Journal of Econometrics 208 (2019) 395–417

Contents lists available at ScienceDirect

Journal of Econometrics

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

Factor GARCH-Itô models for high-frequency data with application to large volatility matrix prediction

Donggyu Kim [a], Jianqing Fan [b,*]

[a] College of Business, Korea Advanced Institute of Science and Technology, Seoul, Korea
[b] Department of Operations Research and Financial Engineering, Princeton University, USA

ARTICLE INFO

Article history:
Received 17 May 2017
Received in revised form 25 June 2018
Accepted 9 October 2018
Available online 5 November 2018

JEL classification:
C13
C32
C53
C55
C58

Keywords:
Factor model
GARCH
Low-rank
POET
Quasi-maximum likelihood estimator
Sparsity

ABSTRACT

Several novel large volatility matrix estimation methods have been developed based on the high-frequency financial data. They often employ the approximate factor model that leads to a low-rank plus sparse structure for the integrated volatility matrix and facilitates estimation of large volatility matrices. However, for predicting future volatility matrices, these nonparametric estimators do not have a dynamic structure to implement. In this paper, we introduce a novel Itô diffusion process based on the approximate factor models and call it a factor GARCH-Itô model. We then investigate its properties and propose a quasi-maximum likelihood estimation method for the parameter of the factor GARCH-Itô model. We also apply it to estimating conditional expected large volatility matrices and establish their asymptotic properties. Simulation studies are conducted to validate the finite sample performance of the proposed estimation methods. The proposed method is also illustrated by using data from the constituents of the S&P 500 index and an application to constructing the minimum variance portfolio with gross exposure constraints.

1. Introduction

Volatility analysis for high-frequency financial data is a vibrant research area in financial econometrics and statistics. The high-frequency financial data allow us to study market microstructures and to estimate volatilities using the relatively short time horizon. Examples include two-time scale realized volatility (TSRV) (Zhang et al., 2005), multi-scale realized volatility (MSRV) (Zhang, 2006, 2011), pre-averaging realized volatility (PRV) (Christensen et al., 2010; Jacod et al., 2009), kernel realized volatility (KRV) (Barndorff-Nielsen et al., 2008, 2011), a quasi-maximum likelihood estimator (QMLE) (Aït-Sahalia et al., 2010; Xiu, 2010), local method of moments (Bibinger et al., 2014), and robust pre-averaging realized volatility (RPRV) (Fan and Kim, 2018). For the finite number of assets, these estimators perform well. However, in financial practices and studies, we often encounter a large number of assets, and it is known that to obtain the efficient and effective estimator for a large volatility matrix, we need to impose some sparse or factor structure on large volatility matrices. For example, several estimation methods for factor-based high-dimensional Itô processes have been proposed (Fan et al., 2016a; Aït-Sahalia and Xiu, 2017; Fan and Kim, 2018; Kim et al., 2018). They assume that the dependence of stock returns is driven by a few common factors, which leads to a low-rank plus sparse structure for integrated volatility matrix.

* Corresponding author.
E-mail addresses: donggyukim@kaist.ac.kr (D. Kim), jqfan@princeton.edu (J. Fan).

https://doi.org/10.1016/j.jeconom.2018.10.003
0304-4076/

In practice, we often need to predict the future volatility given the current information, and unless the volatility is stable, the nonparametric volatility matrix estimator does not capture the dynamics of the future volatility. One of the stylized features of returns is that the log-returns are not significantly autocorrelated while the squared log-returns are positively autocorrelated. Also we often observe that large changes of returns tend to be followed by large changes, and small changes of returns tend to be followed by small changes, which is the so-called volatility clustering (Mandelbrot, 1963). These stylized features indicate that the volatilities are heterogeneous and autocorrelated, which prompts us to develop parametric models to account for the dynamics of time-dependent volatilities in the stock market.

In this paper, we develop a parametric Itô diffusion model based on high-dimensional factor-based Itô processes whose volatility matrices consist of the factor (low-rank) volatility matrix and idiosyncratic (sparse) volatility matrix. Specifically, latent factor loading matrices are assumed to be relatively stable: namely, the eigenvectors of the factor volatility matrix used to construct estimated latent factors do not vary over a short time period (e.g. within a day). On the other hand, we allow the eigenvalues to evolve with time and impose a parametric dynamic structure. In particular, we assume that the eigenvalue sequence of the latent factor volatility matrices admits some unified GARCH-Itô model structure (Kim and Wang, 2016b) so that the dynamics of the volatility can be explained by the r-factors. Their daily integrated conditional volatility given the information up to that time point is a function of the weighted latent factor daily returns and previous conditional expectations. Thus, it has a structure that is similar to the famous GARCH (Bollerslev, 1986).

To estimate the model parameters of the factor GARCH-Itô process, we first construct eigenvalue estimators for the latent factor volatility matrices by a nonparametric method using high-frequency intra-daily data. Then using the relationship between the eigenvalue estimator and integrated eigenvalue calculated from the factor GARCH-Itô model, we propose a quasi-maximum likelihood estimation procedure and establish its asymptotic properties for estimated parameters. Finally, with the quasi-maximum likelihood estimator, we construct the conditional expected large volatility matrix estimator. Further improvements are also possible when the idiosyncratic volatility matrices are a martingale sequence and the eigenvectors of the latent factor volatility matrices are constant. See Section 4 for details.

The rest of the paper is organized as follows. Section 2 introduces a factor GARCH-Itô model based on the high-dimensional factor-based Itô diffusion process and studies its properties. Section 3 proposes a quasi-maximum likelihood estimation procedure and establishes its asymptotic results. In Section 4, we show how to estimate the conditional expected large volatility matrix and establish its asymptotic properties. Section 5 conducts Monte Carlo simulation to check the finite sample performance and applies the real data to the factor GARCH-Itô model. Proofs are collected in Section 6.

2. Factor GARCH-Itô model

We first introduce some notations. For any given $d_1 \times d_2$ matrix $\mathbf{U} = \left(U_{ij}\right)$, denote its Frobenius norm by $\|\mathbf{U}\|_F = \sqrt{\mathrm{tr}(\mathbf{U}^\top \mathbf{U})}$, its matrix spectral norm $\|\mathbf{U}\|_2$, and $\|\mathbf{U}\|_{\max} = \max_{i,j} |U_{ij}|$. For any given vector $\mathbf{a}$, $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix using elements of $\mathbf{a}$. Finally, for any given squared matrix $\mathbf{A}$, $\det(\mathbf{A})$ is the determinant of the matrix $\mathbf{A}$.

Denote by $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))^\top$ the vector of true log-stock prices at time t. To account for dependence, we assume that the true log-stock prices follow the factor model:

$$d\mathbf{X}(t) = \boldsymbol{\mu}(t)\,dt + \mathbf{B}(t)d\mathbf{f}(t) + d\mathbf{u}(t), \tag{2.1}$$

where $\boldsymbol{\mu}(t) \in \mathbb{R}^p$ is the drift vector, $\mathbf{B}(t) \in \mathbb{R}^{p\times r}$ the unknown loading matrix, $\mathbf{f}(t) \in \mathbb{R}^r$ the unobservable factor process, and $\mathbf{u}(t)$ the idiosyncratic process. Suppose that the latent factor and idiosyncratic processes $\mathbf{f}(t)$ and $\mathbf{u}(t)$ follow continuous-time diffusion models:

$$d\mathbf{f}(t) = \boldsymbol{\vartheta}^\top(t)d\mathbf{W}(t) \quad \text{and} \quad d\mathbf{u}(t) = \boldsymbol{\sigma}^\top(t)d\mathbf{W}^*(t),$$

where $\boldsymbol{\sigma}(t)$ is a p by p matrix, $\boldsymbol{\vartheta}(t)$ an r by r matrix, $\mathbf{W}(t)$ and $\mathbf{W}^*(t)$ are r-dimensional and p-dimensional independent Brownian motions, respectively. Stochastic processes $\boldsymbol{\mu}(t)$, $\mathbf{X}(t)$, $\mathbf{f}(t)$, $\mathbf{u}(t)$, $\mathbf{B}(t)$, $\boldsymbol{\sigma}(t)$, and $\boldsymbol{\vartheta}(t)$ are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0,\infty)\}, P)$ with filtration $\mathcal{F}_t$ satisfying the usual conditions. It is helpful to think that the time unit in our applications is day and we have high-frequency intra-daily data for the assets. The instantaneous volatility of $\mathbf{X}(t)$ is

$$\boldsymbol{\varsigma}(t) = (\varsigma_{ij}(t)) = \mathbf{B}(t)\boldsymbol{\vartheta}^\top(t)\boldsymbol{\vartheta}(t)\mathbf{B}(t)^\top + \boldsymbol{\sigma}^\top(t)\boldsymbol{\sigma}(t), \tag{2.2}$$

and the integrated volatility over time $[d-1, d]$ (e.g., intra-day d)

$$\boldsymbol{\Gamma}_d = \int_{d-1}^{d} \boldsymbol{\varsigma}(t)dt = \boldsymbol{\Psi}_d + \boldsymbol{\Sigma}_d,$$

where $\boldsymbol{\Psi}_d = \int_{d-1}^{d} \mathbf{B}(t)\boldsymbol{\vartheta}^\top(t)\boldsymbol{\vartheta}(t)\mathbf{B}(t)^\top dt$ and $\boldsymbol{\Sigma}_d = \int_{d-1}^{d} \boldsymbol{\sigma}^\top(t)\boldsymbol{\sigma}(t)dt$.

The idiosyncratic volatility matrices $\boldsymbol{\Sigma}_d$ usually come from the risk of the individual firm, which is generally unpredictable, but we can reduce its risk through diversification in portfolio allocations. On the other hand, the factor volatility matrices $\boldsymbol{\Psi}_d$ are governed by the factor-driven component of returns which are related to market factors such as industry sectors, inflation reports, Fed rate hikes, and oil prices, and it is impossible to completely avoid their risk in portfolio allocations (unless factor-neutral constraints are imposed). Thus, it is important to develop a parametric model to explain their dynamics in the stock market. In light of this, we propose a factor GARCH-Itô process to model the latent factor process as follows.

Let $\mathbf{q}_{d,1}, \ldots, \mathbf{q}_{d,r}$ be the eigenvectors of the factor volatility matrix $\boldsymbol{\Psi}_d$. We will assume that they are $\mathcal{F}_{d-1}$-measurable eigenvectors. This essentially requires that the factor loading matrix $\mathbf{B}(t)$ for $t \in [d-1, d)$ is $\mathcal{F}_{d-1}$-adaptive (predictable on day $d-1$). However, their corresponding instantaneous eigenvalues, $\lambda_{t,1}(\boldsymbol{\theta}_1), \ldots, \lambda_{t,r}(\boldsymbol{\theta}_r)$, have some specific parametric structure defined in Definition 1.

Definition 1. We call a log-stock price vector $\mathbf{X}(t)$, $t \in [0, \infty)$, to follow a factor GARCH-Itô model if it satisfies for any $i = 1, \ldots, r$ and $t \in [d-1, d)$,

$$d\mathbf{q}_{d,i}^{\top}\mathbf{X}(t) = \sqrt{p}\mu_i dt + \sqrt{\lambda_{t,i}(\boldsymbol{\theta}_i)}dW_i(t) + \mathbf{q}_{d,i}^{\top}\boldsymbol{\sigma}^{\top}(t)d\mathbf{W}^*(t),$$

$$\lambda_{t,i}(\boldsymbol{\theta}_i) = \lambda_{[t],i}(\boldsymbol{\theta}_i) + (t-[t])\{p\omega_i + (\gamma_i - 1)\lambda_{[t],i}(\boldsymbol{\theta}_i)\} + \sum_{l=1}^{r} \beta_{i,l}\left\{\int_{[t]}^{t}\sqrt{\lambda_{s,l}(\boldsymbol{\theta}_l)}dW_l(s)\right\}^2,$$

where $[t]$ denotes the integer part of t except that $[t] = t-1$ when t is an integer, the drift $\mu_i = p^{-1/2}\mathbf{q}_{d,i}^{\top}\boldsymbol{\mu}(t)$ is restricted to a non-negative value, and $\boldsymbol{\theta}_i = (\omega_i, \gamma_i, \beta_{i,1}, \ldots, \beta_{i,r})$ is the model parameter.

Remark 1. Since the eigenvector $\mathbf{q}_{d,i}$ has the sign problem, that is, $\mathbf{q}_{d,i}$ and $-\mathbf{q}_{d,i}$ are not distinguishable, we restrict the sign of the drift term μ_i in order to identify the model uniquely. Also, to focus on developing the volatility process, we assume that the drift term μ_i is constant over time.

Note that in the above definition, we assume that the weights (eigenvectors) used to construct latent factors do not change within a day or more generally a short time period. Yet, the instantaneous eigenvalues can evolve with the time. For example, the instantaneous eigenvalues are continuous time processes, and when restricted to the integer time $d \in \mathbb{N}$, they have the following GARCH structure

$$\lambda_{d,i}(\boldsymbol{\theta}_i) = p\omega_i + \gamma_i\lambda_{d-1,i}(\boldsymbol{\theta}_i) + \sum_{l=1}^{r}\beta_{i,l}Z_{d,l}^2,$$

where $Z_{d,l} = \int_{d-1}^{d}\sqrt{\lambda_{t,l}(\boldsymbol{\theta}_l)}dW_l(t)$. This has a similar structure to the GARCH model. For intra-daily volatility, eigenvalues in Definition 1 are also a form of GARCH model, except some kind of interpolation is used.

