariance estimator dominates QML.

There is a clear trade-off among the three estimators. Kern is robust to potentially dependent noise in the data, (see Section 6.1.2 of Barndorff-Nielsen et al., 2011), despite a lower convergence

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

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

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

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.

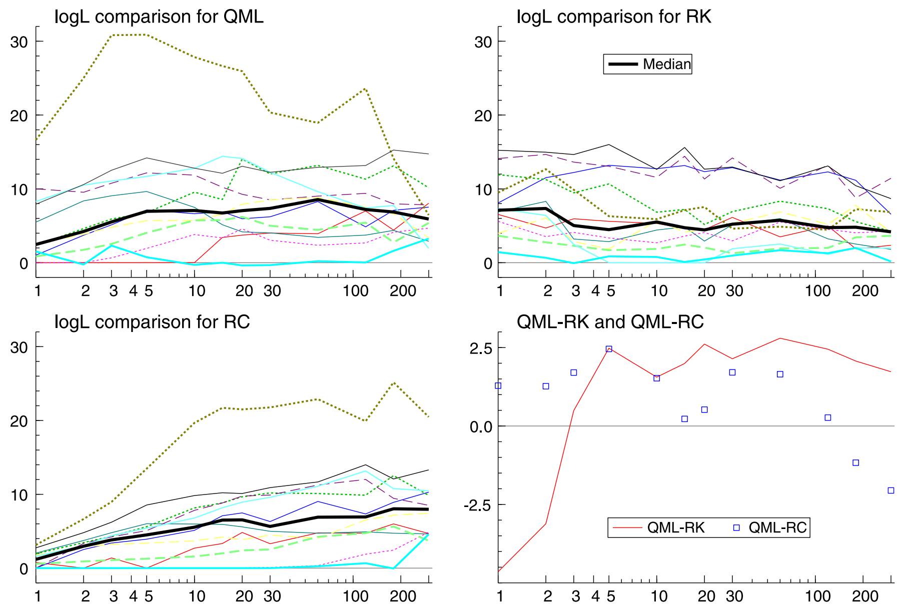

Fig. 7. Gains in logL for pairs involving SPY by including the realised quantities. On the x-axis is the level of sparsity, the heavy line denotes the median. All realised quantities are computed using 13 series. Bottom right is the logL for the model including realised QML minus the corresponding figure for realised kernel. Also drawn is the result for the realised QML against realised covariance.

rate. More sparsity leads to less data, hence it is detrimental to Kern. QML is more efficient but is less robust to the dependence of the noise, so it needs a moderate sparsity to be competitive as the dependence of noise disappears. Among all three estimators, realised covariance converges at the best rate, but is most vulnerable to the presence of noise.

Overall we can see that for the dependence modelling the inclusion of the realised information does add value, but the effects are not enormous. 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.

7.5. Minimum variance portfolio

We now investigate the performance of our estimators including QML, FQML, Kern, and MRC, in building minimum variance portfolios:

$$\min_{\omega} \omega' \widehat{\Sigma} \omega, \quad \text{subject to} \quad \mathbf{1}'\omega = 1.$$

We build these portfolios on a weekly basis. At the beginning of each week, we calculate different covariance matrix estimates with the previous week's data. This amounts to using the most recent covariance estimate as the predictor for the next period. We then hold our optimal portfolio weights for one week. To benchmark, we also implement the popular equal-weighted portfolio strategy, as advocated by DeMiguel et al. (2009).

We adopt the realised (ex-post) risk of an (ex-ante) minimum variance portfolio as an empirically relevant loss function, which sheds light on the matrix-wise properties of the covariance estimates. Fig. 8 compares the time series of the realised volatilities (in logarithm) of different portfolios for different values of sparsity. The realised volatilities are calculated using the sparse realised QML with portfolio returns subsampled every 10 ticks (Spar = 10). Since we estimate based on a whole week of intraday data, choosing a sparsity level of 10 helps reduce the sample size and accelerate the estimation. The comparison results remain the same if we use different frequencies or alternative estimators. In all scenarios, the FQML based optimal portfolio obtains the minimal out-of-sample volatility level by a large margin, among all minimum variance portfolios. Yet the risk of the FQML portfolio is on a par with that of the equal-weighted portfolio.

In addition, we report in Table 7 additional empirical measures of portfolio performance, including the Sharpe ratio, the average and the maximal short positions in percentage of wealth, and the peak-to-valley drawdown, all of which are widely used among practitioners.[8] We find that the FQML based strategy leads to the lowest realised volatilities, the modest short positions, and the highest Sharpe ratios for all choices of Spar over the sampling period, among all minimal variance portfolios. When sparsity is high, FQML also dominates in terms of the drawdown measure. Compared to the equal-weighted portfolio, FQML based strategy also has a larger Sharpe ratio and a smaller drawdown.

8. Conclusion

This paper systematically studies a new likelihood-based method for estimating the dependence amongst financial asset price processes. The realised QML and FQML estimators are robust to certain types of market microstructure noise and can deal with asynchronous time stamps. They are also guaranteed to be positive definite and converge at the optimal asymptotic rate. Neither requires tuning parameters, e.g., bandwidths or local window sizes. This combination of properties is unique in the literature, favourable among practitioners, and so it is worthwhile exploring these estimators in some detail.

In this paper we show how to numerically optimise the likelihood in a simple way even in large dimensions. We also develop some of the theory needed to understand the properties of these estimators and the corresponding results for betas and correlations. Our Monte Carlo experiments are extensive, comparing the estimator to various alternatives. The QML performs well in these comparisons for point-wise estimation accuracy, in particular in unbalanced cases, whereas the FQML dominates all estimators in terms of the matrix-wise loss functions.

Our initial empirical results are somewhat encouraging. The volatilities seem to be robust to the presence of slowly trading

[8] The drawdown is the peak-to-trough decline, measured as the (negative) percentage change between the peak and the trough. It is also a measure of risk.

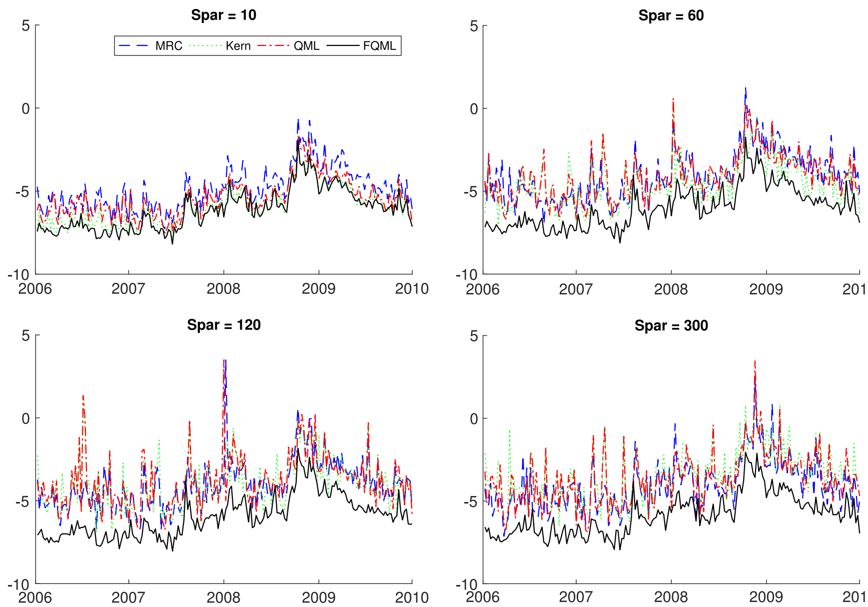

Fig. 8. Comparison of out-of-sample weekly annualised realised volatility (in logarithm) of the optimal portfolios using covariance matrices estimated by MRC (blue dashed), Kern (red dot-dashed), QML (green dotted), and FQML (black solid) estimators, respectively, along with that of the equal-weighted portfolio (yellow thicker dashed). The realised volatilities are calculated using the sparse realised QML with portfolio returns subsampled every 10 ticks.

Table 7
Comparison of out-of-sample portfolio allocation performance. We report the time series medians of realised volatilities, Sharpe ratios, average and maximal short positions in percentage of wealth, along with the (negative) peak-to-valley drawdown for all minimal variance portfolios with various values of Spar. We also report these statistics for the equal-weighted portfolio (EQW). The numbers in bold indicate the optimal results in comparison. The realised volatilities are calculated using the sparse realised QML with portfolio returns subsampled every 10 ticks.

	QML	FQML	Kern	MRC	QML	FQML	Kern	MRC	EQW
	Spar = 10				Spar = 60				
Realised volatility	0.048	**0.044**	0.063	0.084	0.094	**0.047**	0.114	0.117	0.044
Sharpe ratio	1.870	**2.692**	0.085	0.300	1.995	**3.075**	2.310	2.282	1.851
Average short position	0.040	**0.030**	0.078	0.111	0.138	**0.039**	0.188	0.186	0
Maximal short position	0.092	**0.056**	0.179	0.264	0.308	**0.073**	0.436	0.430	0
Drawdown	**0.523**	0.540	0.650	0.574	**0.444**	0.499	0.670	0.699	1.196
	Spar = 120				Spar = 300				
Realised volatility	0.130	**0.048**	0.139	0.127	0.176	**0.050**	0.152	0.119	
Sharpe ratio	−0.308	**3.297**	0.538	2.327	−2.178	**3.562**	−0.217	1.419	
Average short position	0.197	**0.038**	0.224	0.203	0.295	**0.044**	0.214	0.183	
Maximal short position	0.489	**0.078**	0.478	0.462	0.732	**0.089**	0.570	0.442	
Drawdown	0.572	**0.473**	1.905	3.623	1.274	**0.430**	2.816	1.433	

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 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. When it comes to a large dimensional data, FQML estimator is more favourable in that it implements faster, that it is more accurate when a factor model serves as a good approximation, and that it has desirable performance in portfolio allocation exercises.

Acknowledgements

We thank the co-editor Oliver Linton, the associate editor, and two anonymous referees for their useful comments. We are particularly grateful for help we have received from Francesco Audrino, Markus Bibinger, Fulvio Corsi, Siem Jan Koopman, Cheng Liu, Stefano Peluso, Mark Podolskij, 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, David Veredas, and Zhongjun Qu. This research was supported in part by the FMC and IBM Faculty Scholar Funds at the University of Chicago Booth School of Business.

Appendix A. Collection of proofs

A.1. Proof of Theorem 1

Proof. To clarify our notation in this proof, we use superscript $\star$ to mark quantities that are made of true values. Σ, Λ, etc.

without any special marks represent any parameter values within the parameter space, which is assumed to be a compact set. For matrix function $F(X)$, we use $\mathrm{D}_X(F)$ to denote the Jacobian matrix of F with respect to X, and use $\mathrm{H}_X(F)$ for the corresponding Hessian matrix.