The model assumption and structure have some connections with pre-existing factor models and unified GARCH models (Kim and Wang, 2016b). For example, to identify the latent factor volatility matrix, we often assume that the factor loading matrix $\mathbf{B}(t)$ in model (2.1) is piecewise constant within a day and orthogonal (Aït-Sahalia and Xiu, 2017; Fan et al., 2013, 2016). Then the eigenvectors of the factor volatility matrix $\boldsymbol{\Psi}_d$ become $\mathbf{q}_{d,k} = \mathbf{b}_{[t],k}/\|\mathbf{b}_{[t],k}\|_2$, $k = 1, \ldots, r$, where $\mathbf{B}(t) = (\mathbf{b}_{[t],1}, \ldots, \mathbf{b}_{[t],r})$ and $t \in [d-1, d)$, and the instantaneous volatility matrix of the latent factor process $\boldsymbol{\vartheta}^{\top}(t)\boldsymbol{\vartheta}(t) = \mathrm{diag}(\lambda_{t,1}(\boldsymbol{\theta}_1)/\|\mathbf{b}_{[t],1}\|_2^2, \ldots, \lambda_{t,r}(\boldsymbol{\theta}_r)/\|\mathbf{b}_{[t],r}\|_2^2)$. Thus, the factor GARCH-Itô model assumes that the latent factor process $\mathbf{f}(t)$ consists of r diffusion processes and they follow some unified GARCH-Itô process (Kim and Wang, 2016b). That is, the factor GARCH-Itô model assumes that the dynamics of the stock prices are governed by the latent factor process $\mathbf{f}(t)$. This latent factor process $\mathbf{f}(t)$ is identifiable under the pervasive condition of the factor volatility matrices and sparse condition of the idiosyncratic volatility matrices (Fan et al., 2013, 2016; Aït-Sahalia and Xiu, 2017). We will discuss more details in Section 3.

In the literature of the high-frequency volatility analysis, we often estimate the integrated volatility matrix $\boldsymbol{\Gamma}_d$ by using high-frequency intra-daily data (Aït-Sahalia et al., 2010; Barndorff-Nielsen et al., 2011; Bibinger et al., 2014; Christensen et al., 2010; Fan and Kim, 2018; Jacod et al., 2009; Zhang, 2006, 2011). Using a nonparametric estimator, we are able to estimate integrated eigenvalues and infer the parameters in their GARCH structure. This facilitates volatility prediction. In the following proposition, we establish some properties of the integrated eigenvalues, which will be used for statistical inferences.

Proposition 2.1. *Let $\boldsymbol{\beta} = (\beta_{i,j})_{1\le i,j\le r}$, with coefficients $\beta_{i,j}$ defined in Definition 1. Then we have the following iterative relations for a factor GARCH-Itô model.*

(a) For $\|\boldsymbol{\beta}\|_2 < 1$ and $\det(\boldsymbol{\beta}) \neq 0$, we have for $d \in \mathbb{N}$

$$\int_{d-1}^{d}\boldsymbol{\lambda}_t(\boldsymbol{\theta})dt = \mathbf{h}_d(\boldsymbol{\theta}) + \mathbf{D}_d \ \text{a.s.},$$

where $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = (\lambda_{t,1}(\boldsymbol{\theta}_1), \ldots, \lambda_{t,r}(\boldsymbol{\theta}_r))^{\top}$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_r)^{\top}$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)^{\top}$,

$$\mathbf{h}_d(\boldsymbol{\theta}) = p\boldsymbol{\beta}^{-2}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r - \boldsymbol{\beta}\right)\boldsymbol{\omega} + \boldsymbol{\varrho}\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta}), \tag{2.3}$$

$$\mathbf{D}_d = \sum_{k=0}^{\infty}\boldsymbol{\beta}^{k+1}\left(2\int_{d-1}^{d}\frac{(d-t)^{k+1}}{(k+1)!}\int_{d-1}^{t}\sqrt{\lambda_{s,i}(\boldsymbol{\theta}_i)}dW_i(s)\sqrt{\lambda_{t,i}(\boldsymbol{\theta}_i)}dW_i(t)\right)_{i=1,\ldots,r}^{\top},$$

$\boldsymbol{\varrho} = \boldsymbol{\beta}^{-2}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r - \boldsymbol{\beta}\right)(\mathrm{diag}(\boldsymbol{\gamma}) - \mathbf{I}_r) + \boldsymbol{\beta}^{-1}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r\right)$, $e^{\boldsymbol{\beta}} = \sum_{k=0}^{\infty}\boldsymbol{\beta}^k/k!$, and $\mathbf{I}_r$ is an r-dimensional identity matrix.

(b) For $\|\boldsymbol{\beta}\|_2 < 1$, $\det(\boldsymbol{\beta}) \neq 0$, *and* $d \in \mathbb{N}$,

$$\mathrm{E}\left[\int_{d-1}^{d} \boldsymbol{\lambda}_t(\boldsymbol{\theta})dt \middle| \mathcal{F}_{d-1}\right] = \mathbf{h}_d(\boldsymbol{\theta}) \text{ a.s.}$$

Proposition 2.1 indicates that the integrated eigenvalues $\int_{d-1}^{d} \boldsymbol{\lambda}_t(\boldsymbol{\theta})dt$ can be decomposed into $\mathbf{h}_d(\boldsymbol{\theta})$ and $\mathbf{D}_d$. Also, $\mathbf{h}_d(\boldsymbol{\theta})$ is adapted to the filtration $\mathcal{F}_{d-1}$, and $\mathbf{D}_d$ is a martingale difference. Thus, the conditional expectation of the integrated eigenvalues given the past information $\mathcal{F}_{d-1}$ is $\mathbf{h}_d(\boldsymbol{\theta})$ which shares some similarities of the GARCH structure. For example, if $\boldsymbol{\beta}$ is a diagonal matrix or γ_i's are the same, $\mathbf{h}_d(\boldsymbol{\theta})$ has the standard GARCH structure. That is, when $\boldsymbol{\beta} = \mathrm{diag}(\beta_1, \ldots, \beta_r)$ or $\mathrm{diag}(\boldsymbol{\gamma}) = \gamma \mathbf{I}_r$, we have

$$\begin{aligned}\mathrm{E}\left[\int_{d-1}^{d} \boldsymbol{\lambda}_t(\boldsymbol{\theta})dt \middle| \mathcal{F}_{d-1}\right] &= \mathbf{h}_d(\boldsymbol{\theta}) \\ &= \boldsymbol{\beta}^{-1}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r\right)\boldsymbol{\omega} + \mathrm{diag}(\boldsymbol{\gamma})\mathbf{h}_{d-1}(\boldsymbol{\theta}) + \varrho\boldsymbol{\beta}\mathbf{Z}_{d-1}^2 \text{ a.s.,}\end{aligned}$$

where $\mathbf{Z}_d^2 = (Z_{d,1}^2, \ldots, Z_{d,r}^2)^\top$.

Remark 2. Generalized dynamic factor models have been developed to estimate and forecast large volatility matrices (Boivin and Ng, 2006; Forni et al., 2000, 2015; Stock and Watson, 2002). See also Barigozzi and Hallin (2016), Barigozzi and Hallin (2017), Connor et al. (2006), Diebold and Nerlove (1989), Harvey et al. (1992), Ng et al. (1992), Rangel and Engle (2012), Sentana et al. (2008) and Van der Weide (2002). These models have some VAR or GARCH structure for the latent factor volatilities, which are usually developed based on the discrete time models. For example, for large panels of stock returns, Barigozzi and Hallin (2016) proposed the nonparametric and model-free two-step generalized dynamic factor model, and Barigozzi and Hallin (2017) studied how to extend the two-step generalized dynamic model to predicting future volatilities. As seen in Proposition 2.1, for the daily log-returns, the factor GARCH-Itô model also has some generalized dynamic factor model structure. Thus, the proposed factor GARCH-Itô model can be considered as a specific class of the generalized dynamic factor models. The main effort of this paper is to connect these well-developed discrete time models to the continuous Itô diffusion process. Thus, the main difference from the generalized dynamic factor models is that the factor GARCH-Itô model is developed under the continuous time diffusion process. Thanks to this connection, we can make inferences using the high-frequency financial data, which provides more accurate parameter estimators.

3. Parameter estimation for the factor GARCH-Itô model

3.1. A model set-up

Suppose that the true log-stock prices follow the factor GARCH-Itô model in Definition 1. We assume that the integrated volatility matrix $\boldsymbol{\Gamma}_d$ has the low-rank plus sparse structure (Aït-Sahalia and Xiu, 2017; Fan et al., 2016a). Specifically, the factor volatility matrices $\boldsymbol{\Psi}_d$'s have the finite rank r, and the idiosyncratic volatility matrices $\boldsymbol{\Sigma}_d = (\Sigma_{d,ij})_{i,j=1,\ldots,p}$, $d = 1, \ldots, n$, satisfy the sparse condition,

$$\max_{1\le d\le n}\max_{1\le i\le p}\sum_{j=1}^{p}|\Sigma_{d,ij}|^{\delta}(\Sigma_{d,ii}\Sigma_{d,jj})^{(1-\delta)/2} \le M_\sigma s_p, \tag{3.1}$$

where $\delta \in [0, 1)$, M_σ is a bounded positive random variable, and the sparsity measure s_p diverges slowly with the dimensionality p. When $\delta = 0$ and $\Sigma_{d,ii}$ is bounded from below, s_p measures the maximum number of nonvanishing elements in each row of the idiosyncratic volatility matrix $\boldsymbol{\Sigma}_d$.

Assume that we observe the true log-stock prices $\mathbf{X}(t)$, $t = 0, \ldots, n$, at the low frequency (e.g., daily). In addition, we have high-frequency intra-daily data that are contaminated by microstructural noise. To capture this stylized feature, we assume that the observed log-stock prices are masked with the additive noises:

$$Y_i(t_{d,k}) = X_i(t_{d,k}) + \epsilon_i(t_{d,k}), \quad i = 1, \ldots, p, d = 1, \ldots, n, k = 0, \ldots, m, \tag{3.2}$$

where $d - 1 = t_{d,0} < \cdots < t_{d,m} = d$, the microstructural noises are independent random variables which are independent of the price process and volatility process, and for each the ith asset and dth day, $\epsilon_i(t_{d,k})$, $k = 1, \ldots, m$, are i.i.d. with $\mathrm{E}\{\epsilon_i(t_{d,k})\} = 0$. Furthermore, we assume that the observation time points are synchronized and equally spaced, that is, $t_{d,k} - t_{d,k-1} = m^{-1}$ for $d = 1, \ldots, n$ and $k = 1, \ldots, m$.

Remark 3. In practice, the observed time points are non-synchronized and unequally spaced. This non-synchronization problem has been well studied in the literature by using refresh time (see also Aït-Sahalia et al. (2010), Barndorff-Nielsen et al. (2011), Bibinger et al. (2014), Christensen et al. (2010) and Zhang (2011)). Thus, to focus on development of the parametric model, we assume that the observed time points are synchronized and equally spaced for simplicity so that the key techniques can be better highlighted.

3.2. Nonparametric estimation methods

To develop a parametric estimation method for the factor GARCH-Itô models, we need a good nonparametric estimators for the eigenvalues and eigenvectors of the factor volatility matrix $\boldsymbol{\Psi}_d$. In this section, we first investigate asymptotic behaviors of the nonparametric estimators.

Let $\widehat{\boldsymbol{\Gamma}}_d$ be the dth day integrated volatility matrix estimator which can be one of multi-scale realized volatility matrix (Zhang, 2006), pre-averaging realized volatility matrix (Christensen et al., 2010), and kernel realized volatility matrix (Barndorff-Nielsen et al., 2011). By the eigen-decomposition, the integrated volatility matrix $\widehat{\boldsymbol{\Gamma}}_d$ can be decomposed as

$$\widehat{\boldsymbol{\Gamma}}_d = \sum_{i=1}^{p} \widehat{\xi}_{d,i} \widehat{\mathbf{q}}_{d,i} \widehat{\mathbf{q}}_{d,i}^{\top},$$

where the realized eigenvalue $\widehat{\xi}_{d,i}$ is the ith largest eigenvalue of $\widehat{\boldsymbol{\Gamma}}_d$ with $\widehat{\mathbf{q}}_{d,i}$ as its associated eigenvector. We use the first r eigenvalues and eigenvectors, $\widehat{\xi}_{d,i}$ and $\widehat{\mathbf{q}}_{d,i}$, $i = 1, \ldots, r$, as the estimators for the eigenvalues and eigenvectors of the factor volatility matrix $\boldsymbol{\Psi}_d$. To investigate their asymptotic behaviors, we need the following technical conditions. Denote by C's generic constants whose values are free of $\boldsymbol{\theta}$, n, m, and p and may change from occurrence to occurrence.

Assumption 1.

(a) The instantaneous volatility, drift processes, and microstructural noises satisfy, for some $\alpha \geq 2$,

$$\max_{1 \leq t \leq n} \max_{1 \leq i \leq p} \mathrm{E}\{\varsigma_{ii}(t)^{3\alpha}\} < \infty, \qquad \max_{1 \leq t \leq n} \mathrm{E}\{\|\boldsymbol{\varsigma}(t)\|_2^{3\alpha}\} \leq Cp^{3\alpha},$$
$$\max_{1 \leq t \leq n} \mathrm{E}\{\|\boldsymbol{\sigma}^{\top}(t)\boldsymbol{\sigma}(t)\|_2^{3\alpha}\} \leq Cs_p^{3\alpha}, \qquad \max_{1 \leq t \leq n} \max_{1 \leq i \leq p} \mathrm{E}\{|\mu_i(t)|^{6\alpha}\} < \infty,$$
$$\max_{1 \leq d \leq n} \max_{1 \leq i \leq p} \mathrm{E}\{|\epsilon_i(t_{d,k})|^{6\alpha}\} < \infty,$$

where $\boldsymbol{\varsigma}(t) = (\varsigma_{ij}(t))_{i,j=1,\ldots,p}$ is the instantaneous volatility process of the log-price $\mathbf{X}(t)$ defined in (2.2).

(b) Let $D_\xi = \min\{\xi_{d,i} - \xi_{d,i+1}, i = 1, \ldots, r, d = 1, \ldots, n\}$, where the integrated eigenvalue $\xi_{d,i} = \int_{d-1}^{d} \lambda_{t,i}(\boldsymbol{\theta}_i)dt$ and $\xi_{d,r+1} = 0$. There is a fixed positive constant C such that $D_\xi \geq Cp$ a.s.

Remark 4. Assumption 1(b) is called the pervasive condition which is often imposed on analyzing the approximate factor models (Aït-Sahalia and Xiu, 2017; Fan et al., 2016). Under the factor GARCH-Itô model, the integrated eigenvalue is a function of the model parameters $\boldsymbol{\theta}_i$'s, and there exist the parameters $\boldsymbol{\theta}_i$'s satisfying Assumption 1(b). For example, for any $i > i'$, we choose the parameters $\boldsymbol{\theta}_i$'s such that $\beta_{i,1} > \beta_{i',1}, \ldots, \beta_{i,r} > \beta_{i',r}, \gamma_i > \gamma_{i'}$, and $\omega_i > \omega_{i'}$. Then the eigen-gap D_ξ diverges with the order p.