Step (1) By Theorems 8.2 and 8.3, Equation (4) of Section 2.4, and Ex. 1 of Section 3.7 of Magnus and Neudecker (1999), we have

$$\begin{aligned}\mathrm{vec}(dL_n) =&\ \mathrm{vec}\left(-\frac{1}{2}tr(\Omega^{-1}d\Omega)+\frac{1}{2}tr(\Omega^{-1}rr^\intercal\Omega^{-1}(d\Omega))\right)\\ =& -\frac{1}{2}\mathrm{vec}(\Omega^{-1})^\intercal\mathrm{vec}(d\Omega)\\ &+\frac{1}{2}\mathrm{vec}(\Omega^{-1}rr^\intercal\Omega^{-1})^\intercal\mathrm{vec}(d\Omega),\\ \mathrm{vec}(d\Omega)=&\ \Delta\mathrm{vec}(d\Sigma\otimes\mathbb{I}_n)+\mathrm{vec}\,(d\Lambda\otimes\mathbb{J}_n)\\ =&\ \Delta(\mathbb{I}_d\otimes G)\mathrm{vec}(d\Sigma)+(\mathbb{I}_d\otimes G')\mathrm{vec}(d\Lambda),\end{aligned}$$

where $G=(\mathbb{K}_{n,d}\otimes\mathbb{I}_n)(\mathbb{I}_d\otimes\mathrm{vec}(\mathbb{I}_n))$, $G'=(\mathbb{K}_{n,d}\otimes\mathbb{I}_n)(\mathbb{I}_d\otimes\mathrm{vec}(\mathbb{J}_n))$, and $\mathbb{K}_{n,d}$ is the commutation matrix introduced in Section 3.7 of Magnus and Neudecker (1999). Suppose $\mathbb{D}_d$ is the duplication matrix such that

$$\mathbb{D}_d\mathrm{vech}(C)=\mathrm{vec}(C), \tag{A.1}$$

and $\mathbb{E}_d=(e_1e_1^\intercal : e_2e_2^\intercal : \ldots : e_de_d^\intercal)^\intercal$ is $d^2\times d$ matrix, then by the first identification theorem, see Theorem 5.11 of Magnus and Neudecker (1999), we have

$$\begin{aligned}\Psi_n &= -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{vech}(\Sigma)}(L_n)\\ &=\frac{\Delta}{2\sqrt{n}}\big(\mathrm{vec}(\Omega^{-1})-\mathrm{vec}(\Omega^{-1}rr^\intercal\Omega^{-1})\big)^\intercal(\mathbb{I}_d\otimes G)\mathbb{D}_d, \qquad (A.2)\\ \Phi_n &= -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{diag}(\Lambda)}(L_n)\\ &=\frac{1}{2\sqrt{n}}\big(\mathrm{vec}(\Omega^{-1})-\mathrm{vec}(\Omega^{-1}rr^\intercal\Omega^{-1})\big)^\intercal(\mathbb{I}_d\otimes G')\mathbb{E}_d. \qquad (A.3)\end{aligned}$$

To simplify the notation, we omit the dependence of Ψ_n and Φ_n on Σ or Λ whenever possible. Further, by Theorems 2.2 and 8.3 of Magnus and Neudecker (1999), we have

$$\begin{aligned}d\Psi_n =& -\frac{\Delta}{2\sqrt{n}}\Big(\mathrm{vec}(\Omega^{-1}d\Omega\Omega^{-1})-\mathrm{vec}(\Omega^{-1}d\Omega\Omega^{-1}rr^\intercal\Omega^{-1})\\ &-\mathrm{vec}(\Omega^{-1}rr^\intercal\Omega^{-1}d\Omega\Omega^{-1})\Big)^\intercal(\mathbb{I}_d\otimes G)\mathbb{D}_d\\ =& -\frac{\Delta}{2\sqrt{n}}\Big(\big(\Omega^{-1}\otimes\Omega^{-1}-\big(\Omega^{-1}rr^\intercal\Omega^{-1}\big)\otimes\Omega^{-1}-\Omega^{-1}\\ &\otimes\ \big(\Omega^{-1}rr^\intercal\Omega^{-1}\big)\big)\,\mathrm{vec}(d\Omega)\Big)^\intercal(\mathbb{I}_d\otimes G)\mathbb{D}_d.\end{aligned}$$

Therefore, by the first identification theorem again, we obtain

$$\begin{aligned}&\mathrm{D}_{\mathrm{vech}(\Sigma)}(\Psi_n^\intercal) \qquad (A.4)\\ =& -\frac{\Delta^2}{2\sqrt{n}}\mathbb{D}_d^\intercal(\mathbb{I}_d\otimes G^\intercal)\big(\Omega^{-1}\otimes\Omega^{-1}-(\Omega^{-1}rr^\intercal\Omega^{-1})\\ &\otimes\ \Omega^{-1}-\Omega^{-1}\otimes(\Omega^{-1}rr^\intercal\Omega^{-1})\big)(\mathbb{I}_d\otimes G)\mathbb{D}_d.\end{aligned}$$

As a result, by (A.4), (A.2) and (A.3), we have

$$\begin{aligned}\bar\Psi_n &:= \mathrm{E}(\Psi_n)=\frac{\Delta}{2\sqrt{n}}\big(\mathrm{vec}(\Omega^{-1})-\mathrm{vec}(\Omega^{-1}\Omega^\star\Omega^{-1})\big)^\intercal(\mathbb{I}_d\otimes G)\mathbb{D}_d,\\ \bar\Phi_n &:= \mathrm{E}(\Phi_n)=\frac{1}{2\sqrt{n}}\big(\mathrm{vec}(\Omega^{-1})-\mathrm{vec}(\Omega^{-1}\Omega^\star\Omega^{-1})\big)^\intercal(\mathbb{I}_d\otimes G')\mathbb{E}_d,\\ V_n &:= \mathrm{E}\left(\mathrm{H}_{\mathrm{vech}(\Sigma)}(L_n)\right)=\mathrm{E}\left(\mathrm{D}_{\mathrm{vech}(\Sigma)}(\mathrm{D}_{\mathrm{vech}(\Sigma)}(L_n)^\intercal)\right)\\ &= \frac{\Delta^2}{2\sqrt{n}}\mathbb{D}_d^\intercal\left(\mathbb{I}_d\otimes G^\intercal\right)\left(\Omega^{-1}\otimes\Omega^{-1}\right)\left(\mathbb{I}_d\otimes G\right)\mathbb{D}_d.\end{aligned}$$

Step (2) We now calculate $\lim_{n\to\infty}V_n^{-1}$. The calculation will be useful in other parts of the proof, and in the calculation of the asymptotic variance of $\mathrm{vech}(\hat\Sigma)$.

We need some preparations. Note that there exists an orthogonal matrix O such that $D=O\Lambda^{-1/2}\Sigma\Lambda^{-1/2}O^\intercal$ is diagonal. If we write

$$\begin{aligned}&U_{ij}=\sqrt{\frac{2}{n+1}}\sin\Big(\frac{i\cdot j}{n+1}\pi\Big),\\ &1\le i,j\le n,\quad W=\left(U\otimes\left(O\Lambda^{-1/2}\right)\right)\mathbb{K}_{n,d},\end{aligned}$$

then we have $UU^\intercal=U^\intercal U=\mathbb{I}_n$,

$$\begin{aligned}\tilde D &:= U\mathbb{J}_nU^\intercal = 2\cdot\mathrm{diag}\left(\left(1-\cos\frac{\pi}{n+1}\right),\right.\\ &\left.\left(1-\cos\frac{2\pi}{n+1}\right),\ldots,\left(1-\cos\frac{n\pi}{n+1}\right)\right),\end{aligned}$$

and

$$\Gamma := W\Omega W^\intercal=\Delta\mathbb{I}_n\otimes D+\tilde D\otimes\mathbb{I}_d$$

is another diagonal matrix.

By Equation (2.9) and Definition 3.1 of Magnus and Neudecker (1979), we have

$$\begin{aligned}\mathbb{I}_d\otimes G &= \mathbb{I}_d\otimes(\mathbb{K}_{n,d}\otimes\mathbb{I}_n)(\mathbb{I}_d\otimes\mathrm{vec}(\mathbb{I}_n))\\ &=\sum_{i=1}^n\sum_{j=1}^d\mathbb{I}_d\otimes(e_i\epsilon_j^\intercal)\otimes\epsilon_j\otimes e_i,\end{aligned}$$

then by tedious calculations repeatedly using Equation (4) of Section 2.2 and Theorem 3.9 of Magnus and Neudecker (1999), we have

$$\begin{aligned}V_n =& \frac{\Delta^2}{2\sqrt{n}}\sum_{i,r=1}^n\sum_{j,k=1}^d\left\{\left(\left(e_i^\intercal U^\intercal\right)\otimes\left(\epsilon_j^\intercal\Lambda^{-1/2}O^\intercal\right)\right)\right.\\ &\times\left.\Gamma^{-1}\left((Ue_r)\otimes\left(O\Lambda^{-1/2}\epsilon_k\right)\right)\right\}\\ &\times\mathbb{D}_d^\intercal\mathbb{K}_{d,d}\left\{\left(\left(\epsilon_je_i^\intercal U^\intercal\right)\otimes\left(\Lambda^{-1/2}O^\intercal\right)\right)\right.\\ &\times\left.\Gamma^{-1}\left(\left(Ue_r\epsilon_k^\intercal\right)\otimes\left(O\Lambda^{-1/2}\right)\right)\right\}\mathbb{K}_{d,d}\mathbb{D}_d\\ :=&\mathbb{D}_d^\intercal\tilde V_n\mathbb{D}_d,\end{aligned}$$

where the (u,v)-entry of the (j,k)-block of the $d^2\times d^2$ matrix $\tilde V_n$ is given by

$$\begin{aligned}\tilde V_{n,uv}^{(j,k)} =& \frac{\Delta^2}{2\sqrt{n}}\sum_{l,m=1}^d O_{mj}O_{mk}O_{lu}O_{lv}\Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}\\ &\times\sum_{i=1}^n\left(\Delta D_{mm}+\tilde D_{ii}\right)^{-1}\left(\Delta D_{ll}+\tilde D_{ii}\right)^{-1}.\end{aligned}$$

Also, note that

$$\begin{aligned}&n^{-1/2}\Delta^2\sum_{i=1}^n\left(\Delta D_{mm}+\tilde D_{ii}\right)^{-1}\left(\Delta D_{ll}+\tilde D_{ii}\right)^{-1}\\ =&\frac{T^2}{\sqrt{n}}\sum_{i=1}^n\frac{1}{(TD_{mm}+2n(1-\cos\frac{i\pi}{n+1}))}\frac{1}{(TD_{ll}+2n(1-\cos\frac{i\pi}{n+1}))}.\end{aligned}$$

To simplify it, we notice that the summand decreases as i increases. Moreover, for any $0<\delta<1/2$,

$$\begin{aligned}&\frac{T^2}{\sqrt{n}}\sum_{i=n^{1-\delta}+1}^n\frac{1}{(TD_{mm}+2n(1-\cos\frac{i\pi}{n+1}))}\frac{1}{(TD_{ll}+2n(1-\cos\frac{i\pi}{n+1}))}\\ \le&\frac{K}{\sqrt{n}}(n-n^{1-\delta})\left(TD_{mm}+4n\sin^2\frac{n^{1-\delta}\pi}{2(n+1)}\right)^{-1}\end{aligned}$$

$$\times\left(TD_{ll}+4n\sin^2\frac{n^{1-\delta}\pi}{2(n+1)}\right)^{-1}$$
$$\leq Kn^{4\delta-3/2},$$

where K is some generic constant that may change from line to line. We choose $\delta < 3/8$, so that this term diminishes, as $n \to \infty$.

$$\frac{T^2}{\sqrt{n}}\sum_{i=1}^{n^{1-\delta}}\frac{1}{\left(TD_{mm}+4n\sin^2\frac{i\pi}{2(n+1)}\right)}\frac{1}{\left(TD_{ll}+4n\sin^2\frac{i\pi}{2(n+1)}\right)}$$
$$=\frac{T^2}{\sqrt{n}}\sum_{i=1}^{n^{1-\delta}}\frac{1}{\left(TD_{mm}+n\frac{i^2\pi^2}{(n+1)^2}\right)}\frac{1}{\left(TD_{ll}+n\frac{i^2\pi^2}{(n+1)^2}\right)}+o(1)$$
$$=\int_0^\infty\frac{T^2dx}{(TD_{mm}+\pi^2x^2)(TD_{ll}+\pi^2x^2)}+o(1)$$
$$=\frac{T^{1/2}}{2\left(D_{mm}D_{ll}^{1/2}+D_{mm}^{1/2}D_{ll}\right)}+o(1),$$

where the second equality uses $\sin\frac{i\pi}{2(n+1)}\sim\frac{i\pi}{2(n+1)}$, for $i\leq n^{1-\delta}$, and the third equality follows from the definition of the Riemann integral, since $n^{1-\delta}/\sqrt{n}\to\infty$.

Therefore, using the above equality, the (p,q)-entry of the (g,h)-block of the matrix $(O\Lambda^{1/2}\otimes O\Lambda^{1/2})\tilde{V}_n(\Lambda^{1/2}O^\intercal\otimes\Lambda^{1/2}O^\intercal)$ is given by

$$\sum_{l,m,j,u,k,v=1}^{d}O_{pj}O_{gu}O_{qk}O_{hv}\Lambda_{jj}^{1/2}\Lambda_{uu}^{1/2}\Lambda_{kk}^{1/2}\Lambda_{vv}^{1/2}O_{mj}O_{mk}O_{lu}O_{lv}$$
$$\times\Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}$$
$$\times\left(\frac{T^{1/2}}{4\left(D_{mm}D_{ll}^{1/2}+D_{mm}^{1/2}D_{ll}\right)}+o(1)\right)$$
$$=\frac{T^{1/2}}{4\left(D_{pp}D_{gg}^{1/2}+D_{pp}^{1/2}D_{gg}\right)}\delta_{p,q}\delta_{g,h}+o(1),$$

where $\delta_{p,q}$ is equal to 1, if $p=q$, and 0 otherwise. This implies that $(O\Lambda^{1/2}\otimes O\Lambda^{1/2})\tilde{V}_n(\Lambda^{1/2}O^\intercal\otimes\Lambda^{1/2}O^\intercal)$ is asymptotically a $d^2\times d^2$ diagonal matrix, and that its inverse is given by

$$\left\{(O\Lambda^{1/2}\otimes O\Lambda^{1/2})\tilde{V}_n(\Lambda^{1/2}O^\intercal\otimes\Lambda^{1/2}O^\intercal)\right\}^{-1}$$
$$=4T^{-1/2}\left(D\otimes D^{1/2}+D^{1/2}\otimes D\right)+o(1).$$

Therefore, we have

$$\begin{aligned}\tilde{V}_n^{-1}&=4T^{-1/2}(\Lambda^{1/2}O^\intercal\otimes\Lambda^{1/2}O^\intercal)\left(D\otimes D^{1/2}+D^{1/2}\otimes D\right)\\&\quad\times(O\Lambda^{1/2}\otimes O\Lambda^{1/2})+o(1)\\&=4T^{-1/2}\left\{\Sigma\otimes\left(\Lambda^{1/2}\left(\Lambda^{-1/2}\Sigma\Lambda^{-1/2}\right)^{1/2}\Lambda^{1/2}\right)\right.\\&\quad\left.+\left(\Lambda^{1/2}\left(\Lambda^{-1/2}\Sigma\Lambda^{-1/2}\right)^{1/2}\Lambda^{1/2}\right)\otimes\Sigma\right\}+o(1)\\&:=H+o(1).\end{aligned}$$

By Theorems 3.9, 3.11, Equations (4), (6), and (7) of Section 3.8 in Magnus and Neudecker (1999), we have

$$\begin{aligned}\left(\mathbb{D}_d^+H\mathbb{D}_d^{+\intercal}\right)\left(\mathbb{D}_d^\intercal H^{-1}\mathbb{D}_d\right)&=\mathbb{D}_d^+H\left(\mathbb{D}_d\mathbb{D}_d^+\right)^\intercal H^{-1}\mathbb{D}_d\\&=\mathbb{D}_d^+H\mathbb{N}_dH^{-1}\mathbb{D}_d=\mathbb{D}_d^+\mathbb{N}_dHH^{-1}\mathbb{D}_d\\&=\mathbb{I}_{d(d+1)/2},\\\left(\mathbb{D}_d^\intercal H^{-1}\mathbb{D}_d\right)\left(\mathbb{D}_d^+H\mathbb{D}_d^{+\intercal}\right)&=\mathbb{D}_d^\intercal H^{-1}\left(\mathbb{D}_d\mathbb{D}_d^+\right)H\mathbb{D}_d^{+\intercal}\\&=\mathbb{D}_d^\intercal H^{-1}H\mathbb{N}_d\mathbb{D}_d^{+\intercal}=\mathbb{D}_d^\intercal\mathbb{D}_d^{+\intercal}=\mathbb{I}_{d(d+1)/2},\end{aligned}$$

which imply that

$$V_n^{-1}=\left(\mathbb{D}_d^\intercal H^{-1}\mathbb{D}_d\right)^{-1}+o(1)=\left(\mathbb{D}_d^+H\mathbb{D}_d^{+\intercal}\right)+o(1).\tag{A.5}$$

Step (3) We now prove the consistency and derive the asymptotic distribution of $\mathrm{vech}(\widehat{\Sigma})$. It is easy to see that $\Omega=\Omega^\star$ (i.e., $\Sigma=\Sigma^\star$, $\Lambda=\Lambda^\star$) is the unique solution to $\bar{\Psi}_n=0$ and $\bar{\Phi}_n=0$. To establish the consistency, it is sufficient to show the uniform convergence in probability:

$$\sup_{\Sigma,\Lambda}\|\Psi_n-\bar{\Psi}_n\|\xrightarrow{p}0,\quad\sup_{\Sigma,\Lambda}\|\Phi_n-\bar{\Phi}_n\|\xrightarrow{p}0.\tag{A.6}$$

In fact, using (A.2), Theorem 2.2 of Magnus and Neudecker (1999), and Definition 3.1 of Magnus and Neudecker (1979), we have

$$\begin{aligned}\Psi_n-\bar{\Psi}_n&=-\frac{\Delta}{2\sqrt{n}}\mathrm{vec}\left(\Omega^{-1}\left(rr^\intercal-\Omega^\star\right)\Omega^{-1}\right)^\intercal(\mathbb{I}_d\otimes G)\mathbb{D}_d\\&=-\frac{\Delta}{2\sqrt{n}}\mathrm{vec}\left(rr^\intercal-\Omega^\star\right)^\intercal\left(\Omega^{-1}\otimes\Omega^{-1}\right)(\mathbb{I}_d\otimes G)\mathbb{D}_d\\&=-\frac{\Delta}{2\sqrt{n}}\sum_{j=1}^{d}\sum_{i=1}^{n}\mathrm{vec}\left(\left(\Omega^{-1}\left(\epsilon_j\otimes e_i\right)\right)^\intercal\right.\\&\quad\times\left.\left(rr^\intercal-\Omega^\star\right)\Omega^{-1}\left(\mathbb{I}_d\otimes\left(e_i\epsilon_j^\intercal\right)\right)\right)^\intercal\mathbb{D}_d.\end{aligned}\tag{A.7}$$

Because that the $(p-1)d+q$th entry of

$$\mathrm{vec}\left(\left(\Omega^{-1}\left(\epsilon_j\otimes e_i\right)\right)^\intercal\left(rr^\intercal-\Omega^\star\right)\Omega^{-1}\left(\mathbb{I}_d\otimes\left(e_i\epsilon_j^\intercal\right)\right)\right)$$

is

$$\sum_{k,l=1}^{d}e_i^\intercal(\Omega^{-1})^{(j,k)}(rr^\intercal-\Omega^\star)^{(k,l)}(\Omega^{-1})^{(l,p)}e_i\epsilon_j^\intercal\epsilon_q$$
$$=\sum_{k,l=1}^{d}\sum_{s,t=1}^{n}(\Omega^{-1})^{(j,k)}_{i,s}(rr^\intercal-\Omega^\star)^{(k,l)}_{s,t}(\Omega^{-1})^{(l,p)}_{t,i}\delta_{j,q},$$

we have

$$n^{1/4}\left(\Psi_n-\bar{\Psi}_n\right)=-\left(M_n^1+M_n^2+M_n^3\right)\mathbb{D}_d,$$

where for any $1\leq p,q\leq d$,

$$\begin{aligned}M^1_{n,(p-1)d+q}&=\frac{\Delta}{2n^{1/4}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij}\\&\quad\times\left(\left(\Delta_i^ny_k\right)\left(\Delta_j^ny_l\right)-\delta_{i,j}\Sigma_{kl}\Delta\right),\\M^2_{n,(p-1)d+q}&=\frac{\Delta}{2n^{1/4}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij}\\&\quad\times\left(\Delta_i^ny_k\Delta_j^n\varepsilon_l+\Delta_j^ny_l\Delta_i^n\varepsilon_k\right),\\M^3_{n,(p-1)d+q}&=\frac{\Delta}{2n^{1/4}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij}\\&\quad\times\left(\left(\Delta_i^n\varepsilon_k\right)\left(\Delta_j^n\varepsilon_l\right)-\mathrm{E}\left(\Delta_i^n\varepsilon_k\Delta_j^n\varepsilon_l\right)\right).\end{aligned}$$

Moreover, note that

$$\Omega^{-1}=W^\intercal\Gamma^{-1}W=\tilde{W}^\intercal\tilde{\Gamma}^{-1}\tilde{W},\tag{A.8}$$

where $\tilde{W}=\left(O\Lambda^{-1/2}\right)\otimes U$ and $\tilde{\Gamma}=\Delta D\otimes\mathbb{I}_n+\mathbb{I}_d\otimes\tilde{D}$. Hence, for any fixed k,l,p,q, we have

$$\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij}$$
$$=\sum_{g,h=1}^{d}\left(\tilde{W}^{\intercal(k,g)}(\tilde{\Gamma}^{-1})^{(g,g)}\tilde{W}^{(g,q)}\tilde{W}^{\intercal(p,h)}(\tilde{\Gamma}^{-1})^{(h,h)}\tilde{W}^{(h,l)}\right)_{ij}$$

$$= \sum_{g,h=1}^{d} O_{gk}O_{gq}O_{hp}O_{hl}\Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2} \times \left(U^\intercal(\tilde{\Gamma}^{-1})^{(g,g)}(\tilde{\Gamma}^{-1})^{(h,h)}U\right)_{ij}.$$

Since d and the entries of O and Λ are independent of n and bounded, the magnitude of the left hand side is solely determined by the following term:

$$\left(U^\intercal(\tilde{\Gamma}^{-1})^{(g,g)}(\tilde{\Gamma}^{-1})^{(h,h)}U\right)_{ij} = \sum_{l=1}^{n} U_{li}U_{lj}(\tilde{\Gamma}^{-1})_{ll}^{(g,g)}(\tilde{\Gamma}^{-1})_{ll}^{(h,h)}.$$