The following theorem provides the convergence rates of $\widehat{\xi}_{d,i}$ and $\widehat{\mathbf{q}}_{d,i}$.

Theorem 3.1. *Suppose that the true log-stock prices follow the Factor GARCH-Itô model, and Assumption 1 and the sparsity condition (3.1) are met, and that for any $d = 1, \ldots, n$,*

$$\mathrm{E}\left(\left|\widehat{\Gamma}_{d,ij} - \Gamma_{d,ij}\right|^{3\alpha}\right) \leq Cm^{-3\alpha/4} \quad \text{for all } i, j \in \{1, \ldots, p\}. \tag{3.3}$$

Then we have, for any $d = 1, \ldots, n$,

$$\max_{1 \leq i \leq r} \mathrm{E}\left(|\widehat{\xi}_{d,i} - \xi_{d,i}|^{\alpha}\right) \leq Cp^{\alpha}\{m^{-\alpha/4} + (s_p/p)^{\alpha}\}, \tag{3.4}$$
$$\max_{1 \leq i \leq r} \mathrm{E}\left(\|\widehat{\mathbf{q}}_{d,i} - sign(\langle \widehat{\mathbf{q}}_{d,i}, \mathbf{q}_{d,i}\rangle)\mathbf{q}_{d,i}\|_2^{3\alpha}\right) \leq C\left\{m^{-3\alpha/4} + (s_p/p)^{3\alpha}\right\}. \tag{3.5}$$

Remark 5. If the input volatility matrix $\widehat{\boldsymbol{\Gamma}}_d$ satisfies the condition (3.3), we can enjoy the asymptotic properties obtained in Theorem 3.1. For example, multi-scale realized volatility (MSRV) (Zhang, 2006, 2011), pre-averaging realized volatility (PRV) (Christensen et al., 2010; Jacod et al., 2009), and kernel realized volatility (KRV) (Barndorff-Nielsen et al., 2008, 2011) satisfy the condition (3.3) under Assumption 1 (see Christensen et al. (2010) and Kim et al. (2016) and Tao et al. (2013a)).

Remark 6. Theorem 3.1 shows that the nonparametric eigen-factor estimators have convergence rates, consisting of two terms $m^{-1/4}$ and s_p/p. The first term comes from estimating the integrated volatility matrix $\boldsymbol{\Gamma}_d$, which is optimal in presence of the microstructural noise. The second term s_p/p appears because we cannot observe the latent factors $\mathbf{f}(t)$. It is the cost that we need to separate the factor volatility matrix $\boldsymbol{\Psi}_d$ and idiosyncratic volatility matrix $\boldsymbol{\Sigma}_d$ from the integrated volatility matrix $\boldsymbol{\Gamma}_d$. Thus, the sparsity condition such as $s_p/p = o(1)$ is required.

3.3. Quasi-maximum likelihood estimator

In this section, we propose a quasi-maximum likelihood estimator for the parameters of the factor GARCH-Itô model and establish its asymptotic convergence rate. We denote the true parameters by $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{0,1}, \dots, \boldsymbol{\theta}_{0,r})$ and $\boldsymbol{\theta}_{0,i} = (\omega_{0,i}, \gamma_{0,i}, \mu_{0,i}, \beta_{0,i,1}, \dots, \beta_{0,i,r})$.

To estimate the true parameters $\boldsymbol{\theta}_0$, we use the direct relationship between the integrated eigenvalue estimator $\widehat{\xi}_{d,i}$ and integrated eigenvalue calculated from the factor GARCH-Itô model. For example, $p^{-1}\widehat{\xi}_{d,i}$ converges to the integrated eigenvalue $p^{-1}\xi_{d,i} = p^{-1}\int_{d-1}^{d} \lambda_{t,i}(\boldsymbol{\theta}_0)dt$ with the convergence rate $m^{-1/4} + s_p/p$ (see Theorem 3.1). Let

$$\phi_{d,i}(\boldsymbol{\theta}) = \frac{p\omega_i + (\gamma_i+1)\lambda_{d-1,i}(\boldsymbol{\theta})}{2} + \sum_{l=1}^{r} \beta_{i,l}\,\kappa_{d,l}(\mu_l),$$

where

$$\lambda_{d,i}(\boldsymbol{\theta}) = p\omega_i + \gamma_i\lambda_{d-1,i}(\boldsymbol{\theta}) + \sum_{l=1}^{r}\beta_{i,l}\left(Z_{d,l} + \sqrt{p}\mu_{0,l} - \sqrt{p}\mu_l\right)^2,$$

$$\kappa_{d,i}(\mu_i) = \int_{d-1}^{d}\left[\mathbf{q}_{d,i}^{\top}\{\mathbf{X}(t) - \mathbf{X}(d-1)\} - \mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t) - \mathbf{u}(d-1)\} - (t-d+1)\sqrt{p}\mu_i\right]^2 dt,$$

and, under the unified GARCH-Itô model, the integrated eigenvalue $\xi_{d,i}$ is the same as $\phi_{d,i}(\boldsymbol{\theta}_0)$. Using this relationship, we can construct a quasi-likelihood function as follows:

$$L_{n,m,i}(\boldsymbol{\theta}) = -\frac{1}{2n}\sum_{d=1}^{n}\left\{\log\phi_{d,i}(\boldsymbol{\theta}) + \frac{\widehat{\xi}_{d,i}}{\phi_{d,i}(\boldsymbol{\theta})}\right\}.$$

To evaluate the quasi-likelihood function, we need a good nonparametric estimator for $\kappa_{d,i}(\mu_i)$. The difficulty of estimating these quantities is to identify the latent factor process. Under the pervasive condition (Assumption 1(b)) and the sparsity condition (3.1), we can separate the factor part from the log-price process $\mathbf{X}(t)$ (see Aït-Sahalia and Xiu (2017)). Since the microstructural noise is assumed to have a sparse correlation structure, using the similar argument, we can also separate the factor part from the noisy observation $\mathbf{Y}(t_{d,i})$. Using this technique, we can estimate $\kappa_{d,i}(\mu_i)$ nonparametrically by

$$\widehat{\kappa}_{d,i}(\mu_i) = \sum_{l=1}^{m}\{\widehat{\mathbf{q}}_{d,i}^{\top}(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)) - (t_{d,l} - d + 1)\sqrt{p}\mu_i\}^2\Delta_{d,l},$$

where $\Delta_{d,l} = t_{d,l} - t_{d,l-1}$. It can then be shown that $p^{-1}\widehat{\kappa}_{d,i}(\mu_i)$ uniformly converges to $p^{-1}\kappa_{d,i}(\mu_i)$ with the convergence rate $m^{-1/4} + (s_p/p)^{1/2}$ (see Lemma 6.1 in Section 6).

With the nonparametric estimator $\widehat{\kappa}_{d,i}$, we estimate the quasi-likelihood function by

$$\widehat{L}_{n,m,i}(\boldsymbol{\theta}) = -\frac{1}{2n}\sum_{d=1}^{n}\left\{\log\widehat{\phi}_{d,i}(\boldsymbol{\theta}) + \frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\right\},$$

where $\widehat{\lambda}_{d,i}(\boldsymbol{\theta})$ is the solution to (3.6),

$$\widehat{\lambda}_{d,i}(\boldsymbol{\theta}) = p\omega_i + \gamma_i\widehat{\lambda}_{d-1,i}(\boldsymbol{\theta}) + \sum_{l=1}^{r}\beta_{i,l}\left[\widehat{\mathbf{q}}_{d,l}^{\top}\{\mathbf{X}(d) - \mathbf{X}(d-1)\} - \sqrt{p}\mu_l\right]^2, \tag{3.6}$$

and

$$\widehat{\phi}_{d,i}(\boldsymbol{\theta}) = \frac{p\omega_i + (\gamma_i+1)\widehat{\lambda}_{d-1,i}(\boldsymbol{\theta})}{2} + \sum_{l=1}^{r}\beta_{i,l}\,\widehat{\kappa}_{d,l}(\mu_l).$$

Then the true parameters $\boldsymbol{\theta}_0$ are estimated by maximizing the quasi-likelihood function $\widehat{L}_{n,m}(\boldsymbol{\theta}) = \sum_{i=1}^{r}\widehat{L}_{n,m,i}(\boldsymbol{\theta})$, that is,

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\widehat{L}_{n,m}(\boldsymbol{\theta}), \tag{3.7}$$

where $\boldsymbol{\Theta}$ is the parameter space of $\boldsymbol{\theta}$. To investigate the asymptotic behavior of quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$, we make some technical conditions as follows.

Assumption 2.

(a) Assume that the parameter space $\boldsymbol{\Theta}$ for $\boldsymbol{\theta}$ is compact, $\|\boldsymbol{\beta}\|_2 < 1$, $\det(\boldsymbol{\beta}) \neq 0$, $0 < \gamma_i < 1$ for all $i = 1, \dots, r$, and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$, where $\boldsymbol{\beta} = (\beta_{i,j})_{i,j=1,\dots,r}$.
(b) The initial value $\boldsymbol{\lambda}_0$ in Definition 1 is given.

Remark 7. We impose the initial value condition (Assumption 2(b)) to investigate the asymptotic behaviors with the finite period. However, even if this condition is violated, the effect of the initial value is negligible with the convergence rate n^{-1} (see Lemma 1 in Kim and Wang (2016b)) and the convergence rate in Theorem 3.2 has one additional term n^{-1}.

The following theorem provides the convergence rate of the quasi-maximum likelihood estimator (QMLE) $\widehat{\boldsymbol{\theta}}$.

Theorem 3.2. *Under the assumptions of Theorem 3.1 and Assumption 2, for $n \geq 2+4r$, we have*

$$\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_{\max} = O_p\left(m^{-1/4}+(s_p/p)^{1/2}\right).$$

Remark 8. Theorem 3.2 shows that the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ has the convergence rate $m^{-1/4}+\sqrt{s_p/p}$. The first term is coming from the high-frequency observation in presence of the microstructural noise, and the second term is due to identifying the latent factors in the volatility matrix. If we can observe the latent factor part, we may have the optimal convergence rate $m^{-1/4}$. Or when the number, p, of assets is big, for example, $m^{1/2}s_p = O(p)$, we have the convergence rate $m^{-1/4}$.

Remark 9. Asymptotic distribution of the QMLE $\widehat{\boldsymbol{\theta}}$ would be useful for its inference. However, to establish the asymptotic distribution, we face two challenges. One is to identify the latent factor part $\mathbf{f}(t)$. If the number of assets is big enough to make the term $\sqrt{s_p/p}$ negligible, this issue will be solved and we may be able to derive the asymptotic distribution of the non-parametric eigenvalue estimator $\widehat{\xi}_{d,i}$ under some conditions. The other issue is to estimate the parametric integrated eigenvalue $\phi_{d,i}(\boldsymbol{\theta}_0)$. Since its asymptotic convergence rate is $m^{-1/4}+\sqrt{s_p/p}$ (see Lemma 6.1), it is not negligible and we need to derive its asymptotic distribution. The major challenge of deriving the asymptotic distribution of $\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)$ is to handle the eigenvector estimator $\widehat{\mathbf{q}}_{d,i}$. For example, we need to manage the terms $\{\widehat{\mathbf{q}}_{d,i}^{\top}(\mathbf{Y}(t_{d,l})-\mathbf{Y}(d-1))\}^2$, which is not easy. In short, if we know the asymptotic joint distribution of $\widehat{\xi}_{d,i}$ and $\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)$, $d=1,\ldots,n$, $i=1,\ldots,r$, we can possibly derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$. All these problems are very challenging and we leave this issue for a future study.

4. Large volatility matrix estimation

In financial practices such as portfolio allocation, it is important to predict future volatilities. In this section, we show how to predict the future large volatility matrix under the factor GARCH-Itô model.

Given the current information $\mathcal{F}_n$, the conditional expectation $\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)$ is the best predictor for the future volatility matrix. To evaluate the conditional expectation under the factor GARCH-Itô model, we need some additional structure for the idiosyncratic volatility matrix $\boldsymbol{\Sigma}_{n+1}$ and eigenvectors of the factor volatility matrix $\boldsymbol{\Psi}_{n+1}$. As discussed in Section 2, the idiosyncratic risk is unpredictable, but it can be mitigated by diversification. Hence it may not be harmful to assume that the idiosyncratic volatility matrices $\boldsymbol{\Sigma}_d$ are martingale processes: $\mathrm{E}\left(\boldsymbol{\Sigma}_{n+1}|\mathcal{F}_n\right)=\boldsymbol{\Sigma}_n$ a.s. Also, we assume that the eigenvectors $\mathbf{q}_{d,i}$'s are constant and $\mathbf{q}_{d,i}=\mathbf{q}_i$ for all $d=1,\ldots,n+1$. Then the conditional expected large volatility matrix $\boldsymbol{\Gamma}_{n+1}$ is

$$\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)=\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^{\top}+\boldsymbol{\Sigma}_n \text{ a.s.},$$

where $\mathbf{h}_n(\boldsymbol{\theta}_0)=(h_{n,1}(\boldsymbol{\theta}_0),\ldots,h_{n,r}(\boldsymbol{\theta}_0))^{\top}$ is defined in (2.3). We use the conditional expectation $\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)$ as the future volatility matrix estimator.