By calculating the variances of M_n^i for i = 1, 2, 3, using similar integration approximations in Step (2) and Chebyshev's inequality, we can establish the point-wise convergence:

$$\Psi_n - \bar{\Psi}_n = O_p(n^{-1/4}).$$

To prove the desired uniform convergence, we first note that maximising over Σ is equivalent to over (O, D) as both contain $d(d+1)/2$ number of parameters. Then, for two sets of parameters (Σ, Λ) and $(\breve{\Sigma}, \breve{\Lambda})$, or correspondingly, Ω and $\breve{\Omega}$, using (A.7) we have

$$\begin{aligned}&(\Psi_n(\Sigma,\Lambda) - \bar{\Psi}_n(\Sigma,\Lambda)) - (\Psi_n(\breve{\Sigma},\breve{\Lambda}) - \bar{\Psi}_n(\breve{\Sigma},\breve{\Lambda})) \qquad (A.9)\\ = & -\frac{\Delta}{2\sqrt{n}}\mathrm{vec}\left((\Omega^{-1} - \breve{\Omega}^{-1})\left(rr^\intercal - \Omega^\star\right)\Omega^{-1}\right.\\ &\left. - \breve{\Omega}^{-1}\left(rr^\intercal - \Omega^\star\right)(\breve{\Omega}^{-1} - \Omega^{-1})\right)^\intercal(\mathbb{I}_d \otimes G)\mathbb{D}_d.\end{aligned}$$

By (A.8), we can decompose

$$\begin{aligned}&\Omega^{-1} - \breve{\Omega}^{-1}\\ = & \left(\left(O\Lambda^{-1/2} - \breve{O}\breve{\Lambda}^{-1/2}\right)\otimes U\right)^\intercal\left(\Delta D\otimes \mathbb{I}_n + \mathbb{I}_d\otimes\tilde{D}\right)^{-1}\\ &\times\left(\left(O\Lambda^{-1/2}\right)\otimes U\right)\\ &+\left(\left(\breve{O}\breve{\Lambda}^{-1/2}\right)\otimes U\right)^\intercal\left(\left(\Delta D\otimes\mathbb{I}_n + \mathbb{I}_d\otimes\tilde{D}\right)^{-1}\right.\\ &\left. - \left(\Delta\breve{D}\otimes\mathbb{I}_n + \mathbb{I}_d\otimes\tilde{D}\right)^{-1}\right)\left(\left(O\Lambda^{-1/2}\right)\otimes U\right)\\ &+\left(\left(\breve{O}\breve{\Lambda}^{-1/2}\right)\otimes U\right)^\intercal\left(\Delta\breve{D}\otimes\mathbb{I}_n + \mathbb{I}_d\otimes\tilde{D}\right)^{-1}\\ &\times\left(\left(O\Lambda^{-1/2} - \breve{O}\breve{\Lambda}^{-1/2}\right)\otimes U\right).\end{aligned}$$

Then to bound (A.9), we separately consider each of these three terms and follow similar variance calculations as above. This yields

$$\begin{aligned}&\|(\Psi_n(\Sigma,\Lambda) - \bar{\Psi}_n(\Sigma,\Lambda)) - (\Psi_n(\breve{\Sigma},\breve{\Lambda}) - \bar{\Psi}_n(\breve{\Sigma},\breve{\Lambda}))\|\\ &\le O_p(1)\times\left(\|O\lambda^{-1/2} - \breve{O}\breve{\lambda}^{-1/2}\| + \|D - \breve{D}\|\right),\end{aligned}$$

which in turn leads to stochastic equicontinuity, see, e.g., Newey and McFadden (1994). As a result, we obtain the desired uniform convergence of Ψ_n to $\bar{\Psi}_n$. Similarly, we can prove the uniform convergence of Φ_n and $\bar{\Phi}_n$, and hence establish (A.6) and the desired consistency.

Moreover, by invoking the classic central limit theorem of quadratic forms, e.g., de Jong (1987), and tedious verifications of the Lindeberg-type condition and similar calculations of the asymptotic variance as in Step (2), we can show that

$$M_n^1 + M_n^2 + M_n^3 \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, V^1(\Sigma,\Lambda) + V^2(\Sigma,\Lambda) + V^3(\Sigma,\Lambda)\right),$$

where

for any $1 \le i, j, i', j' \le d$, we have

$$\begin{aligned}&\left((O\Lambda^{1/2}\otimes O\Lambda^{1/2})\left(V^1(\Sigma,\Lambda) + V^2(\Sigma,\Lambda) + V^3(\Sigma,\Lambda)\right)\right.\\ &\left.\times\ (\Lambda^{1/2}O^\intercal\otimes\Lambda^{1/2}O^\intercal)\right)_{(i-1)d+j,(i'-1)d+j'}\\ = & \frac{T^{1/2}}{8}\left(\delta_{i,i'}\delta_{j,j'} + \delta_{i,j'}\delta_{i',j}\right)\times\frac{1}{D_{ii}^{1/2}D_{jj}^{1/2}(D_{ii}^{1/2} + D_{jj}^{1/2})}.\end{aligned}$$

Using (A.5), the fact that $\mathbb{N}_d(A\otimes A)\mathbb{D}_d^{+\intercal} = (A\otimes A)\mathbb{D}_d^{+\intercal}$ for any $d\times d$ matrix A, and the sandwich theorem, we obtain that the asymptotic variance of the estimator for vech($\widehat{\Sigma}$):

$$\begin{aligned}&\left(\mathbb{D}_d^+ H\mathbb{D}_d^{+\intercal}\right)\mathbb{D}_d^\intercal\left(V^1(\Sigma,\Lambda) + V^2(\Sigma,\Lambda) + V^3(\Sigma,\Lambda)\right)\mathbb{D}_d\left(\mathbb{D}_d^+ H\mathbb{D}_d^{+\intercal}\right)\\ &= \mathbb{D}_d^+ H\mathbb{D}_d^{+\intercal},\end{aligned}$$

which concludes the proof. □

A.2. Proof of Theorem 2

Proof. Again, we use superscript $\star$ to mark quantities that are made of true values, in particular, $\Sigma^\star = \frac{1}{T}\int_0^T \Sigma_s ds$. The pseudo-true parameters are marked with a superscript such as Σ^* and Λ^*, and the QMLs are marked as $\widehat{\Sigma}$ and $\widehat{\Lambda}$.

Step (1) Recall that in (11), we have

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

Now we consider the following function:

$$L_n^* = -n\log(2\pi) - \frac{1}{2}\log(\det\Omega) - \frac{1}{2}tr(\Omega^{-1}\Omega^\star),$$

where the (i, j)th entry of the (p, q)-block is given by

$$\Omega_{ij}^{\star,(p,q)} = \int_{t_{i-1}}^{t_i}\Sigma_{pq,t}dt + \left(2\Lambda_{pp}^\star\delta_{i,j} - \Lambda_{pp}^\star\delta_{i,j+1} - \Lambda_{pp}^\star\delta_{i,j-1}\right)\delta_{p,q}. \qquad (A.10)$$

Writing

$$\Psi_n = -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{vech}(\Sigma)}(L_n), \quad \Psi_n^* = -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{vech}(\Sigma)}(L_n^*),$$

so by the same derivation as in the previous proof, we can obtain the following decomposition:

$$n^{1/4}\left(\Psi_n - \Psi_n^*\right) = -\left(M_n^1 + M_n^2 + M_n^3\right)\mathbb{D}_d,$$

where for any $1 \le p, q \le d$,

$$\begin{aligned}M_{n,(p-1)d+q}^1 = & \frac{\Delta}{2n^{1/4}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left((\Omega^{-1})^{(k,q)}(\Omega^{-1})^{(p,l)}\right)_{ij}\\ &\times\left(\left(\Delta_i^n y_k\right)\left(\Delta_j^n y_l\right) - \delta_{i,j}\int_{\tau_{i-1}}^{\tau_i}\Sigma_{kl,s}ds\right),\end{aligned}$$

and $M_{n,(p-1)d+q}^2$ and $M_{n,(p-1)d+q}^3$ take on the identical forms as in the previous proof. We further decompose $M_{n,(p-1)d+q}^1$ as

$$\begin{aligned}M_{n,(p-1)d+q}^1 = & \frac{\Delta}{2n^{1/4}}\sum_{i=1}^{n}\sum_{j>i}\sum_{k,l=1}^{d}\left(\left((\Omega^{-1})^{(k,q)}(\Omega^{-1})^{(p,l)}\right)_{ij}\right.\\ &\left.+\left((\Omega^{-1})^{(l,q)}(\Omega^{-1})^{(p,k)}\right)_{ji}\right)\\ &\times\left(\Delta_i^n y_k\right)\left(\Delta_j^n y_l\right)\\ &+\frac{\Delta}{2n^{1/4}}\sum_{i=1}^{n}\sum_{k,l=1}^{d}\left(\left((\Omega^{-1})^{(k,q)}(\Omega^{-1})^{(p,l)}\right)_{ii}\right.\\ &\left.+\left((\Omega^{-1})^{(l,q)}(\Omega^{-1})^{(p,k)}\right)_{ii}\right)\\ &\times\left(\left(\Delta_i^n y_k\right)\left(\Delta_i^n y_l\right) - \int_{\tau_{i-1}}^{\tau_i}\Sigma_{kl,s}ds\right).\end{aligned}$$

By tedious calculations, we can show that the second term above is dominated by the first term asymptotically. Following the same calculations previously, we can show that

$$\Psi_n - \Psi_n^* = O_p(n^{-1/4}),$$

and by Jacod (2012, Thm 7.1), we can show that M_n^i, $i = 1, 2, 3$ are asymptotically independent, and that

$$n^{1/4}\left(\Psi_n - \Psi_n^*\right) = \left(M_n^1 + M_n^2 + M_n^3\right)\mathbb{D}_d \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \mathbb{D}_d^{\top}\left(V^1(\Sigma, \Lambda) + V^2(\Sigma, \Lambda) + V^3(\Sigma, \Lambda)\right)\mathbb{D}_d\right), \tag{A.11}$$

where

$$V^1(\Sigma, \Lambda)_{(p-1)d+q,(p'-1)d+q'} \tag{A.12}$$
$$= \frac{T^3}{4}\sum_{k,k',l,l',g,g',h,h'=1}^{d} O_{gk}O_{gq}O_{hp}O_{hl}O_{g'k'}O_{g'q'}O_{h'p'}O_{h'l'} \times \Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2}\Lambda_{k'k'}^{-1/2}\Lambda_{l'l'}^{-1/2}\Lambda_{q'q'}^{-1/2}\Lambda_{p'p'}^{-1/2}$$
$$\times \int_0^\infty \left(D_{gg}T + \pi^2x^2\right)^{-1}\left(D_{hh}T + \pi^2x^2\right)^{-1}\left(D_{g'g'}T + \pi^2x^2\right)^{-1} \times \left(D_{h'h'}T + \pi^2x^2\right)^{-1}dx$$
$$\times \left(\int_0^T \Sigma_{kk',s}\Sigma_{ll',s}ds + \int_0^T \Sigma_{kl',s}\Sigma_{lk',s}ds\right),$$