Remark 10. In this paper, we mainly focus on developing the parametric model of the factor volatility process. When analyzing the factor volatility matrix, we do not need any strong condition for the idiosyncratic volatility matrix. In fact, we only impose the sparse structure (3.1) on it. On the other hand, in order to predict the future volatilities, we need some parametric structure for the idiosyncratic volatility matrix. To focus on the development of the factor volatility process, we simply assume that the idiosyncratic volatility process is a martingale. However, several studies show that the idiosyncratic volatility exhibits heteroskedasticity and its proportion of the total volatility is significant (Connor et al., 2006; Herskovic et al., 2016; Rangel and Engle, 2012). Also, the empirical study of Barigozzi and Hallin (2016) shows that the market volatility shocks are hitting not only factor volatilities but also idiosyncratic volatilities. Thus, it would be interesting and important to develop a parametric model for the idiosyncratic volatility process based on the continuous time model. Since the factor volatility process proposed in this paper is developed independently from the idiosyncratic volatility process, and so it would be easy to combine other idiosyncratic volatility process with the proposed factor GARCH-Itô model. However, modeling idiosyncratic volatility process is hard when sparsity condition is imposed, which is necessary for identifiability and high-dimensional covariance matrix regularization. One possible idea is to appeal special sparsity structure. It was documented in Fan et al. (2016a) that the volatility matrix of idiosyncratic components admits a block diagonal structure when sorted by sectors or industries. We can then model the dynamics of the idiosyncratic volatility matrices for each sector or industry by using a multivariate GARCH or DCC (dynamic conditional correlation) model (Bollerslev, 1990; Bollerslev et al., 1988; Engle, 2002; Engle and Kroner, 1995). This part of dynamics can also be modeled by the GARCH-Itô process in the same way as we model the factor volatility.

To estimate the idiosyncratic volatility matrix $\boldsymbol{\Sigma}_n$, we employ the principal orthogonal component thresholding (POET) (Fan et al., 2013) procedure as follows. For any given integrated volatility estimator $\widehat{\boldsymbol{\Gamma}}_n$, the spectral-decomposition provides

$$\widehat{\boldsymbol{\Gamma}}_n = \sum_{i=1}^{p} \widehat{\xi}_{n,i} \widehat{\mathbf{q}}_{n,i} \widehat{\mathbf{q}}_{n,i}^{\top},$$

where $\widehat{\xi}_{n,i}$ is the ith largest eigenvalue of $\widehat{\boldsymbol{\Gamma}}_n$ and $\widehat{\mathbf{q}}_{n,i}$ is its corresponding eigenvector. The sparse volatility matrix is estimated by the thresholding of the input idiosyncratic volatility matrix $\widetilde{\boldsymbol{\Sigma}}_n = (\widetilde{\Sigma}_{n,ij})_{1\le i,j\le p} = \widehat{\boldsymbol{\Gamma}}_n - \sum_{i=1}^{r} \widehat{\xi}_{n,i} \widehat{\mathbf{q}}_{n,i} \widehat{\mathbf{q}}_{n,i}^{\top}$:

$$\widehat{\Sigma}_{n,ij} = \begin{cases} \widetilde{\Sigma}_{n,ij} \vee 0, & \text{if } i = j \\ s_{ij}(\widetilde{\Sigma}_{n,ij}) \mathbf{1}(|\widetilde{\Sigma}_{n,ij}| \ge \varpi_{ij}), & \text{if } i \ne j \end{cases} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_n = (\widehat{\Sigma}_{n,ij})_{1\le i,j\le p}, \tag{4.1}$$

where the thresholding function $s_{ij}(\cdot)$ satisfies that $|s_{ij}(x) - x| \le \varpi_{ij}$, and the thresholding level $\varpi_{ij} = \varpi_m \sqrt{(\widetilde{\Sigma}_{ii} \vee 0)(\widetilde{\Sigma}_{jj} \vee 0)}$ is the same as applying thresholding ϖ_m to the correlation matrix. Interesting examples of the thresholding function $s_{ij}(x)$ include the soft thresholding and hard thresholding function. Under some conditions, Fan and Kim (2018) showed that $\widehat{\boldsymbol{\Sigma}}_n$ converges to $\boldsymbol{\Sigma}_n$ in terms of the spectral norm. We use $\widehat{\boldsymbol{\Sigma}}_n$ as the idiosyncratic volatility matrix estimator.

Next, using the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ in (3.7), we estimate the conditional expected factor volatility matrix by

$$\widehat{\boldsymbol{\Psi}}_{n+1} = \sum_{i=1}^{r} \widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}}) \widehat{\mathbf{q}}_{n,i} \widehat{\mathbf{q}}_{n,i}^{\top}, \tag{4.2}$$

where by Proposition 2.1,

$$\widehat{\mathbf{h}}_{n+1}(\widehat{\boldsymbol{\theta}}) = (\widehat{h}_{n+1,1}(\widehat{\boldsymbol{\theta}}), \ldots, \widehat{h}_{n+1,r}(\widehat{\boldsymbol{\theta}}))^{\top} = p\widehat{\boldsymbol{\beta}}^{-2} \left(e^{\widehat{\boldsymbol{\beta}}} - \mathbf{I}_r - \widehat{\boldsymbol{\beta}}\right) \widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\varrho}} \widehat{\boldsymbol{\lambda}}_n(\widehat{\boldsymbol{\theta}}),$$

$\widehat{\boldsymbol{\lambda}}_n(\widehat{\boldsymbol{\theta}}) = (\widehat{\lambda}_{n,1}(\widehat{\boldsymbol{\theta}}), \ldots, \widehat{\lambda}_{n,r}(\widehat{\boldsymbol{\theta}}))^{\top}$, $\widehat{\lambda}_{n,i}(\boldsymbol{\theta})$'s are defined in (3.6), $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_{i,j})_{1\le i,j\le r}$, $\widehat{\boldsymbol{\omega}} = (\widehat{\omega}_1, \ldots, \widehat{\omega}_r)^{\top}$, $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_r)^{\top}$, and $\widehat{\boldsymbol{\varrho}} = \widehat{\boldsymbol{\beta}}^{-2} \left(e^{\widehat{\boldsymbol{\beta}}} - \mathbf{I}_r - \widehat{\boldsymbol{\beta}}\right) (\mathrm{diag}(\widehat{\boldsymbol{\gamma}}) - \mathbf{I}_r) + \widehat{\boldsymbol{\beta}}^{-1} \left(e^{\widehat{\boldsymbol{\beta}}} - \mathbf{I}_r\right)$. Then, combining the idiosyncratic volatility matrix estimator $\widehat{\boldsymbol{\Sigma}}_n$ and conditional expected factor volatility matrix estimator $\widehat{\boldsymbol{\Psi}}_{n+1}$, we predict the conditional expected volatility matrix as follows:

$$\widetilde{\boldsymbol{\Gamma}}_{n+1} = \widehat{\boldsymbol{\Psi}}_{n+1} + \widehat{\boldsymbol{\Sigma}}_n. \tag{4.3}$$

To investigate its asymptotic behavior, we make the following technical conditions.

Assumption 3.

(a) For some fixed constant C_1, we have

$$\frac{p}{r} \max_{1\le i\le p} \sum_{j=1}^{r} q_{ij}^2 \le C_1,$$

where $\mathbf{q}_j = (q_{1j}, \ldots, q_{pj})^{\top}$ is the jth eigenvector of $\boldsymbol{\Psi}_n$;

(b) There is some fixed positive constant C_2 such that $\xi_{n,1}/D_{\xi} \le C_2$ a.s., and the smallest eigenvalue of $\boldsymbol{\Sigma}_n$ stays away from zero;

(c) $s_p/\sqrt{p} + \sqrt{\log p / m^{1/2}} = o(1)$.

Remark 11. Assumption 3(a) is incoherence condition which is usually imposed on investigating the low-rank matrix inferences (Candès et al., 2011; Fan et al., 2016).

The following theorem shows the asymptotic behaviors of the conditional expected volatility matrix estimators.

Theorem 4.1. *Under the assumptions of Theorem 3.2, suppose that*

$$\Pr\left\{ \max_{1\le i,j\le p} |\widehat{\Gamma}_{n,ij} - \Gamma_{n,ij}| \ge C \sqrt{\frac{\log(p \vee m)}{m^{1/2}}} \right\} \le p^{-1}, \tag{4.4}$$

and Assumption 3 are met. Take $\varpi_m = C_{\varpi} \tau_m$ for some large fixed constant C_{ϖ}, where $\tau_m = s_p/p + \sqrt{\log(p \vee m)/m^{1/2}}$. Then we have

$$\|\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\|_2 = O_p\left(s_p \tau_m^{1-\delta}\right), \tag{4.5}$$

$$\|\widehat{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n\|_{\max} = O_p\left(\tau_m\right), \tag{4.6}$$

$$\|\widetilde{\boldsymbol{\Gamma}}_{n+1}-\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)\|_{\boldsymbol{\Gamma}^*}=O_p\left(p^{1/2}m^{-1/2}+s_p/p^{1/2}+s_p\tau_m^{1-\delta}\right), \tag{4.7}$$

where the relative Frobenius norm $\|\mathbf{A}\|_{\boldsymbol{\Gamma}^*}^2=p^{-1}\|\boldsymbol{\Gamma}^{*-1/2}\mathbf{A}\boldsymbol{\Gamma}^{*-1/2}\|_F^2$ *and* $\boldsymbol{\Gamma}^*=\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)$.

Remark 12. Condition (4.4) can be obtained under the bounded instantaneous volatility condition (Tao et al., 2013b). Thanks to the localization argument made in Section 4.4.1 of Jacod and Protter (2011), we assume the bounded condition without loss of generality. So this condition is not restrictive at all.

Remark 13. Theorem 4.1 shows that the conditional expected volatility matrix estimator $\widetilde{\boldsymbol{\Gamma}}_{n+1}$ is consistent in terms of relative Frobenius norm as long as $p=o(n)$. Its convergence rate is comparable with the convergence rates obtained in Fan and Kim (2018). The difference is the additional term $s_p/p^{1/2}$ which comes from estimating the parameters $\boldsymbol{\theta}_0$. As discussed in Remark 8, it is due to identifying the latent factor volatility, and so if the latent factor part is observable, the term $s_p/p^{1/2}$ is removed.

To evaluate the conditional expected volatility matrix $\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)$, we assume that the eigenvectors of the latent factor volatility matrices $\boldsymbol{\Psi}_n$'s are constant. Under this condition, we are able to estimate eigenvectors using the whole period information. For example, we can estimate the eigenvectors $\mathbf{q}_1,\ldots,\mathbf{q}_r$ by the first r eigenvectors, $\widehat{\mathbf{q}}_1^c,\ldots,\widehat{\mathbf{q}}_r^c$, of $\frac{1}{n}\sum_{d=1}^n\widehat{\boldsymbol{\Gamma}}_d$. With this aggregated estimation, the eigenvalues for the dth day are now estimated by

$$\widehat{\xi}_{d,i}^c=(\widehat{\mathbf{q}}_i^c)^\top\widehat{\boldsymbol{\Gamma}}_d\widehat{\mathbf{q}}_i^c \quad \text{for } i=1,\ldots,r.$$

Then we apply the quasi-maximum likelihood estimation procedure proposed in Section 3 with $\widehat{\mathbf{q}}_i^c$ and $\widehat{\xi}_{d,i}^c$ instead of $\widehat{\mathbf{q}}_{d,i}$ and $\widehat{\xi}_{d,i}$, and $\widehat{\boldsymbol{\theta}}^c$ denotes the resulting quasi-maximum likelihood estimator. Let us call it the aggregated QMLE. Using $\widehat{\boldsymbol{\theta}}^c$, $\widehat{\xi}_{d,i}^c$, and $\widehat{\mathbf{q}}_i^c$, we estimate the conditional expected factor volatility matrix by

$$\widehat{\boldsymbol{\Psi}}_{n+1}^c=\sum_{i=1}^r\widehat{h}_{n+1,i}^c(\widehat{\boldsymbol{\theta}}^c)\widehat{\mathbf{q}}_i^c(\widehat{\mathbf{q}}_i^c)^\top, \tag{4.8}$$

where

$$\begin{aligned}
\widehat{\mathbf{h}}_{n+1}^c(\widehat{\boldsymbol{\theta}}^c)&=(\widehat{h}_{n+1,1}^c(\widehat{\boldsymbol{\theta}}^c),\ldots,\widehat{h}_{n+1,r}^c(\widehat{\boldsymbol{\theta}}^c))^\top=p(\widehat{\boldsymbol{\beta}}^c)^{-2}\left(e^{\widehat{\boldsymbol{\beta}}^c}-\mathbf{I}_r-\widehat{\boldsymbol{\beta}}^c\right)\widehat{\boldsymbol{\omega}}^c+\widehat{\boldsymbol{\varrho}}^c\widehat{\boldsymbol{\lambda}}_n^c(\widehat{\boldsymbol{\theta}}^c),\\
\widehat{\boldsymbol{\lambda}}_n^c(\widehat{\boldsymbol{\theta}}^c)&=(\widehat{\lambda}_{n,1}^c(\widehat{\boldsymbol{\theta}}^c),\ldots,\widehat{\lambda}_{n,r}^c(\widehat{\boldsymbol{\theta}}^c))^\top,\\
\widehat{\lambda}_{n,i}^c(\widehat{\boldsymbol{\theta}}^c)&=p\widehat{\omega}_i^c+\widehat{\gamma}_i^c\widehat{\lambda}_{n-1,i}^c(\widehat{\boldsymbol{\theta}}^c)+\sum_{l=1}^r\widehat{\beta}_{i,l}^c\left[(\widehat{\mathbf{q}}_l^c)^\top\{\mathbf{X}(n)-\mathbf{X}(n-1)\}-\sqrt{p}\widehat{\mu}_l^c\right]^2,
\end{aligned}$$

and $\widehat{\boldsymbol{\beta}}^c$, $\widehat{\boldsymbol{\gamma}}^c$, $\widehat{\boldsymbol{\omega}}^c$, and $\widehat{\boldsymbol{\varrho}}^c$ are estimated using $\widehat{\boldsymbol{\theta}}^c$ instead of $\widehat{\boldsymbol{\theta}}$. Finally, the conditional expected volatility matrix is estimated by

$$\widetilde{\boldsymbol{\Gamma}}_{n+1}^c=\widehat{\boldsymbol{\Psi}}_{n+1}^c+\widehat{\boldsymbol{\Sigma}}_n, \tag{4.9}$$

where $\widehat{\boldsymbol{\Sigma}}_n$ is defined in (4.1). The alternative estimator $\widetilde{\boldsymbol{\Gamma}}_{n+1}^c$ can enjoy the same asymptotic properties in Theorem 4.1 under the constant eigenvector condition.

5. Numerical study

5.1. A simulation study

We conducted simulations to verify the finite sample performances of the proposed estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^c$ and conditional expected volatility matrix estimators $\widetilde{\boldsymbol{\Gamma}}_{n+1}$ and $\widetilde{\boldsymbol{\Gamma}}_{n+1}^c$ given the past n period observations. We generated the log-prices $\mathbf{X}(t_{i,j})$ for n days with frequency $1/m$ on each day: $t_{i,j}=i-1+j/m$, $i=1,\ldots,n$, $j=1,\ldots,m$, from the factor GARCH-Itô model in Definition 1 with the following form:

$$\begin{aligned}
d\mathbf{X}(t)&=\mathbf{Q}\boldsymbol{\lambda}^{1/2}(t)d\mathbf{W}(t)+\boldsymbol{\sigma}^\top dW^*(t),\\
\lambda_{t,i}(\boldsymbol{\theta}_0)&=\lambda_{[t],i}(\boldsymbol{\theta}_0)+(t-[t])\{p\omega_{0,i}+(\gamma_{0,i}-1)\lambda_{[t],i}(\boldsymbol{\theta}_0)\}+\sum_{l=1}^r\beta_{0,i,l}\left\{\int_{[t]}^t\sqrt{\lambda_{t,l}(\boldsymbol{\theta}_0)}dW_l(t)\right\}^2,
\end{aligned}$$

where $\boldsymbol{\lambda}(t)=\mathrm{diag}(\lambda_{t,1}(\boldsymbol{\theta}_0),\ldots,\lambda_{t,r}(\boldsymbol{\theta}_0))$ with $r=3$,

$$\begin{aligned}
\boldsymbol{\theta}_0&=(\omega_{0,1},\ldots,\omega_{0,r},\gamma_{0,1},\ldots,\gamma_{0,r},\beta_{0,1,1},\ldots,\beta_{0,r,r})\\
&=(0.15,0.125,0.1,0.2,0.15,0.1,0.25,0.2,0.15,0,0.15,0.125,0,0,0.1),
\end{aligned}$$

and $\mathbf{W}(t)=(W_1(t),\ldots,W_r(t))^\top$ and $\mathbf{W}^*(t)$ are r- and p-dimensional independent Brownian motions. To generate the eigenvector matrix $\mathbf{Q}$, we first made a p by p matrix whose elements were generated by i.i.d. uniform [0, 1], and chose its

Table 1
The MADEs of QMLEs estimated with/without aggregation for $p = 200$, $r = 3$, $n = 100$, $m = 400, 1000, 2000, 5000$.

		MADE$\times 10^2$											
		ω		γ		$\beta_{i,1}$		$\beta_{i,2}$		$\beta_{i,3}$		μ	
	m	$\widehat{\theta}$	$\widehat{\theta}^c$	$\widehat{\theta}$	$\widehat{\theta}^c$	$\widehat{\theta}$	$\widehat{\theta}^c$	$\widehat{\theta}$	$\widehat{\theta}^c$	$\widehat{\theta}$	$\widehat{\theta}^c$	$\widehat{\theta}$	$\widehat{\theta}^c$
λ_1	400	4.39	2.18	7.83	8.12	3.52	3.36	4.39	4.12	5.61	5.89	2.41	2.13
	1000	3.34	1.96	6.08	6.16	2.95	2.30	3.62	3.35	4.50	4.79	1.95	1.64
	2000	2.85	1.86	5.15	5.31	2.60	2.00	3.18	2.87	4.38	4.31	1.81	1.66
	5000	2.33	1.72	4.21	4.22	2.33	1.91	2.57	2.41	3.66	3.48	1.55	1.57
λ_2	400	3.02	1.42	10.86	10.69	1.85	0.44	3.75	3.69	4.18	3.63	2.48	1.76
	1000	2.39	1.37	9.86	8.78	1.23	0.33	2.70	2.51	3.15	2.62	1.87	1.71
	2000	1.97	1.36	8.09	7.92	0.89	0.30	2.39	2.07	2.79	2.53	1.67	1.46
	5000	1.70	1.22	7.32	6.86	0.56	0.22	1.69	1.48	2.48	2.19	1.31	1.32
λ_3	400	1.61	0.84	9.30	9.26	0.39	0.23	2.05	0.58	3.33	3.57	2.95	1.82
	1000	1.27	0.81	8.01	8.09	0.31	0.20	1.54	0.43	2.46	2.59	2.32	1.73
	2000	1.29	0.94	7.67	7.66	0.22	0.17	1.24	0.37	2.16	1.77	2.21	1.45
	5000	1.06	0.85	6.22	6.38	0.17	0.13	0.84	0.33	1.65	1.34	1.54	1.33

first r eigenvectors as $\mathbf{Q}$. To obtain the sparse volatility matrix $\boldsymbol{\Sigma} = (\Sigma_{ij})_{1\le i,j\le p}$, its off-diagonal elements were generated as follows:

$$\Sigma_{ij} = 0.5^{|i-j|}\sqrt{\Sigma_{ii}\Sigma_{jj}},$$

the diagonal elements are $\Sigma_{ii} = 1$, $i = 1, \ldots, p$, and $\boldsymbol{\sigma}$ is the Cholesky decomposition of $\boldsymbol{\Sigma}$. We took the initial values $\mathbf{X}_0 = 0$ and $\boldsymbol{\lambda}_0 = p \times \mathrm{diag}(1.5, 1, 0.5)$, the rank $r = 3$, $n = 100$, and $p = 200$. We varied m from 400 to 5,000.

The low-frequency data were taken to be $\mathbf{X}(i)$, $i = 0, \ldots, n$. The high-frequency data $\mathbf{Y}(t_{d,j})$ were simulated from the model (3.2), where the true log-stock prices $\mathbf{X}(t_{d,j})$ were taken from the generated log-prices described above, and for the ith asset, the market microstructural noises $\epsilon_i(t_{d,j})$ were from i.i.d. normal distribution with mean zero and standard deviation $0.01\sqrt{\Sigma_{ii}}$. To estimate the integrated volatility matrices $\boldsymbol{\Gamma}_d$'s, we employed the pre-averaging realized volatility matrix (PRVM) estimator (Christensen et al., 2010; Jacod et al., 2009) as follows:

$$\widehat{\boldsymbol{\Gamma}}_d = \frac{1}{\psi K}\sum_{k=1}^{m-K+1}\left\{\bar{\mathbf{Y}}(t_{d,k})\bar{\mathbf{Y}}(t_{d,k}) - \varsigma\,\widehat{\boldsymbol{\eta}}\right\}, \tag{5.1}$$

where

$$\widehat{\boldsymbol{\eta}}_d = \mathrm{diag}(\widehat{\eta}_{d,11}, \ldots, \widehat{\eta}_{d,pp}), \qquad \widehat{\eta}_{d,ii} = \frac{1}{2m}\sum_{k=1}^{m}\left\{Y_i(t_{d,k}) - Y_i(t_{d,k-1})\right\}^2,$$

$$\bar{\mathbf{Y}}(t_{d,k}) = \sum_{l=1}^{K-1} g\left(\frac{l}{K}\right)\left\{\mathbf{Y}(t_{d,k+l}) - \mathbf{Y}(t_{d,k+l-1})\right\},$$

$$\varsigma = \sum_{l=0}^{K-1}\left\{g\left(\frac{l}{K}\right) - g\left(\frac{l+1}{K}\right)\right\}^2, \qquad \psi = \int_0^1 g(t)^2 dt,$$

the bandwidth size $K = \lfloor m^{1/2}\rfloor$, and the weight function $g(x) = x \wedge (1-x)$. We estimated the model parameters by the proposed QMLE procedure in Section 3.3 with/without aggregation. The whole simulation procedure was repeated 500 times.

Table 1 reports the mean absolute deviation errors (MADE) of the model parameter estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^c$. As expected, it shows that as m increases, the MADEs decrease; the aggregated QMLE, $\widehat{\boldsymbol{\theta}}^c$, outperforms $\widehat{\boldsymbol{\theta}}$. This is due to more accurate aggregated estimated eigenvectors $\widehat{\mathbf{q}}_i^c$ under the constant eigenvector condition. These results support the theoretical findings in Section 3.3.

Using the QMLEs, we predicted the conditional expected eigenvalues, $h_{n+1,i}(\boldsymbol{\theta}_0)$, and conditional expected integrated volatility matrix, $\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)$, given the past n period observations. For example, the conditional expected eigenvalues are estimated by $\widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}})$ and $\widehat{h}^c_{n+1,i}(\widehat{\boldsymbol{\theta}}^c)$ in (4.2) and (4.8), respectively, and the conditional expected integrated volatility matrix by $\widetilde{\boldsymbol{\Gamma}}_{n+1}$ and $\widetilde{\boldsymbol{\Gamma}}^c_{n+1}$ in (4.3) and (4.9), respectively. On the other hand, when not considering parametric models, we often assume the martingale structure on the integrated volatility matrices and estimate the conditional expected volatility matrix using the previous period estimator, that is, $\widehat{\boldsymbol{\Gamma}}_n$. In the light of this, we compare the proposed factor GARCH-Itô estimators with the nonparametric estimator $\widehat{\boldsymbol{\Gamma}}_n$. For the conditional expected eigenvalues $h_{n+1,i}(\boldsymbol{\theta}_0)$, we measured the errors by the

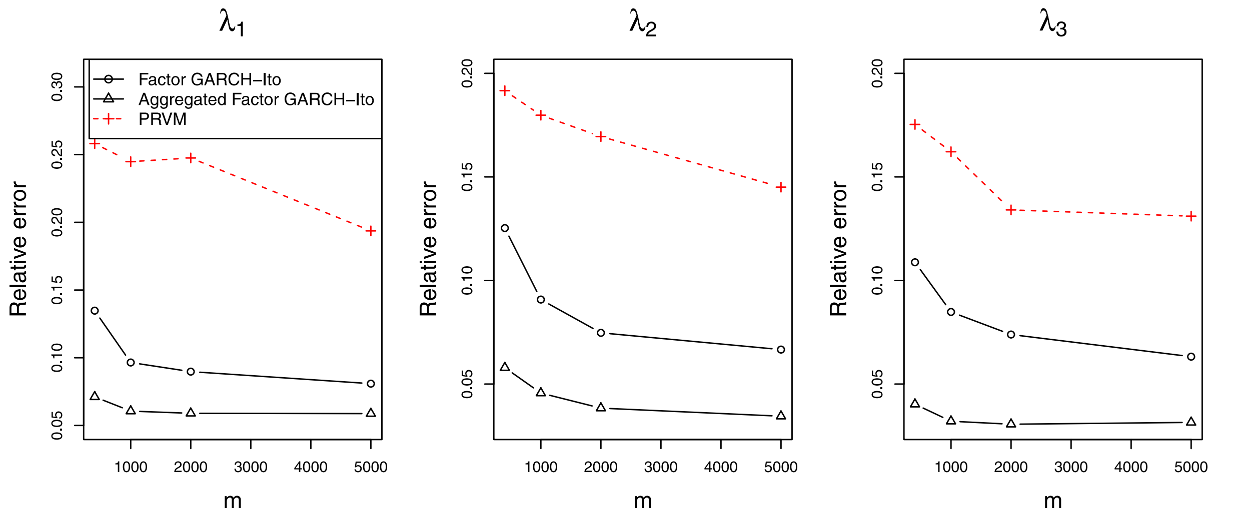

Fig. 1. Average relative estimation errors of estimators for the conditional eigenvalues against m, with $p = 200$, $r = 3$, and $n = 100$.

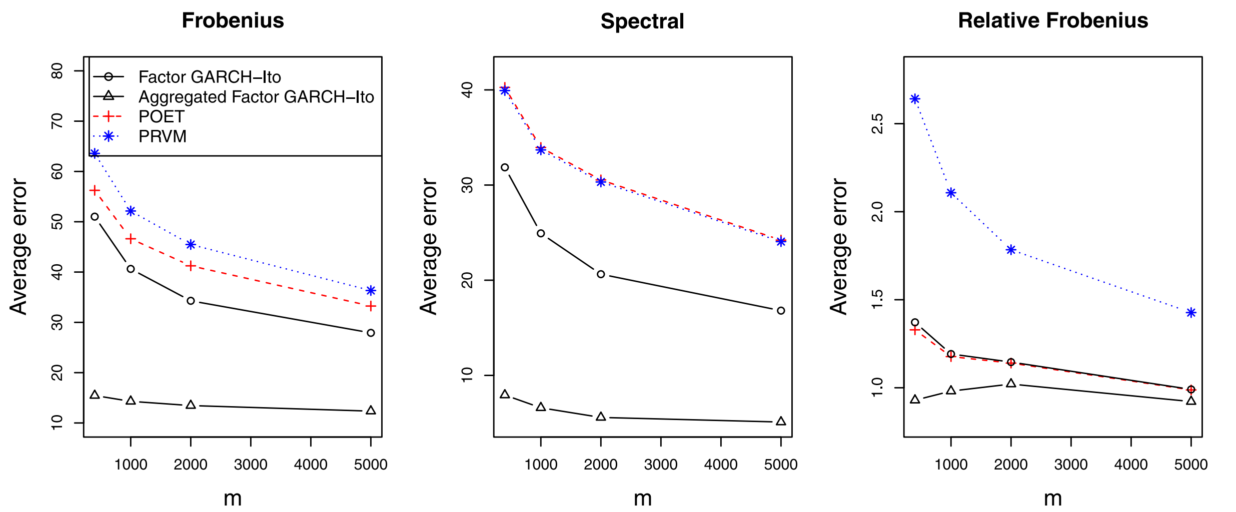

Fig. 2. Average estimation errors of estimators for the conditional integrated volatility matrix against m under different matrix norms, with $p = 200$, $r = 3$, and $n = 100$.

relative errors

$$\frac{|\widehat{\lambda}_i - h_{n+1,i}(\boldsymbol{\theta}_0)|}{h_{n+1,i}(\boldsymbol{\theta}_0)},$$

where $\widehat{\lambda}_i$ could be one of three estimators: unaggregated estimator $\widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}})$, aggregated estimator $\widehat{h}^c_{n+1,i}(\widehat{\boldsymbol{\theta}}^c)$, and the nonparametric estimation, which is the ith eigenvalue $\widehat{\xi}_{n,i}$ of the PRVM estimator $\widehat{\boldsymbol{\Gamma}}_n$. Fig. 1 depicts the relative errors against m. We can find that the estimates based on the factor GARCH-Itô model, $\widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}})$ and $\widehat{h}^c_{n+1,i}(\widehat{\boldsymbol{\theta}}^c)$, are better than the ith eigenvalue, $\widehat{\xi}_{n,i}$, of the PRVM estimator $\widehat{\boldsymbol{\Gamma}}_n$. The results reveal that the aggregated estimator $\widehat{h}^c_{n+1,i}(\widehat{\boldsymbol{\theta}}^c)$ has the smallest relative errors.