$$V^2(\Sigma, \Lambda)_{(p-1)d+q,(p'-1)d+q'} \tag{A.13}$$
$$= \frac{T^2}{4}\sum_{k,k',l,l',g,g',h,h'=1}^{d} O_{gk}O_{gq}O_{hp}O_{hl}O_{g'k'}O_{g'q'}O_{h'p'}O_{h'l'} \times \Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2}\Lambda_{k'k'}^{-1/2}\Lambda_{l'l'}^{-1/2}\Lambda_{q'q'}^{-1/2}\Lambda_{p'p'}^{-1/2}$$
$$\times \int_0^\infty \pi^2x^2\left(D_{gg}T + \pi^2x^2\right)^{-1}\left(D_{g'g'}T + \pi^2x^2\right)^{-1} \times \left(D_{hh}T + \pi^2x^2\right)^{-1}\left(D_{h'h'}T + \pi^2x^2\right)^{-1}dx$$
$$\times \left(\delta_{ll'}\Lambda_{ll}^{\star}\int_0^T \Sigma_{kk',s}ds + \delta_{lk'}\Lambda_{ll}^{\star}\int_0^T \Sigma_{kl',s}ds + \delta_{kl'}\Lambda_{kk}^{\star}\int_0^T \Sigma_{lk',s}ds + \delta_{kk'}\Lambda_{kk}^{\star}\int_0^T \Sigma_{ll',s}ds\right),$$

$$V^3(\Sigma, \Lambda)_{(p-1)d+q,(p'-1)d+q'} \tag{A.14}$$
$$= \frac{T^2}{4}\sum_{k,k',l,l',g,g',h,h'=1}^{d} O_{gk}O_{gq}O_{hp}O_{hl}O_{g'k'}O_{g'q'}O_{h'p'}O_{h'l'} \times \Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2}\Lambda_{k'k'}^{-1/2}\Lambda_{l'l'}^{-1/2}\Lambda_{q'q'}^{-1/2}\Lambda_{p'p'}^{-1/2}$$
$$\times \int_0^\infty \pi^4x^4\left(D_{gg}T + \pi^2x^2\right)^{-1}\left(D_{g'g'}T + \pi^2x^2\right)^{-1} \times \left(D_{hh}T + \pi^2x^2\right)^{-1}\left(D_{h'h'}T + \pi^2x^2\right)^{-1}dx$$
$$\times \left(\delta_{k,k'}\delta_{l,l'} + \delta_{k,l'}\delta_{k',l}\right)\Lambda_{kk}^{\star}\Lambda_{ll}^{\star},$$

where we use the fact that

$$\mathrm{E}\left\{\left[\left(\Delta_i^n\varepsilon_k\right)\left(\Delta_j^n\varepsilon_l\right) - \mathrm{E}\left(\Delta_i^n\varepsilon_k\Delta_j^n\varepsilon_l\right)\right] \times \left[\left(\Delta_{i'}^n\varepsilon_{k'}\right)\left(\Delta_{j'}^n\varepsilon_{l'}\right) - \mathrm{E}\left(\Delta_{i'}^n\varepsilon_{k'}\Delta_{j'}^n\varepsilon_{l'}\right)\right]\right\}$$
$$= +\,\delta_{k,k'}\delta_{l,l'}\left(2\delta_{i,i'} - \delta_{i,i'-1} - \delta_{i,i'+1}\right) \times \left(2\delta_{j,j'} - \delta_{j,j'-1} - \delta_{j,j'+1}\right)\Lambda_{kk}^{\star}\Lambda_{ll}^{\star}$$
$$+\,\delta_{k,l'}\delta_{k',l}\left(2\delta_{i,j'} - \delta_{i,j'-1} - \delta_{i,j'+1}\right) \times \left(2\delta_{j,i'} - \delta_{j,i'-1} - \delta_{j,i'+1}\right)\Lambda_{kk}^{\star}\Lambda_{ll}^{\star}$$
$$+\,\delta_{k,l,k',l'}\mathrm{cum}[\Delta_i^n\varepsilon_k, \Delta_j^n\varepsilon_k, \Delta_{i'}^n\varepsilon_k, \Delta_{j'}^n\varepsilon_k]\Lambda_{kk}^{\star 2},$$

and that

$$\mathrm{cum}\left(\Delta_i^n\varepsilon_k, \Delta_j^n\varepsilon_k, \Delta_{i'}^n\varepsilon_k, \Delta_{j'}^n\varepsilon_k\right) = \begin{cases} 0, & \text{if } \max(i,j,i',j') - \min(i,j,i',j') \geq 2, \\ 2\mathrm{cum}_4(\varepsilon_k), & \text{if } i = j = i' = j', \\ (-1)^s\mathrm{cum}_4(\varepsilon_k), & \text{if } s = \#\left\{\{i,j,i',j'\} = \min(i,j,i',j')\right\}, \end{cases}$$

where # denotes the cardinality of a set.

Similarly, since Λ is a diagonal matrix, we define

$$\Phi_n = -\frac{1}{n}\mathrm{D}_{\mathrm{diag}(\Lambda)}(L_n), \quad \Phi_n^* = -\frac{1}{n}\mathrm{D}_{\mathrm{diag}(\Lambda)}(L_n^*),$$

and we can show that

$$n^{1/2}\left(\Phi_n - \Phi_n^*\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, (4\Lambda^4)^{-1}\left(2\Lambda^{\star 2} + \mathrm{diag}\left(\mathrm{cum}_4(\varepsilon)\right)\right)\right),$$

where $\mathrm{diag}(\mathrm{cum}_4(\varepsilon))$ is a $d \times d$ matrix with $(\mathrm{cum}_4(\varepsilon_1), \ldots, \mathrm{cum}_4(\varepsilon_d))$ on the diagonal.

Step (2) We now need to solve $\Psi_n^* = 0$ and $\Phi_n^* = 0$ for the pseudo-true parameters Σ^* and Λ^*, and show that the distance between them and the values of interest are negligible asymptotically.

Note that

$$0 = \Psi_n^* = -\frac{\Delta}{2\sqrt{n}}\mathrm{vec}\left(\Omega^{-1}\left(\Omega^{\star} - \Omega\right)\Omega^{-1}\right)^{\top}(\mathbb{I}_d \otimes G)\mathbb{D}_d$$
$$= -\frac{\Delta}{2\sqrt{n}}\sum_{j=1}^{d}\sum_{i=1}^{n}\mathrm{vec}\left(\left(\Omega^{-1}\left(\epsilon_j \otimes e_i\right)\right)^{\top}\left(\Omega^{\star} - \Omega\right) \times \Omega^{-1}\left(\mathbb{I}_d \otimes \left(e_i\epsilon_j^{\top}\right)\right)\right)^{\top}\mathbb{D}_d,$$

which implies that, for any $1 \leq p, q \leq d$,

$$0 = \frac{\Delta}{2\sqrt{n}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij} \times \left(\int_{\tau_{i-1}}^{\tau_i}\Sigma_{kl,s}ds - \Sigma_{kl}\Delta\right)\delta_{i,j}$$
$$+ \frac{\Delta}{2\sqrt{n}}\sum_{k,l=1}^{d}\sum_{i,j=1}^{n}\left(\left(\Omega^{-1}\right)^{(k,q)}\left(\Omega^{-1}\right)^{(p,l)}\right)_{ij}\left(\Lambda_{kk}^{\star} - \Lambda_{kk}\right) \times \delta_{k,l}\left(2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1}\right)$$
$$= \frac{T}{2}\sum_{k=1}^{d}\sum_{g,h=1}^{d}O_{gk}O_{gq}O_{hp}O_{hl}\Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2} \times \int_0^\infty \frac{1}{\left(D_{gg}T + \pi^2x^2\right)\left(D_{hh}T + \pi^2x^2\right)}dx \times \left(\int_0^T \Sigma_{kl,s}ds - \Sigma_{kl}T\right)(1 + o_p(1))$$
$$+ \frac{T}{2}\sum_{k=1}^{d}\sum_{g,h=1}^{d}O_{gk}O_{gq}O_{hp}O_{hk}\Lambda_{kk}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{qq}^{-1/2}\Lambda_{pp}^{-1/2} \times \left(\Lambda_{kk}^{\star} - \Lambda_{kk}\right) \times \int_0^\infty \frac{\pi^2x^2}{\left(D_{gg}T + \pi^2x^2\right)\left(D_{hh}T + \pi^2x^2\right)}dx \times (1 + o_p(1)). \tag{A.15}$$

Similarly, we have

$$0 = \Phi_n^* = -\frac{1}{2n}\mathrm{vec}\left(\Omega^{-1}\left(\Omega^\star - \Omega\right)\Omega^{-1}\right)^\intercal (\mathbb{I}_d \otimes G')\mathbb{E}_d$$
$$= -\frac{1}{2n}\sum_{j=1}^{d}\sum_{i=1}^{n}\mathrm{vec}\left(\left(\Omega^{-1}\left(\epsilon_j \otimes (2e_i - e_{i-1} - e_{i+1})\right)\right)^\intercal \times \left(\Omega^\star - \Omega\right)\Omega^{-1}\left(\mathbb{I}_d \otimes \left(e_i\epsilon_j^\intercal\right)\right)\right)^\intercal \mathbb{E}_d,$$

where $G' = (\mathbb{K}_{n,d} \otimes \mathbb{I}_n)(\mathbb{I}_d \otimes \mathrm{vec}(\mathbb{J}_n))$. This implies that for any $1 \le p \le d$,

$$0 = \frac{1}{2n}\sum_{k,l=1}^{d}\sum_{i,j,u,v=1}^{n}\left(\Omega^{-1}\right)_{ju}^{(p,k)}\left(\Omega^{-1}\right)_{vi}^{(l,p)}\left(2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1}\right)$$
$$\times \left(\left(\left(\int_{\tau_{u-1}}^{\tau_u}\Sigma_{kl,s}ds - \Sigma_{kl}\Delta\right)\delta_{u,v} + \left(\Lambda_{kk}^\star - \Lambda_{kk}\right)\delta_{k,l}\left(2\delta_{uv} - \delta_{u,v-1} - \delta_{u,v+1}\right)\right)\right)$$
$$= \frac{1}{2\sqrt{n}}\sum_{k,l,g,h=1}^{d} O_{gp}O_{gk}O_{hl}O_{hp}\Lambda_{pp}^{-1}\Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}$$
$$\times \int_0^\infty \frac{\pi^2x^2dx}{\left(D_{gg}T + \pi^2x^2\right)\left(D_{hh}T + \pi^2x^2\right)}$$
$$\times \left(\int_0^T \Sigma_{kl,s}ds - \Sigma_{kl}T\right) \times (1 + o_p(1))$$
$$+ \frac{1}{2}\left(\Lambda_{pp}^\star - \Lambda_{pp}\right)\Lambda_{pp}^{-2} - \left(\Lambda_{pp}^\star - \Lambda_{pp}\right)(1 + o_p(1))$$
$$\times \frac{1}{2\sqrt{n}}\sum_{k,l,g,h=1}^{d} O_{gp}O_{gk}O_{hl}O_{hp}\Lambda_{pp}^{-1}\Lambda_{kk}^{-1/2}\Lambda_{ll}^{-1/2}$$
$$\times \int_0^\infty \frac{D_{gg}D_{hh}T^2 + 2T\left(D_{gg} + D_{hh}\right)\pi^2x^2}{\left(D_{gg}T + \pi^2x^2\right)\left(D_{hh}T + \pi^2x^2\right)}dx, \tag{A.16}$$

where we use the approximation

$$\sum_{m=1}^{n}\left(\tilde{\Gamma}^{-1}\right)_{mm}^{(g,g)}\left(\tilde{\Gamma}^{-1}\right)_{mm}^{(h,h)}4\left(1 - \cos\left(\frac{m\pi}{n+1}\right)\right)^2$$
$$= n - \sum_{m=1}^{n}\left(\tilde{\Gamma}^{-1}\right)_{mm}^{(g,g)}\left(\tilde{\Gamma}^{-1}\right)_{mm}^{(h,h)} \times \left(\Delta^2 D_{gg}D_{hh} + 2\Delta\left(D_{gg} + D_{hh}\right)\times\left(1 - \cos\left(\frac{m\pi}{n+1}\right)\right)\right)$$
$$= n - n^{1/2}\int_0^\infty \frac{D_{gg}D_{hh}T^2 + 2T\left(D_{gg} + D_{hh}\right)\pi^2x^2}{\left(D_{gg}T + \pi^2x^2\right)\left(D_{hh}T + \pi^2x^2\right)}dx \times (1 + o(1)).$$