Finally, we consider the conditional expected integrated volatility matrix $\mathrm{E}\,(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n)$. The integrated volatility matrix has the low-rank plus sparse structure, and so we employ the POET procedure to account for the such structure. Denote the POET estimator at the nth day by $\widehat{\boldsymbol{\Gamma}}^{POET}_n$. For the thresholding step, we used the thresholding level $\sqrt{2\log p/m^{1/2}}$. We used $\widetilde{\boldsymbol{\Gamma}}_{n+1}$, $\widetilde{\boldsymbol{\Gamma}}^c_{n+1}$, $\widehat{\boldsymbol{\Gamma}}_n$, and $\widehat{\boldsymbol{\Gamma}}^{POET}_n$ as the estimator of the conditional expected integrated volatility matrix $\mathrm{E}\,(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n)$. We measured the average errors of matrix estimation using the Frobenius norm, spectral norm, and relative Frobenius norm (Fan et al., 2013). Fig. 2 depicts the average errors under different matrix norms against m. It is clear that the estimators based on the factor GARCH-Itô model, $\widetilde{\boldsymbol{\Gamma}}_{n+1}$ and $\widetilde{\boldsymbol{\Gamma}}^c_{n+1}$, outperform the other two estimators. When comparing $\widetilde{\boldsymbol{\Gamma}}_{n+1}$ with $\widetilde{\boldsymbol{\Gamma}}^c_{n+1}$, the aggregated estimator $\widetilde{\boldsymbol{\Gamma}}^c_{n+1}$ has smaller average errors, as expected.

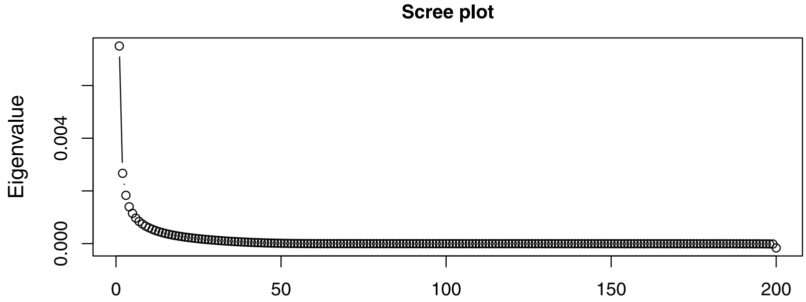

Fig. 3. The scree plot for average eigenvalues of PRVM.

5.2. An empirical study

In this section, we applied the proposed estimators to high-frequency trading data for 200 assets from January 1st to December 31st in 2013 ($n = 252$). The data is taken from the Wharton Data Service (WRDS) system. Top 200 large trading volume stocks were selected among S&P 500, and we used 1-min log-returns. To employ the proposed estimators, we first need to find the rank r. To do this, we first calculated 252 daily integrated volatility matrices using the PRVM estimation method in (5.1). We estimated the rank r using the procedure proposed by Aït-Sahalia and Xiu (2017) as follows:

$$\widehat{r} = \arg\min_{1\le j\le r_{\max}} \sum_{d=1}^{252} \left[p^{-1}\widehat{\xi}_{d,j} + j \times c_1 \left\{ \sqrt{\log p/m^{1/2}} + p^{-1}\log p \right\}^{c_2} \right] - 1,$$

where $r_{\max} = 30$, $c_1 = 0.02 \times \widehat{\xi}_{d,30}$, and $c_2 = 0.5$. The minimum value is $\widehat{r} = 3$. Also we draw the scree plot using the average values of eigenvalues from the 252 PRVM estimators. Fig. 3 shows that the possible values of the rank r are 1, 2, 3, 4. From those results, we determined $r = 3$.

We used the open-to-close period high frequency data as the one time period. To evaluate the quasi-likelihood function, we used the first r eigenvalues of the first day as the initial eigenvalue $\boldsymbol{\lambda}_0$. We calculated $\widehat{\boldsymbol{\theta}}$ and estimated the conditional integrated eigenvalues $\mathbf{h}_d(\boldsymbol{\theta}_0)$ by $\widehat{\mathbf{h}}_d(\widehat{\boldsymbol{\theta}})$ in (4.2). The estimated values are $\widehat{\boldsymbol{\theta}} = (\widehat{\omega}_1, \ldots, \widehat{\omega}_3, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_3, \widehat{\beta}_{1,1}, \ldots, \widehat{\beta}_{3,3}, \widehat{\mu}_1, \ldots, \widehat{\mu}_3) = 10^{-3} \times (0.012, 0.006, 0.004, 362.479, 242.343, 315.174, 139.116, 74.370, 63.324, 17.151, 53.201, 75.818, 9.190, 25.905, 38.370, 2.366, 0.554, 0.670)$. Fig. 4 depicts the estimated daily integrated eigenvalues by the integrated eigenvalue estimates $\widehat{\xi}_{d,i}$ and estimated conditional expected integrated eigenvalues $\widehat{h}_{d,i}(\widehat{\boldsymbol{\theta}})$. It shows that the estimated conditional integrated eigenvalue estimators $\widehat{h}_{d,i}(\widehat{\boldsymbol{\theta}})$ can capture the dynamics of the integrated eigenvalue estimates $\widehat{\xi}_{d,i}$ and their path is smoother than the integrated eigenvalue estimates $\widehat{\xi}_{d,i}$.

We consider the constrained portfolio allocation problem using our forecasted volatility:

$$\min_{\mathbf{w} \text{ s.t. } \mathbf{w}^\top \mathbf{J}=1, \|\mathbf{w}\|_1 = c_0} \mathbf{w}^\top \widetilde{\boldsymbol{\Gamma}}_{n+1} \mathbf{w}, \tag{5.2}$$

where $\mathbf{J} = (1, \ldots, 1)^\top \in \mathbb{R}^p$ and c_0 is the gross exposure constraint which varied from 1 to 2. We estimated the conditional expected volatility matrix $\mathrm{E}\,(\boldsymbol{\Gamma}_{n+1}|\,\mathcal{F}_n)$ using the past n-period observations, and we varied n from 148 to 252. That is, we used at least 148 daily observations (7 months) to estimate $\boldsymbol{\theta}_0$. To check the dependency of the split points, we calculated the out-of-sample risk for three different testing periods: from 148 to 252 (5 months), from 169 to 252 (4 months), and from 190 to 252 (3 months). A popular alternative method is to use the integrated volatility matrix in the previous period as the estimator of the future volatility matrix $\boldsymbol{\Gamma}_{n+1}$. Thus, for the comparison, we also used the PRVM estimator $\widehat{\boldsymbol{\Gamma}}_n$ and POET estimator $\widehat{\boldsymbol{\Gamma}}_n^{POET}$ in order to construct the portfolio in (5.2). To make the estimates positive semi-definite, we first projected the input volatility matrix PRVM estimators onto the positive semi-definite cone in the spectral norm. For the thresholding step in estimating the covariance matrix of the idiosyncratic component, we utilized the global industry classification standard (GICS) for sectors, and maintained within-sector volatilities but set others to zero (Fan et al., 2016a). To model the dynamics for the idiosyncratic volatilities, we apply the factor GARCH-Itô model to each block with the full rank. Under the block diagonal structure, the factor GARCH-Itô model is still complex, which causes huge estimation errors. To simplify the model, we assume that each eigenvalue is not affected by other eigenvalue parts, that is, $\boldsymbol{\beta}$ is assumed to be diagonal, and eigenvectors are constant over time. Then we can estimate the model parameters marginally. We call it idiosyncratic GARCH-Itô.

Fig. 5 plots the out-of-sample risk of the portfolio constructed by the factor GARCH-Itô estimator $\widetilde{\boldsymbol{\Gamma}}_{n+1}$, aggregated factor GARCH-Itô estimator $\widetilde{\boldsymbol{\Gamma}}_{n+1}^c$, aggregated factor GARCH-Itô with idiosyncratic GARCH-Itô, POET estimator $\widehat{\boldsymbol{\Gamma}}_n^{POET}$, and PRVM estimator $\widehat{\boldsymbol{\Gamma}}_n$. Here the portfolio risk was measured using the 1-min portfolio log-returns for one day. For the three different periods, we have similar behaviors, and so the results do not significantly depend on the split points. We find that for the purpose of portfolio allocation, the factor GARCH-Itô performs well and improves the performance of the POET. On the other

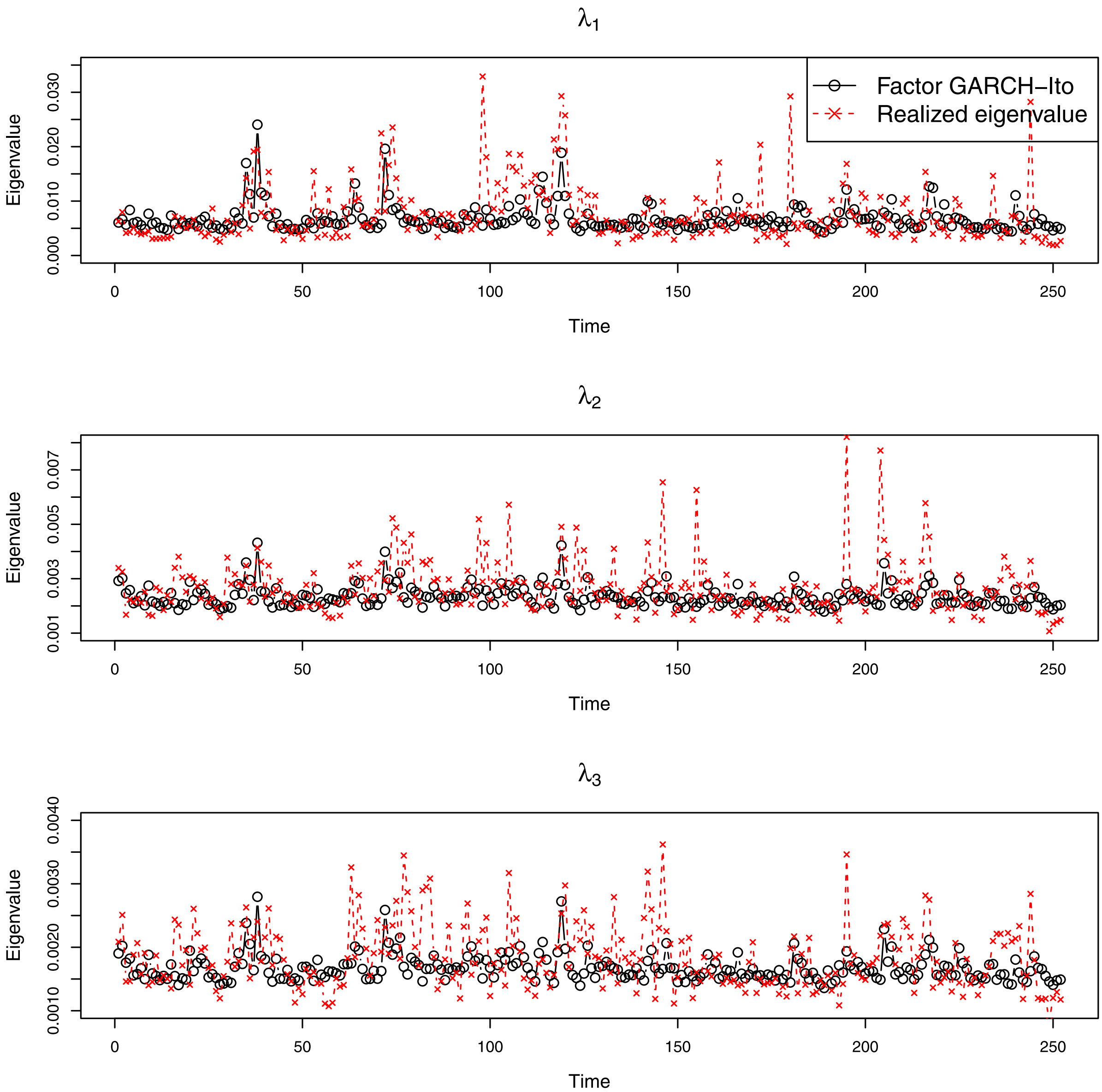

Fig. 4. The plots of daily integrated eigenvalue estimates, the integrated eigenvalue estimate $\widehat{\xi}_{d,i}$ and factor GARCH-Itô eigenvalue $\widehat{h}_{d,i}(\widehat{\boldsymbol{\theta}})$.

hand, the allocation based on the PRVM estimator becomes unstable as the exposure constraint increases and its out-of-sample risk also much bigger than those of the factor GARCH-Itô and POET. Meanwhile, the aggregated factor GARCH-Itô has the smallest risk when the exposure is small, that is, the portfolio is sparse. We also find that the idiosyncratic GARCH-Itô estimates show stable performance over the exposure level. The results suggest that the proposed factor GARCH-Itô model can explain the dynamics of the volatility in the stock market. Also it is important to model the dynamics for the idiosyncratic volatility, and so we need to study its asymptotic behaviors. We leave this for the future study.

6. Proof

6.1. Proof of Proposition 2.1

Proof of Proposition 2.1. (a). Let

$$\mathbf{R}(k) = (R_1(k), \ldots, R_r(k))^\top \quad \text{and} \quad R_i(k) = \int_{d-1}^{d} \frac{(d-t)^k}{k!} \lambda_{t,i}(\boldsymbol{\theta}_i) dt.$$

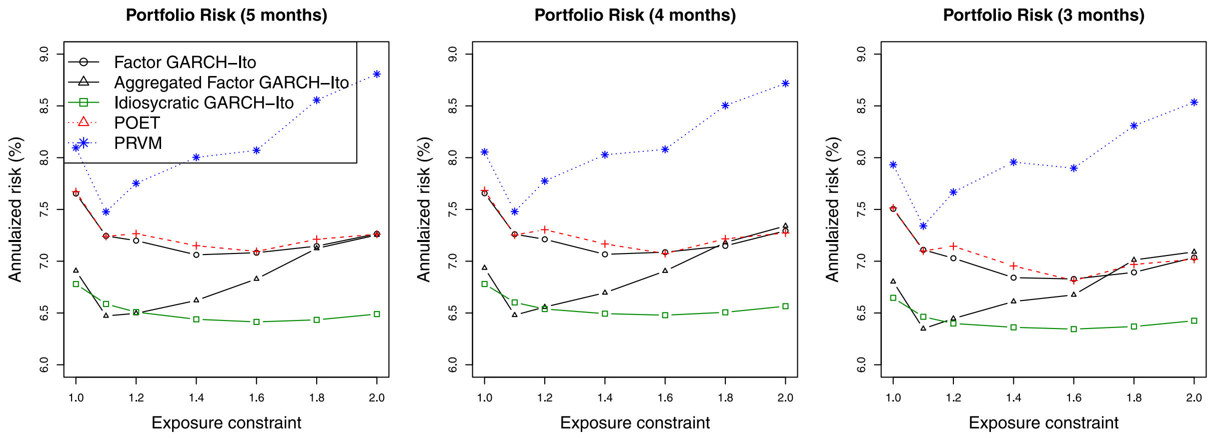

Fig. 5. The out-of-sample risks of the optimal portfolios constructed by using the volatility matrix from the factor GARCH-Itô, aggregated factor GARCH-Itô, aggregated factor GARCH-Itô with idiosyncratic GARCH-Itô, POET, and PRVM estimators.

By the Itô's lemma, we have

$$\begin{aligned} R_i(k) &= \frac{p\omega_i + (\gamma_i + k + 1)\lambda_{d-1,i}(\boldsymbol{\theta}_i)}{(k+2)!} \\ &\quad + 2\sum_{l=1}^{r} \beta_{i,l} \int_{d-1}^{d} \frac{(d-t)^{k+1}}{(k+1)!} \int_{d-1}^{t} \sqrt{\lambda_{s,l}(\boldsymbol{\theta}_l)}dW_l(s)\sqrt{\lambda_{t,l}(\boldsymbol{\theta}_l)}dW_l(t) \\ &\quad + \sum_{l=1}^{r} \beta_{i,l} R_l(k+1) \text{ a.s.,} \end{aligned}$$

and then

$$\begin{aligned} \mathbf{R}(k) &= \frac{p\boldsymbol{\omega} + \{\mathrm{diag}(\boldsymbol{\gamma}) + (k+1)\mathbf{I}_r\}\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta})}{(k+2)!} \\ &\quad + \boldsymbol{\beta}\left(2\int_{d-1}^{d} \frac{(d-t)^{k+1}}{(k+1)!} \int_{d-1}^{t} \sqrt{\lambda_{s,l}(\boldsymbol{\theta}_l)}dW_l(s)\sqrt{\lambda_{t,l}(\boldsymbol{\theta}_l)}dW_l(t)\right)^{\top}_{l=1,\ldots,r} \\ &\quad + \boldsymbol{\beta}\mathbf{R}(k+1) \text{ a.s.} \end{aligned}$$

Thus,

$$\begin{aligned} \mathbf{R}(0) &= \int_{d-1}^{d} \boldsymbol{\lambda}_t(\boldsymbol{\theta})dt \\ &= \sum_{k=0}^{\infty} \boldsymbol{\beta}^k \frac{p\boldsymbol{\omega} + \{\mathrm{diag}(\boldsymbol{\gamma}) + (k+1)\mathbf{I}_r\}\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta})}{(k+2)!} \\ &\quad + \sum_{k=0}^{\infty} \boldsymbol{\beta}^{k+1}\left(2\int_{d-1}^{d} \frac{(d-t)^{k+1}}{(k+1)!} \int_{d-1}^{t} \sqrt{\lambda_{s,i}(\boldsymbol{\theta}_i)}dW_i(s)\sqrt{\lambda_{t,i}(\boldsymbol{\theta}_i)}dW_i(t)\right)^{\top}_{i=1,\ldots,r} \text{ a.s.} \end{aligned}$$

The exponential of $\boldsymbol{\beta}$ is given by the power series

$$e^{\boldsymbol{\beta}} = \sum_{k=0}^{\infty} \frac{\boldsymbol{\beta}^k}{k!}.$$

Using this definition, we have

$$\begin{aligned} &\sum_{k=0}^{\infty} \boldsymbol{\beta}^k \frac{p\boldsymbol{\omega} + \{\mathrm{diag}(\boldsymbol{\gamma}) + (k+1)\mathbf{I}_r\}\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta})}{(k+2)!} \\ &= p\boldsymbol{\beta}^{-2}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r - \boldsymbol{\beta}\right)\boldsymbol{\omega} + \boldsymbol{\beta}^{-2}\left(e^{\boldsymbol{\beta}} - \mathbf{I}_r - \boldsymbol{\beta}\right)(\mathrm{diag}(\boldsymbol{\gamma}) - \mathbf{I}_r)\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta}) \end{aligned}$$

$$
\begin{aligned}
&+\boldsymbol{\beta}^{-1}\left(e^{\boldsymbol{\beta}}-\mathbf{I}_r\right)\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta}) \\
&= p\boldsymbol{\beta}^{-2}\left(e^{\boldsymbol{\beta}}-\mathbf{I}_r-\boldsymbol{\beta}\right)\boldsymbol{\omega}+\varrho\boldsymbol{\lambda}_{d-1}(\boldsymbol{\theta}) \text{ a.s.}
\end{aligned}
$$

(b). Since $\mathbf{h}_d(\boldsymbol{\theta})$ is adapted to $\mathcal{F}_{d-1}$ and $\mathbf{D}_d$ is a martingale difference, the statement is immediately showed. ■

6.2. Proof of Theorem 3.1

Proof of Theorem 3.1. For the simplicity, we omit the subscript d. Without loss of the generality, we assume that $sign(\langle\widehat{\mathbf{q}}_i,\mathbf{q}_i\rangle)=1$ and eigenvectors $\mathbf{q}_i$'s are constants. First we consider (3.4). We have for $i\in\{1,\ldots,r\}$,

$$
\begin{aligned}
\widehat{\xi}_i-\xi_i &= \mathbf{q}_i^\top(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\mathbf{q}_i+(\widehat{\mathbf{q}}_i^\top\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{q}}_i-\mathbf{q}_i^\top\widehat{\boldsymbol{\Gamma}}\mathbf{q}_i)+(\mathbf{q}_i^\top\boldsymbol{\Gamma}\mathbf{q}_i-\xi_i) \\
&= (a)+(b)+(c).
\end{aligned}
$$

For (c), we have

$$
|(c)|=\mathbf{q}_i^\top\boldsymbol{\Sigma}\mathbf{q}_i\le\|\boldsymbol{\Sigma}\|_2\le\|\boldsymbol{\Sigma}\|_1\le\max_{1\le i\le p}\sum_{j=1}^p|\Sigma_{ij}|^\delta(\Sigma_{ii}\Sigma_{jj})^{(1-\delta)/2}\le M_\sigma s_p.
$$

Then we have

$$
\mathrm{E}\{|(c)|^\alpha\}\le Cs_p^\alpha. \tag{6.1}
$$

For (b), we have

$$
\begin{aligned}
|(b)| &\le |(\mathbf{q}_i-\widehat{\mathbf{q}}_i)^\top\widehat{\boldsymbol{\Gamma}}(\mathbf{q}_i-\widehat{\mathbf{q}}_i)|+2|\widehat{\xi}_i\widehat{\mathbf{q}}_i^\top(\widehat{\mathbf{q}}_i-\mathbf{q}_i)| \\
&\le 2\|\widehat{\boldsymbol{\Gamma}}\|_2\|\mathbf{q}_i-\widehat{\mathbf{q}}_i\|_2^2 \\
&\le C\|\widehat{\boldsymbol{\Gamma}}\|_F\frac{\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Psi}\|_2^2}{(\xi_i-\xi_{i+1})^2} \\
&\le C\|\widehat{\boldsymbol{\Gamma}}\|_F\frac{\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_F^2+M_\sigma^2s_p^2}{p^2},
\end{aligned} \tag{6.2}
$$

where the third inequality is due to the Davis–Khan's sine theorem. Then we have

$$
\begin{aligned}
\mathrm{E}\{|(b)|^\alpha\} &\le Cp^{-2\alpha}\,\mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}\|_F^\alpha\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_F^{2\alpha}+\|\widehat{\boldsymbol{\Gamma}}\|_F^\alpha M_\sigma^{2\alpha}s_p^{2\alpha}\right) \\
&\le Cp^{-2\alpha}\left\{\mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}\|_F^{3\alpha}\right)^{1/3}\mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_F^{3\alpha}\right)^{2/3}+\mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}\|_F^{3\alpha}\right)^{1/3}\mathrm{E}\left(M_\sigma^{3\alpha}\right)^{2/3}s_p^{2\alpha}\right\} \\
&\le Cp^{-2\alpha}\left(p^{3\alpha}m^{-\alpha/2}+p^\alpha s_p^{2\alpha}\right) \\
&\le C\left(p^\alpha m^{-\alpha/2}+s_p^{2\alpha}p^{-\alpha}\right),
\end{aligned} \tag{6.3}
$$

where the second and third inequalities are due to the Hölder's inequality and (3.3), respectively.

For (a), similar to the proofs of Theorem 1 (Kim and Wang, 2016a), we can show

$$
\mathrm{E}\left(|(a)|^\alpha\right)\le Cp^\alpha m^{-\alpha/4}. \tag{6.4}
$$

By (6.1), (6.3) and (6.4), we have

$$
\mathrm{E}\left(|\widehat{\xi}_i-\xi_i|^\alpha\right)\le Cp^\alpha\{m^{-\alpha/4}+(s_p/p)^\alpha\}.
$$

Consider (3.5). Similar to the proof of (6.2), we have

$$
\begin{aligned}
\mathrm{E}\left(\|\widehat{\mathbf{q}}_{d,i}-sign(\langle\widehat{\mathbf{q}}_{d,i},\mathbf{q}_{d,i}\rangle)\mathbf{q}_{d,i}\|_2^{3\alpha}\right) &\le Cp^{-3\alpha}\left\{\mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_F^{3\alpha}\right)+\mathrm{E}\left(M_\sigma^{3\alpha}\right)s_p^{3\alpha}\right\} \\
&\le C\{m^{-3\alpha/4}+(s_p/p)^{3\alpha}\},
\end{aligned}
$$

where the last inequality is due to (3.3). ■

6.3. Proof of Theorem 3.2

Define

$$
\begin{aligned}
\widehat{L}_{n,m}(\boldsymbol{\theta}) &= \sum_{i=1}^r\widehat{L}_{n,m,i}(\boldsymbol{\theta})=-\frac{1}{2n}\sum_{i=1}^r\sum_{d=1}^n\log(\widehat{\phi}_{d,i}(\boldsymbol{\theta}))+\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}=-\frac{1}{2n}\sum_{i=1}^r\sum_{d=1}^n\widehat{l}_{d,i}(\boldsymbol{\theta}), \\
L_n(\boldsymbol{\theta}) &= \sum_{i=1}^rL_{n,i}(\boldsymbol{\theta})=-\frac{1}{2n}\sum_{i=1}^r\sum_{d=1}^n\log(\phi_{d,i}(\boldsymbol{\theta}))+\frac{\phi_{d,i}(\boldsymbol{\theta}_0)}{\phi_{d,i}(\boldsymbol{\theta})}=-\frac{1}{2n}\sum_{i=1}^r\sum_{d=1}^nl_{d,i}(\boldsymbol{\theta}),
\end{aligned}
$$

$$\widehat{\psi}_{n,m}(\boldsymbol{\theta}) = \frac{\partial \widehat{L}_{n,m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \psi_n(\boldsymbol{\theta}) = \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

We denote derivatives of any function g at x_0 by

$$\frac{\partial g(x_0)}{\partial x} = \frac{\partial g(x)}{\partial x}\Big|_{x=x_0}.$$

For any given random variable X and constant $c \geq 1$, define $\|X\|_{L_c} = \left\{E\left(|X|^c\right)\right\}^{1/c}$. Let $0 < \rho < 1$ be a generic constant whose values are free of $\boldsymbol{\theta}$, n and m and may change from occurrence to occurrence. By Assumption 1(a), without loss of generality, we assume that there are non-negative constants $\omega_{i,l}, \omega_{i,u}, \gamma_{i,l}, \gamma_{i,u}, \mu_{i,l}, \mu_{i,u}, \beta_{i,1,l}, \beta_{i,1,u}, \ldots, \beta_{i,r,l}, \beta_{i,r,u}$ such that $\omega_{i,l} < \omega_i < \omega_{i,u}, \gamma_{i,l} < \gamma_i < \gamma_{i,u} < 1, \mu_{i,l} < \mu_i < \mu_{i,u}, \beta_{i,1,l} < \beta_{i,1} < \beta_{i,1,u}, \ldots, \beta_{i,r,l} < \beta_{i,r} < \beta_{i,r,u}$

Lemma 6.1. *Under the assumptions of Theorem 3.1, we have*

$$\mathrm{E}\left(\max_{\mu_{i,l}\leq\mu_i\leq\mu_{i,u}} |\widehat{\kappa}_{d,i}(\mu_i) - \kappa_{d,i}(\mu_i)|^{3\alpha/2}\right) \leq Cp^{3\alpha/2}\left(m^{-3\alpha/8} + (s_p/p)^{3\alpha/4}\right).$$