Solving Eq. (A.15) and (A.16), we obtain for any $1 \le p, k, l \le d$,

$$\Lambda_{pp}^* = \Lambda_{pp}^\star + o_p(n^{-1/2}), \quad \text{and} \quad \Sigma_{kl}^* = \frac{1}{T}\int_0^T \Sigma_{kl,s}ds + O_p(n^{-1/2}).$$

Also, we can show by direct calculations that for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\mathrm{plim}\left(\min_{\|\Sigma - \Sigma^*\| \ge \epsilon, \|\Lambda - \Lambda^*\| \ge \epsilon}\|\Psi_n^*\| + \|\Phi_n^*\| \ge \delta\right) \to 1.$$

Moreover, following the same steps as in the previous proof, we can establish uniform convergence of Ψ_n and Φ_n:

$$\sup_{\Sigma,\Lambda}\|\Psi_n - \Psi_n^*\| \xrightarrow{p} 0, \quad \sup_{\Sigma,\Lambda}\|\Phi_n - \Phi_n^*\| \xrightarrow{p} 0.$$

Applying Xiu (2010, Thm 2), $\hat{\Sigma}_{kl} - \frac{1}{T}\int_0^T \Sigma_{kl,s}ds = o_p(1)$ and $\hat{\Lambda}_{pp} - \Lambda_{pp}^\star = o_p(1)$, so consistency follows.

Step (3) Finally, we use the sandwich theorem and consistency to establish the CLT for the QML.

$$\mathrm{D}_{\mathrm{vech}(\Sigma)}(\Psi_n^*) \tag{A.17}$$
$$= \frac{\Delta^2}{2}\mathbb{D}_d^\intercal(\mathbb{I}_d \otimes G^\intercal)\left(\Omega^{-1} \otimes \Omega^{-1} + \left(\Omega^{-1}\left(\Omega^\star - \Omega\right)\Omega^{-1}\right) \otimes \Omega^{-1} + \Omega^{-1} \otimes \left(\Omega^{-1}\left(\Omega^\star - \Omega\right)\Omega^{-1}\right)\right)(\mathbb{I}_d \otimes G)\mathbb{D}_d$$
$$:= \mathbb{D}_d^\intercal\tilde{V}_n\mathbb{D}_d,$$

where the (u, v)-entry of the (j, k)-block of the $d^2 \times d^2$ matrix $\tilde{V}_n$ is given by

$$\tilde{V}_{n,uv}^{(j,k)} = +\frac{\Delta^2}{2\sqrt{n}}\sum_{l,h=1}^{d} O_{lj}O_{lk}O_{hu}O_{hv}\Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}$$
$$\times \sum_{m=1}^{n}\left(\Delta D_{ll} + \tilde{D}_{mm}\right)^{-1}\left(\Delta D_{hh} + \tilde{D}_{mm}\right)^{-1}$$
$$+ \frac{\Delta^2}{\sqrt{n}}\sum_{p,q,l,g,h=1}^{d} O_{gq}O_{hu}O_{lj}O_{hv}O_{gk}O_{lp}\Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}$$
$$\times \Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}\Lambda_{pp}^{-1/2}\Lambda_{qq}^{-1/2}$$
$$\times \sum_{m=1}^{n}\left(\Delta D_{ll} + \tilde{D}_{mm}\right)^{-1}\left(\Delta D_{gg} + \tilde{D}_{mm}\right)^{-1}$$
$$\times \left(\Delta D_{hh} + \tilde{D}_{mm}\right)^{-1}$$
$$\times \left(\sum_{r=1}^{n} U_{mr}^2\left(\int_{\tau_{r-1}}^{\tau_r}\Sigma_{pq,s}ds - \Sigma_{pq}\Delta\right) + \tilde{D}_{mm}\left(\Lambda_{pp}^\star - \Lambda_{pp}\right)\delta_{pq}\right)$$
$$\xrightarrow{p} +\frac{T^2}{2}\sum_{l,h=1}^{d} O_{lj}O_{lk}O_{hu}O_{hv}\Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}$$
$$\times \int_0^\infty \left(TD_{ll} + \pi^2x^2\right)^{-1}\left(TD_{hh} + \pi^2x^2\right)^{-1}dx$$
$$+ T^2\sum_{p,q,l,g,h=1}^{d} O_{gq}O_{hu}O_{lj}O_{hv}O_{gk}O_{lp}$$
$$\times \Lambda_{jj}^{-1/2}\Lambda_{kk}^{-1/2}\Lambda_{uu}^{-1/2}\Lambda_{vv}^{-1/2}\Lambda_{pp}^{-1/2}\Lambda_{qq}^{-1/2}$$
$$\times \left\{\int_0^\infty \left(TD_{ll} + \pi^2x^2\right)^{-1}\left(TD_{gg} + \pi^2x^2\right)^{-1}\right.$$
$$\times \left(TD_{hh} + \pi^2x^2\right)^{-1}dx \times \left(\int_0^T \Sigma_{pq,s}ds - \Sigma_{pq}T\right)$$
$$+ \int_0^\infty \pi^2x^2\left(TD_{ll} + \pi^2x^2\right)^{-1}\left(TD_{gg} + \pi^2x^2\right)^{-1}$$
$$\left.\times \left(TD_{hh} + \pi^2x^2\right)^{-1}dx \times \left(\Lambda_{pp}^\star - \Lambda_{pp}\right)\delta_{pq}\right\}.$$

By consistency, we have

$$\tilde{V}_{n,uv}^{(j,k)}|_{\Sigma=\Sigma^\star,\Lambda=\Lambda^\star}$$
$$\xrightarrow{p} \frac{T^2}{2}\sum_{l,h=1}^{d} O_{lj}^\star O_{lk}^\star O_{hu}^\star O_{hv}^\star\Lambda_{jj}^{\star-1/2}\Lambda_{kk}^{\star-1/2}\Lambda_{uu}^{\star-1/2}\Lambda_{vv}^{\star-1/2}$$
$$\times \int_0^\infty \left(TD_{ll}^\star + \pi^2x^2\right)^{-1}\left(TD_{hh}^\star + \pi^2x^2\right)^{-1}dx$$
$$= \frac{T^{5/2}}{4}\sum_{l,h=1}^{d} O_{lj}^\star O_{lk}^\star O_{hu}^\star O_{hv}^\star\Lambda_{jj}^{\star-1/2}\Lambda_{kk}^{\star-1/2}\Lambda_{uu}^{\star-1/2}\Lambda_{vv}^{\star-1/2}$$
$$\times \left(D_{mm}^\star D_{ll}^{\star1/2} + D_{mm}^{\star1/2}D_{ll}^\star\right)^{-1},$$

where $D^\star = O^\star \Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}O^{\star\intercal}$. Using the same approximation as in the previous proof, we have

$$\begin{aligned}
&\left(\mathrm{D}_{\mathrm{vech}(\Sigma)}(\Psi_n^*)\right)^{-1}\big|_{\Sigma=\Sigma^\star,\Lambda=\Lambda^\star} \xrightarrow{p} 4T^{-1/2}\times \\
&\mathbb{D}_d^+\left\{\Sigma^\star\otimes\left(\Lambda^{\star 1/2}\left(\Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}\right)^{1/2}\Lambda^{\star 1/2}\right)\right. \\
&\left. +\ \left(\Lambda^{\star 1/2}\left(\Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}\right)^{1/2}\Lambda^{\star 1/2}\right)\otimes\Sigma^\star\right\}\mathbb{D}_d^{+\intercal} \\
=:&\ \mathbb{D}_d^+ H(\Sigma^\star,\Lambda^\star)\mathbb{D}_d^{+\intercal}. \qquad \text{(A.18)}
\end{aligned}$$

Note that $\widehat{\Lambda}$ converges at a faster rate, so by Taylor expansion and the sandwich formula, the central limit theorem for $\widehat{\Sigma}$ is

$$n^{1/4}\left(\mathrm{vech}\left(\widehat{\Sigma}\right)-\mathrm{vech}\left(\frac{1}{T}\int_0^T\Sigma_s ds\right)\right)\xrightarrow{\mathcal{L}_X}\mathcal{MN}\left(0,\Pi^{\mathrm{Q}}\right),$$

where

$$\begin{aligned}
\Pi^{\mathrm{Q}} &= \mathbb{D}_d^+ H(\Sigma^\star,\Lambda^\star)\left(V^1\left(\Sigma^\star,\Lambda^\star\right)+V^2\left(\Sigma^\star,\Lambda^\star\right)\right. \\
&\quad \left. +\ V^3\left(\Sigma^\star,\Lambda^\star\right)\right)H(\Sigma^\star,\Lambda^\star)\mathbb{D}_d^{+\intercal}, \qquad \text{(A.19)}
\end{aligned}$$

V^1, V^2, and V^3 are given by (A.12), (A.13) and (A.14), in which $D^\star = O^\star\Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}O^{\star\intercal}$, and

$$\begin{aligned}
&H(\Sigma^\star,\Lambda^\star) \qquad \text{(A.20)} \\
&= 4T^{-1/2}\left\{\Sigma^\star\otimes\left(\Lambda^{\star 1/2}\left(\Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}\right)^{1/2}\Lambda^{\star 1/2}\right)\right. \\
&\quad \left. +\ \left(\Lambda^{\star 1/2}\left(\Lambda^{\star-1/2}\Sigma^\star\Lambda^{\star-1/2}\right)^{1/2}\Lambda^{\star 1/2}\right)\otimes\Sigma^\star\right\}. \qquad \text{(A.21)}
\end{aligned}$$

In the above we use Equation (7) of Section 3.8 in Magnus and Neudecker (1999) and the fact that $\mathbb{N}_d V^i(\Sigma^\star,\Lambda^\star)\mathbb{N}_d = V^i(\Sigma^\star,\Lambda^\star)$, for $i = 1, 2, 3$. □

A.3. Proof of Theorem 3

Proof. Again, by Theorems 2.2, 8.2, and 8.3, Equation (4) of Section 2.4, and Ex. 1 of Section 3.7 of Magnus and Neudecker (1999), we have

$$\begin{aligned}
\mathrm{vec}(d\bar{L}_n) &= -\frac{1}{2}\mathrm{vec}(\bar{\Omega}^{-1})^\intercal\mathrm{vec}(d\bar{\Omega}) \\
&\quad +\frac{1}{2}\mathrm{vec}(\bar{\Omega}^{-1}\bar{r}\bar{r}^\intercal\bar{\Omega}^{-1})^\intercal\mathrm{vec}(d\bar{\Omega}), \qquad \text{(A.22)} \\
\mathrm{vec}(d\bar{\Omega}) &= \Delta(\mathbb{I}_{d+f}\otimes\bar{G})\mathrm{vec}(d\bar{\Sigma})+(\mathbb{I}_{d+f}\otimes\bar{G}')\mathrm{vec}(d\bar{\Lambda}), \qquad \text{(A.23)} \\
\mathrm{vec}(d\bar{\Sigma}) &= \left([\mathbb{I}_d:0]^\intercal\otimes[\mathbb{I}_d:0]^\intercal\right)\mathbb{E}_d\mathrm{vec}(\mathrm{diag}(d\Sigma^e)) \\
&\quad +\left(\left[\beta^\intercal:\mathbb{I}_f\right]^\intercal\otimes\left[\beta^\intercal:\mathbb{I}_f\right]^\intercal\right)\mathrm{vec}(d\Sigma^f) \\
&\quad +\left(\mathbb{I}_{(d+f)^2}+\mathbb{K}_{d+f,d+f}\right)\left(\left[\Sigma^f\beta^\intercal:\Sigma^f\right]^\intercal\right. \\
&\quad \times\left.\otimes[\mathbb{I}_d:0]^\intercal\right)\mathrm{vec}(d\beta), \qquad \text{(A.24)}
\end{aligned}$$

where $\bar{G} = (\mathbb{K}_{n,d+f}\otimes\mathbb{I}_n)(\mathbb{I}_{d+f}\otimes\mathrm{vec}(\mathbb{I}_n))$, $\bar{G}' = (\mathbb{K}_{n,d+f}\otimes\mathbb{I}_n)(\mathbb{I}_{d+f}\otimes\mathrm{vec}(\mathbb{J}_n))$, $[\mathbb{I}_d:0]$ is a $d\times(d+f)$ matrix, and $\mathbb{E}_d = (e_1e_1^\intercal : e_2e_2^\intercal : \ldots : e_de_d^\intercal)^\intercal$ is a $d^2\times d$ matrix. Note that : denotes matrix concatenation horizontally. We define

$$\bar{\Psi}_n = -\frac{1}{\sqrt{n}}\mathrm{D}_{\bar{\Sigma}}(\bar{L}_n),\quad \bar{\Psi}_n^f = -\frac{1}{\sqrt{n}}\mathrm{D}_{\Sigma^f}(\bar{L}_n),$$

$$\bar{\Psi}_n^e = -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{L}_n),\quad \bar{\Psi}_n^\beta = -\frac{1}{\sqrt{n}}\mathrm{D}_\beta(\bar{L}_n),$$

$$\bar{\Psi}_n^* = -\frac{1}{\sqrt{n}}\mathrm{D}_{\bar{\Sigma}}(\bar{L}_n^*),\quad \bar{\Psi}_n^{f*} = -\frac{1}{\sqrt{n}}\mathrm{D}_{\Sigma^f}(\bar{L}_n^*),$$

$$\bar{\Psi}_n^{e*} = -\frac{1}{\sqrt{n}}\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{L}_n^*),\quad \bar{\Psi}_n^{\beta*} = -\frac{1}{\sqrt{n}}\mathrm{D}_\beta(\bar{L}_n^*),$$

where $\bar{\Omega}^\star$ is similarly defined as in (A.10), and

$$\bar{L}_n^* = -n\log(2\pi)-\frac{1}{2}\log(\det\bar{\Omega})-\frac{1}{2}tr(\bar{\Omega}^{-1}\bar{\Omega}^\star).$$

Note that we can apply the same argument as in the proof of Theorem 2 to obtain a similar result as in (A.11):

$$\begin{aligned}
&n^{1/4}\left(\bar{\Psi}_n-\bar{\Psi}_n^*\right) \\
&\xrightarrow{\mathcal{L}_X}\mathcal{MN}\left(0,\left(V^1(\Sigma,\Lambda)+V^2(\Sigma,\Lambda)+V^3(\Sigma,\Lambda)\right)\right). \qquad \text{(A.25)}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&n^{1/4}\left(\bar{\Psi}_n^e-\bar{\Psi}_n^{e*},\bar{\Psi}_n^f-\bar{\Psi}_n^{f*},\bar{\Psi}_n^\beta-\bar{\Psi}_n^{\beta*}\right) \\
&= n^{1/4}\left(\bar{\Psi}_n-\bar{\Psi}_n^*\right)\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma}),\mathrm{D}_{\Sigma^f}(\bar{\Sigma}),\mathrm{D}_\beta(\bar{\Sigma})\right) \\
&\xrightarrow{\mathcal{L}_X}\mathcal{MN}\left(0,\mathbb{C}^\intercal\left(V^1(\Sigma,\Lambda)+V^2(\Sigma,\Lambda)+V^3(\Sigma,\Lambda)\right)\mathbb{C}\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{C} &= \left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma}),\mathrm{D}_{\Sigma^f}(\bar{\Sigma}),\mathrm{D}_\beta(\bar{\Sigma})\right) \\
&= \left(\left([\mathbb{I}_d:0]^\intercal\otimes[\mathbb{I}_d:0]^\intercal\right)\mathbb{E}_d,\left(\left[\beta^\intercal:\mathbb{I}_f\right]^\intercal\otimes\left[\beta^\intercal:\mathbb{I}_f\right]^\intercal\right),\right. \\
&\quad \left.\left(\mathbb{I}_{(d+f)^2}+\mathbb{K}_{d+f,d+f}\right)\left(\left[\Sigma^f\beta^\intercal:\Sigma^f\right]^\intercal\otimes[\mathbb{I}_d:0]^\intercal\right)\right)
\end{aligned}$$

is a $(d+f)^2\times(d+f^2+df)$ matrix. Moreover, since $\bar{\Sigma}^* = \frac{1}{T}\int_0^T\bar{\Sigma}_s ds+O_p(n^{-1/2})$ still holds, it follows by direct calculations that $\beta^* = \beta+O_p(n^{-1/2})$, $\Sigma^{f*} = \frac{1}{T}\int_0^T\Sigma_s^f ds+O_p(n^{-1/2})$, and $\Sigma^{e*} = \frac{1}{T}\int_0^T\Sigma_s^e ds+O_p(n^{-1/2})$, which are the solutions to $\bar{\Psi}_n^* = 0$, so that the consistency holds.

Next, similar to (A.17), we have

$$\begin{aligned}
\bar{V}_n :=&\ \mathrm{D}_{\bar{\Sigma}}(\bar{\Psi}_n^*) \\
=&\ \frac{\Delta^2}{2}(\mathbb{I}_{d+f}\otimes\bar{G}^\intercal)\left(\bar{\Omega}^{-1}\otimes\bar{\Omega}^{-1}+(\bar{\Omega}^{-1}\left(\bar{\Omega}^\star-\bar{\Omega}\right)\bar{\Omega}^{-1}\right) \\
&\times\ \otimes\bar{\Omega}^{-1}+\bar{\Omega}^{-1}\otimes(\bar{\Omega}^{-1}\left(\bar{\Omega}^\star-\bar{\Omega}\right)\bar{\Omega}^{-1}))(\mathbb{I}_{d+f}\otimes\bar{G}),
\end{aligned}$$

where $\bar{V}_n$ is a $(d+f)^2\times(d+f)^2$ matrix. Also, using the expressions of $\mathbb{C}$ and chain rule, we have

$$\begin{aligned}
&\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{e*}),\mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{e*}),\mathrm{D}_\beta(\bar{\Psi}_n^{e*})\right) \\
&= \left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma})\right)^\intercal\bar{V}_n\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma}),\mathrm{D}_{\Sigma^f}(\bar{\Sigma}),\mathrm{D}_\beta(\bar{\Sigma})\right) \\
&\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{f*}),\mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{f*}),\mathrm{D}_\beta(\bar{\Psi}_n^{f*})\right) \\
&= \left(\mathrm{D}_{\Sigma^f}(\bar{\Sigma})\right)^\intercal\bar{V}_n\left\{\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma}),\mathrm{D}_{\Sigma^f}(\bar{\Sigma}),\mathrm{D}_\beta(\bar{\Sigma})\right)\right. \\
&\left.+\left(0,0,\left(\mathbb{I}_{f^2}\otimes\bar{\Psi}_n^*\right)\mathrm{D}_\beta\left(\mathrm{D}_{\Sigma^f}(\bar{\Sigma})^\intercal\right)\right)\right\} \\
&\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{\beta*}),\mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{\beta*}),\mathrm{D}_\beta(\bar{\Psi}_n^{\beta*})\right) \\
&= \left(\mathrm{D}_\beta(\bar{\Sigma})\right)^\intercal\bar{V}_n\left\{\left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Sigma}),\mathrm{D}_{\Sigma^f}(\bar{\Sigma}),\mathrm{D}_\beta(\bar{\Sigma})\right)\right. \\
&\left.+\left(0,\left(\mathbb{I}_{df}\otimes\bar{\Psi}_n^*\right)\mathrm{D}_{\Sigma^f}\left(\mathrm{D}_\beta(\bar{\Sigma})^\intercal\right),\left(\mathbb{I}_{f^2}\otimes\bar{\Psi}_n^*\right)\mathrm{D}_\beta\left(\mathrm{D}_\beta(\bar{\Sigma})^\intercal\right)\right)\right\}.
\end{aligned}$$

Since $\bar{\Psi}_n^*\big|_{\bar{\Sigma}=\bar{\Sigma}^\star,\bar{\Lambda}=\bar{\Lambda}^\star} = o_p(1)$, using the same derivation as for (A.18), we have

$$\begin{aligned}
&\left.\begin{pmatrix}
\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{e*}) & \mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{e*}) & \mathrm{D}_\beta(\bar{\Psi}_n^{e*}) \\
\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{f*}) & \mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{f*}) & \mathrm{D}_\beta(\bar{\Psi}_n^{f*}) \\
\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\bar{\Psi}_n^{\beta*}) & \mathrm{D}_{\Sigma^f}(\bar{\Psi}_n^{\beta*}) & \mathrm{D}_\beta(\bar{\Psi}_n^{\beta*})
\end{pmatrix}^{-1}\right|_{\bar{\Sigma}=\bar{\Sigma}^\star,\bar{\Lambda}=\bar{\Lambda}^\star} \\
&= \left(\mathbb{C}^{\star\intercal}(H(\bar{\Sigma}^\star,\bar{\Lambda}^\star))^{-1}\mathbb{C}^\star\right)^{-1}+o_p(1),
\end{aligned}$$

where $\mathbb{C}^\star = \mathbb{C}|_{\bar{\Sigma}=\bar{\Sigma}^\star,\bar{\Lambda}=\bar{\Lambda}^\star}$. By element-wise mean-value theorem, we have

$$n^{1/4}\begin{pmatrix}
\mathrm{diag}\left(\widehat{\Sigma}^e\right)-\mathrm{diag}\left(\frac{1}{T}\int_0^T\Sigma_s^e ds\right) \\
\mathrm{vec}\left(\widehat{\Sigma}^f\right)-\mathrm{vec}\left(\frac{1}{T}\int_0^T\Sigma_s^f ds\right) \\
\mathrm{vec}\left(\widehat{\beta}\right)-\mathrm{vec}(\beta)
\end{pmatrix}\xrightarrow{\mathcal{L}_X}\mathcal{MN}\left(0,G^\star\right),$$

where

$$G^{\star} = \left(\mathbb{C}^{\star\intercal} H(\bar{\Sigma}^{\star}, \bar{\Lambda}^{\star})^{-1} \mathbb{C}^{\star}\right)^{-1} \mathbb{C}^{\star\intercal}\left(V^1\left(\bar{\Sigma}^{\star}, \bar{\Lambda}^{\star}\right) + V^2\left(\bar{\Sigma}^{\star}, \bar{\Lambda}^{\star}\right) + V^3\left(\bar{\Sigma}^{\star}, \bar{\Lambda}^{\star}\right)\right) \mathbb{C}^{\star}\left(\mathbb{C}^{\star\intercal} H(\bar{\Sigma}^{\star}, \bar{\Lambda}^{\star})^{-1} \mathbb{C}^{\star}\right)^{-1}.$$

It then implies that

$$\Pi^e = [\mathbb{I}_d, 0, 0] G^{\star} [\mathbb{I}_d, 0, 0]^{\intercal},$$
$$\Pi^f = \mathbb{D}_f^{+}[0, \mathbb{I}_{f^2}, 0] G^{\star} [0, \mathbb{I}_{f^2}, 0]^{\intercal} \mathbb{D}_f^{+\intercal},$$
$$\Pi^{\beta} = [0, 0, \mathbb{I}_{df}] G^{\star} [0, 0, \mathbb{I}_{df}]^{\intercal}. \tag{A.26}$$

Since $\widehat{\Sigma} = \widehat{\beta}\widehat{\Sigma}^f\widehat{\beta}^{\intercal} + \widehat{\Sigma}^e$, it follows that

$$W^{\star} := \left(\mathrm{D}_{\mathrm{diag}(\Sigma^e)}(\Sigma), \mathrm{D}_{\Sigma^f}(\Sigma), \mathrm{D}_{\beta}(\Sigma)\right)\Big|_{\bar{\Sigma}=\bar{\Sigma}^{\star}} = \left(\mathbb{E}_d, \beta^{\star} \otimes \beta^{\star}, (\mathbb{I}_{d^2} + \mathbb{K}_{d,d})(\beta^{\star}\Sigma^{f\star} \otimes \mathbb{I}_d)\right)$$

is a $d^2 \times (d + f^2 + df)$ matrix, which by the delta method, leads to

$$n^{1/4}\left(\mathrm{vech}\left(\widehat{\Sigma}\right) - \mathrm{vech}\left(\frac{1}{T}\int_0^T \Sigma_s ds\right)\right) \xrightarrow{\mathcal{L}_X} \mathcal{MN}\left(0, \Pi^F\right),$$

where, using (A.1) and Equation (4) of Section 3.8 in Magnus and Neudecker (1999),

$$\Pi^F = \mathbb{D}_d^{+} W^{\star} G^{\star} W^{\star\intercal} \mathbb{D}_d^{+\intercal}. \quad \square \tag{A.27}$$

Appendix 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 $v_i = x_i - Z_i\widehat{y}_{t_i}$, $F_i = Z_i\,(P_i + \Lambda)\,Z_i^{\intercal}$, $K_i = P_i Z_i^{\intercal} F_i^{-1}$, $L_i = I - K_i Z_i$ then $\widehat{y}_{t_{i+1}} = \widehat{y}_{t_i} + K_i v_i$, $P_{i+1} = P_i L_i^{\intercal} + \Delta_{i+1}^n \Sigma$. Here $\widehat{y}_{t_i} = \mathrm{E}(y_{t_i}|x_{i:i-1})$ and $F_i = \mathrm{Cov}(x_i|x_{i:i-1})$. These recursions need some initial conditions $\widehat{y}_{t_1}$ and P_1. Throughout we will assume their choice does not depend upon Σ or Λ. A typical selection for $\widehat{y}_{t_1}$ 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 $d \times d_i$, $\widehat{y}_{t_{i+1}|t_i}$ 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 $\log f(x_{1:n}; \Lambda, \Sigma) = c - \frac{1}{2}\sum_{i=1}^n \log|F_i| - \frac{1}{2}\sum_{i=1}^n v_i^{\intercal} F_i^{-1} v_i$.

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

Computing $\widehat{\bar{e}}_{i|n}$, $\widehat{\bar{u}}_{i|n}$, $\bar{D}_{i|n}$ and $\bar{N}_{i|n}$ uses the same formulas, with all notations replaced by those for stacked variables.

References

Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High frequency covariance estimates with noisy and asynchronous financial data. J. Amer. Statist. Assoc. 105, 1504–1517.

Aït-Sahalia, Y., Jacod, J., Li, J., 2012. Testing for jumps in noisy high frequency data. J. Econometrics 168, 207–222.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18, 351–416.

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

Aït-Sahalia, Y., Xiu, D., 2016a. Increased correlation among asset classes: Are volatility or jumps to blame?. J. Econometrics 194, 205–219.

Aït-Sahalia, Y., Xiu, D., 2016b. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. J. Econometrics (forthcoming).

Altmeyer, R., Bibinger, M., 2015. Functional stable limit theorems for quasi-efficient spectral covolatility estimators. Stochastic Process. Appl. 125, 4556–4600.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a. The distribution of realized stock return volatility. J. Financ. Econom. 61, 43–76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001b. The distribution of exchange rate volatility. J. Amer. Statist. Assoc. 96, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Anderson, T.W., Amemiya, Y., 1988. The asymptotic normal distribution of estimators in factor analysis under general conditions. Ann. Statist. 16 (2), 759–771.

Ansley, C.F., Kohn, R., 1983. Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. Biometrika 70 (1), 275–278.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Ann. Statist. 40 (1), 436–465.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Rev. Econom. Stud. 75, 339–369.

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

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics 162, 149–169.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 253–280.

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

Bibinger, M., 2012. An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory. Stochastic Process. Appl. 122, 2411–2453.

Bibinger, M., Hautsch, N., Malec, P., Reiß, M., 2014. Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. Ann. Statist. 42 (4), 1312–1346.

Brownlees, C.T., Gallo, G.M., 2010. Comparison of volatility measures: a risk management perspective. J. Financ. Econom. 8, 29–56.

Cartea, A., Karyampas, D., 2011. Volatility and covariation of financial assets: A high-frequency analysis. J. Banking Financ. 35, 3319–3334.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica 51, 1305–1324.

Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. J. Econometrics 159, 116–133.

Christensen, K., Podolskij, M., Vetter, M., 2013. On covariation estimation for multivariate continuous ito semimartingales with noise in non-synchronous observation schemes. J. Multivariate Anal. 120, 59–84.

Corsi, F., Peluso, S., Audrino, F., 2015. Missing asynchronicity: a Kalman-EM approach to multivariate realized covariance estimation. J. Appl. Econometrics 30 (3), 377–397.

de Jong, P., 1987. A central limit theorem for generalized quadratic forms. Probab. Theory Related Fields 75, 261–277.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. Rev. Financ. Stud. 22, 1915–1953.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Engle, R.F., 2009. Anticipating Correlations. Princeton University Press.

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

Engle, R.F., Russell, J.R., 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. J. Amer. Statist. Assoc. 74, 291–296.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econom. 33, 3–56.

Fan, J., Furger, A., Xiu, D., 2016. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. J. Bus. Econom. Statist. 34, 489–503.

Gloter, A., Jacod, J., 2001a. Diffusions with measurement errors. I — local asymptotic normality. ESAIM Probab. Stat. 5, 225–242.

Gloter, A., Jacod, J., 2001b. Diffusions with measurement errors. II — measurement errors. ESAIM Probab. Stat. 5, 243–260.

Hansen, P.R., Horel, G., 2009. Quadratic Variation by Markov Chains. Department of Economics, Stanford University, (Unpublished).

Hansen, P.R., Huang, Z., Shek, H.H., 2011. Realized GARCH: a joint model for returns and realized measures of volatility. J. Appl. Econometrics 27, 877–906.

Hansen, P.R., Large, J., Lunde, A., 2008. Moving average-based estimators of integrated variance. Econometric Rev. 27, 79–111.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise (with discussion). J. Bus. Econom. Statist. 24, 127–218.

Harvey, C.R., Liu, Y., Zhu, H., 2016. . . . and the cross-section of expected returns. Rev. Financ. Stud. 29 (1), 5–68.

Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11, 359–379.

Jacod, J., 2012. Statistics and high frequency data. In: Kessler, M., Lindner, A., Sorensen, M. (Eds.), Statistical Methods for Stochastic Differential Equations. Chapman and Hall, pp. 191–310.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.

Jacod, J., Podolskij, M., Vetter, M., 2010. Limit theorems for moving averages of discretized processes plus noise. Ann. Statist. 38, 1478–1545.

Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of correlated measurement error. J. Econometrics 147, 47–59.

Koike, Y., 2016. Estimation of integrated covariances in the simultaneous presence of nonsynchronicity, microstructure noise and jumps. Econometric Theory 32 (3), 533–611.

Kunitomo, N., Sato, S., 2013. Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise. N. Am. J. Econom. Financ. 26, 282–309.

Li, Y., Mykland, P., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. Econometric Theory 30, 580–605.

Li, J., Xiu, D., 2016. Generalized method of integrated moments for high-frequency data. Econometrica 84, 1613–1633.

Liu, C., Tang, C.Y., 2014. A quasi-maximum likelihood approach to covariance matrix with high frequency data. J. Econometrics 180 (2), 217–232.

Lunde, A., Shephard, N., Sheppard, K.K., 2012. Econometric analysis of vast covariance matrices using composite realized kernels. Department of Economics, University of Oxford, (Unpublished).

Magnus, J.R., Neudecker, H., 1979. The commutation matrix: Some properties and applications. Ann. Statist. 7 (2), 381–394.

Magnus, J.R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Economics. Wiley.

Malliavin, P., Mancino, M.E., 2002. Fourier series method for measurement of multivariate volatilities. Finance Stoch. 6, 49–61.

Malliavin, P., Mancino, M.E., 2009. A fourier transform method for nonparametric estimation of multivariate volatility. Ann. Statist. 37, 1983–2010.

Mancino, M.E., Sanfelici, S., 2008. Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. Comput. Statist. Data Anal. 52, 2966–2989.

Mancino, M.E., Sanfelici, S., 2011. Covariance estimation and dynamic asset allocation under microstructure effects via Fourier methodology. J. Financ. Econometrics 9, 367–408.

McAleer, M., Medeiros, M.C., 2008. Realized volatility: a review. Econometric Rev. 27, 10–45.

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

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

Park, S., Hong, S.Y., Linton, O., 2016. Estimating the quadratic covariation matrix for an asynchronously observed continuous time signal masked by additive noise. J. Econometrics 19, 325–347.

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

Peluso, S., Corsi, F., Mira, A., 2012. A Bayesian High-Frequency Estimator of the Multivariate Covariance of Noisy and Asynchronous Returns. University of Lugano, Switzerland, (Unpublished).

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

Reiß, M., 2011. Asymptotic equivalence for inference on the volatility from noisy observations. Ann. Statist. 39, 772–802.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. J. Econom. Theory 13, 341–360.

Sanfelici, S., Mancino, M.E., 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., Sheppard, K.K., 2010. Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. J. Appl. Econometrics 25, 197–231.

Stein, M.L., 1987. Minimum norm quadratic estimation of spatial variograms. J. Amer. Statist. Assoc. 82, 765–772.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.

Tse, Y.K., Tsui, K.C., 2002. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. J. Bus. Econom. Statist. 20, 351–362.

Varneskov, R.T., 2016. Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. J. Bus. Econom. Statist. 34 (1), 1–22.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

Wu, C.F.J., 1983. On the convergence properties of the EM algorithm. Ann. Statist. 11, 95–103.

Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.

Zhang, L., 2011. Estimating covariation: Epps effect and microstructure noise. J. Econometrics 160, 33–47.

Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.

Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. J. Bus. Econom. Statist. 14, 45–52.

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