Proof of Lemma 6.1. Without loss of the generality, we assume that $sign(\langle \widehat{\mathbf{q}}_{d,i}, \mathbf{q}_{d,i}\rangle) = 1$. We have

$$\begin{aligned}
&\left|\widehat{\kappa}_{d,i}(\mu_i) - \sum_{l=1}^{m}\{\mathbf{q}_{d,i}^\top(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)) - (t_{d,l} - d + 1)\sqrt{p}\mu_i\}^2 \Delta_{d,l}\right| \\
&\leq \sum_{l=1}^{m} |(\widehat{\mathbf{q}}_{d,i} - \mathbf{q}_{d,i})^\top(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1))| \\
&\qquad \times \left\{|(\widehat{\mathbf{q}}_{d,i} + \mathbf{q}_{d,i})^\top(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1))| + |2(t_{d,l} - d + 1)\sqrt{p}\mu_i|\right\} \Delta_{d,l} \\
&\leq 2\sum_{l=1}^{m} \|\widehat{\mathbf{q}}_{d,i} - \mathbf{q}_{d,i}\|_2 \|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2 \left\{\|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2 + \sqrt{p}\mu_{i,u}\right\} \Delta_{d,l} \\
&\leq C\|\widehat{\mathbf{q}}_{d,i} - \mathbf{q}_{d,i}\|_2 \sum_{l=1}^{m}\{\|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2^2 + \sqrt{p}\|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2\}\Delta_{d,l},
\end{aligned}$$

and thus,

$$\begin{aligned}
&\mathrm{E}\left[\left|\widehat{\kappa}_{d,i}(\mu_i) - \sum_{l=1}^{m}\{\mathbf{q}_{d,i}^\top(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)) - (t_{d,l} - d + 1)\sqrt{p}\mu_i\}^2 \Delta_{d,l}\right|^{3\alpha/2}\right] \\
&\leq C\,\mathrm{E}\left[\|\widehat{\mathbf{q}}_{d,i} - \mathbf{q}_{d,i}\|_2^{3\alpha}\right]^{1/2} \left(\mathrm{E}\left[\left|\sum_{l=1}^{m} \|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2^2 \Delta_{d,l}\right|^{3\alpha}\right]^{1/2}\right. \\
&\qquad \left. + \mathrm{E}\left[\left|\sum_{l=1}^{m} p^{1/2}\|\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)\|_2 \Delta_{d,l}\right|^{3\alpha}\right]^{1/2}\right) \\
&\leq Cp^{3\alpha/2}\left(m^{-3\alpha/8} + (s_p/p)^{3\alpha/2}\right),
\end{aligned} \tag{6.5}$$

where the second and last inequalities are due to the Hölder's inequality and Theorem 3.1, respectively.

Simple algebra shows

$$\begin{aligned}
&\sum_{l=1}^{m}\{\mathbf{q}_{d,i}^\top(\mathbf{Y}(t_{d,l}) - \mathbf{Y}(d-1)) - (t_{d,l} - d + 1)\sqrt{p}\mu_i\}^2 \Delta_{d,l} - \kappa_{d,i}(\mu_i) \\
&= \sum_{l=1}^{m}\{\mathbf{q}_{d,i}^\top(\mathbf{X}(t_{d,l}) - \mathbf{X}(d-1)) - (t_{d,l} - d + 1)\sqrt{p}\mu_i\}^2 \Delta_{d,l} - \kappa_{d,i}(\mu_i) \\
&+ \sum_{l=1}^{m}\{\mathbf{q}_{d,i}^\top\boldsymbol{\varepsilon}(t_{d,l})\}^2 \Delta_{d,l} + 2\sum_{l=1}^{m} \mathbf{q}_{d,i}^\top\boldsymbol{\varepsilon}(t_{d,l})(\mathbf{X}(t_{d,l}) - \mathbf{X}(d-1))^\top\mathbf{q}_{d,i}\Delta_{d,l} \\
&- 2\sum_{l=1}^{m} \mathbf{q}_{d,i}^\top\boldsymbol{\varepsilon}(t_{d,l})(t_{d,l} - d + 1)\sqrt{p}\mu_i\Delta_{d,l} \\
&= (I) + (II) + 2(III) + 2(IV).
\end{aligned}$$

Consider (I). For $t_{d,l-1} \le t \le t_{d,l}$, we have

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathbf{q}_{d,i}^{\top}\left\{\mathbf{X}(t_{d,l})-\mathbf{X}(d-1)\right\}-\mathbf{q}_{d,i}^{\top}\{\mathbf{X}(t)-\mathbf{X}(d-1)\}+\mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t)-\mathbf{u}(d-1)\}\right|^{6\alpha}\right] \\
&\le \mathrm{E}\left[\left|\mathbf{q}_{d,i}^{\top}\left\{\mathbf{X}(t_{d,l})-\mathbf{X}(t)\right\}+\mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t)-\mathbf{u}(d-1)\}\right|^{6\alpha}\right] \\
&\le C\left(\mathrm{E}\left[\left\{\int_{t}^{t_{d,l}}\|\varsigma(t)\|_2 dt\right\}^{3\alpha}\right]+\mathrm{E}\left[\left\{\int_{d-1}^{t_{d,l}}\|\boldsymbol{\sigma}^{\top}(t)\boldsymbol{\sigma}(t)\|_2 dt\right\}^{3\alpha}\right]+m^{-6\alpha}p^{3\alpha}\right) \\
&\le C\left(m^{-3\alpha}p^{3\alpha}+s_p^{3\alpha}\right),
\end{aligned} \tag{6.6}$$

where the second inequality is due to the Burkholder's inequality, and

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathbf{q}_{d,i}^{\top}\left\{\mathbf{X}(t_{d,l})-\mathbf{X}(d-1)\right\}+\mathbf{q}_{d,i}^{\top}\{\mathbf{X}(t)-\mathbf{X}(d-1)\}-\mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t)-\mathbf{u}(d-1)\}\right|^{6\alpha}\right] \\
&\le Cp^{3\alpha}.
\end{aligned}$$

Then

$$\mathrm{E}\{|(I)|^{3\alpha}\} \le Cp^{3\alpha}\left(m^{-3\alpha/2}+(s_p/p)^{3\alpha/2}\right). \tag{6.7}$$

For (II), by the application of the Burkholder's inequality, we can show

$$\mathrm{E}\{|(II)|^{3\alpha}|\mathbf{q}_{d,i}\} \le Cm^{-1}\sum_{l=1}^{m}\mathrm{E}\left[\{\mathbf{q}_{d,i}^{\top}\boldsymbol{\varepsilon}(t_{d,l})\}^{6\alpha}|\mathbf{q}_{d,i}\right] \le C \text{ a.s.} \tag{6.8}$$

For (III), we have

$$\begin{aligned}
\mathrm{E}\{|(III)|^{3\alpha}|\mathbf{X}\} &\le Cm^{-3\alpha/2-1}\sum_{l=1}^{m}\mathrm{E}\{|\mathbf{q}_{d,i}^{\top}\boldsymbol{\varepsilon}(t_{d,l})|^{3\alpha}|\mathbf{X}\}\left|(\mathbf{X}(t_{d,l})-\mathbf{X}(d-1))^{\top}\mathbf{q}_{d,i}\right|^{3\alpha} \\
&\le Cm^{-3\alpha/2-1}\sum_{l=1}^{m}\left|(\mathbf{X}(t_{d,l})-\mathbf{X}(d-1))^{\top}\mathbf{q}_{d,i}\right|^{3\alpha} \text{ a.s.},
\end{aligned}$$

where the first inequality is due to the Burkholder's inequality, and

$$\begin{aligned}
\mathrm{E}\{|(III)|^{3\alpha}\} &\le Cm^{-3\alpha/2-1}\sum_{l=1}^{m}\mathrm{E}\left\{\left|(\mathbf{X}(t_{d,l})-\mathbf{X}(d-1))^{\top}\mathbf{q}_{d,i}\right|^{3\alpha}\right\} \\
&\le Cm^{-3\alpha/2}p^{3\alpha/2},
\end{aligned} \tag{6.9}$$

where the last inequality can be derived similar to the proof of (6.6). Similarly, we can show

$$\mathrm{E}\{|(IV)|^{3\alpha}\} \le Cm^{-3\alpha/2}p^{3\alpha/2}. \tag{6.10}$$

Combining (6.5) and (6.7)–(6.10), we have

$$\mathrm{E}\left(\max_{\mu_{i,l}\le\mu_i\le\mu_{i,u}}|\widehat{\kappa}_{d,i}-\kappa_{d,i}|^{3\alpha/2}\right) \le Cp^{3\alpha/2}\left(m^{-3\alpha/8}+(s_p/p)^{3\alpha/4}\right). \quad \blacksquare$$

Lemma 6.2. *Under the assumptions of Theorem 3.1, we have*

$$\begin{aligned}
&\mathrm{E}\Bigg(\max_{\mu_{l,l}\le\mu_l\le\mu_{l,u}}\Bigg|\left[\mathbf{q}_{d,l}^{\top}\{\mathbf{X}(d)-\mathbf{X}(d-1)\}-\mathbf{q}_{d,l}^{\top}\{\mathbf{u}(d)-\mathbf{u}(d-1)\}-\sqrt{p}\mu_l\right]^2 \\
&\qquad\qquad -\left[\widehat{\mathbf{q}}_{d,l}^{\top}\{\mathbf{X}(d)-\mathbf{X}(d-1)\}-\sqrt{p}\mu_l\right]^2\Bigg|^{3\alpha/2}\Bigg) \\
&\le Cp^{3\alpha/2}\left(m^{-3\alpha/8}+(s_p/p)^{3\alpha/4}\right).
\end{aligned}$$

Proof of Lemma 6.2. Similar to the proofs of Lemma 6.1, we can show the statement. ■

Lemma 6.3. *Under the assumptions of Theorem 3.2, we have*

$$\max_{1\le i\le r}\max_{1\le d\le n}\mathrm{E}\left\{\sup_{\boldsymbol{\theta}\in\Theta}|\widehat{\phi}_{d,i}(\boldsymbol{\theta})-\phi_{d,i}(\boldsymbol{\theta})|\right\} \le Cp\left\{m^{-1/4}+(s_p/p)^{1/2}\right\}.$$

Proof of Lemma 6.3. By the compactness of Θ and Lemmas 6.1 and 6.2, the statement is showed. ■

Lemma 6.4. *Under the assumptions of Theorem 3.2, we have*

$$\sup_{\theta\in\Theta} |\widehat{L}_{n,m,i}(\boldsymbol{\theta}) - L_{n,i}(\boldsymbol{\theta})| = O_p\left(m^{-1/4} + (s_p/p)^{1/2}\right) \text{ for } i = 1,\dots,r.$$

Proof of Lemma 6.4. Since $p^{-1}\phi_{d,i}(\boldsymbol{\theta})$ and $p^{-1}\widehat{\phi}_{d,i}(\boldsymbol{\theta})$ stay way from zero, we have

$$\begin{aligned}
&|\widehat{L}_{n,m,i}(\boldsymbol{\theta}) - L_{n,i}(\boldsymbol{\theta})| \\
&\le \frac{1}{2n}\sum_{d=1}^{n}\left[\left|\log\left\{\widehat{\phi}_{d,i}(\boldsymbol{\theta})/\phi_{d,i}(\boldsymbol{\theta})\right\}\right| + \left|\frac{\widehat{\xi}_{d,i} - \phi_{d,i}(\boldsymbol{\theta}_0)}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\right|\right. \\
&\qquad\qquad \left. + \phi_{d,i}(\boldsymbol{\theta}_0)\left|\frac{\phi_{d,i}(\boldsymbol{\theta}) - \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})\phi_{d,i}(\boldsymbol{\theta})}\right|\right] \\
&\le Cn^{-1}p^{-1}\sum_{d=1}^{n}\left\{\left|\widehat{\xi}_{d,i} - \phi_{d,i}(\boldsymbol{\theta}_0)\right| + p^{-1}\phi_{d,i}(\boldsymbol{\theta}_0)\left|\phi_{d,i}(\boldsymbol{\theta}) - \widehat{\phi}_{d,i}(\boldsymbol{\theta})\right|\right\} \text{ a.s.}
\end{aligned}$$

Then the statement is showed from Theorem 3.1 and Lemma 6.3. ■

Lemma 6.5. *Under the assumptions of Theorem 3.2, we have*

(a) for $n \ge 2 + 2r$, $-\triangledown\psi_n(\boldsymbol{\theta}_0)$ is almost surely a positive definite matrix;
(b) $\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial^3\widehat{l}_{d,i}(\boldsymbol{\theta})}{\partial b_j\partial b_k\partial b_l}\right| = O_p(1)$ for all $i = 1,\dots,r$, $d = 1,\dots,n$, and any $j,k,l \in \{1,\dots,3r+r^2\}$, where $\boldsymbol{\theta} = (b_1,\dots,b_{3r+r^2}) = (\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_r)$.

Proof of Lemma 6.5. (a). Simple algebraic manipulations show

$$-\triangledown\psi_n(\boldsymbol{\theta}_0) = \frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\frac{\partial\phi_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\phi_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}^{\top}\phi_{d,i}(\boldsymbol{\theta}_0)^{-2} = \frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\phi_{\boldsymbol{\theta},d,i}\phi_{\boldsymbol{\theta},d,i}^{\top},$$

where $\phi_{\boldsymbol{\theta},d,i} = \frac{\partial\phi_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\phi_{d,i}(\boldsymbol{\theta}_0)^{-1}$. First, we suppose that $-\triangledown\psi_n(\boldsymbol{\theta}_0)$ is not a positive definite matrix. Then there is some non-zero constant vector $\mathbf{a} \in \mathbb{R}^{3r+r^2}$ such that

$$\frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\mathbf{a}^{\top}\phi_{\boldsymbol{\theta},d,i}\phi_{\boldsymbol{\theta},d,i}^{\top}\mathbf{a} = 0 \text{ a.s.},$$

which implies that

$$\phi_{\boldsymbol{\theta},d,i}^{\top}\mathbf{a} = 0 \text{ a.s. for all } d = 1,\dots,n,\ i = 1,\dots,r.$$

Since $\mathbf{X}(t)$ is nondegenerate, for $n \ge 2 + 2r$, the vector $\mathbf{a}$ should be the zero vector in order to satisfy the above equation, which contradicts $\mathbf{a} \neq 0$. ■

(b). Simple algebras show

$$\begin{aligned}
&\frac{\partial^3\widehat{l}_{d,i}(\boldsymbol{\theta})}{\partial b_j\partial b_k\partial b_l} \\
&= \left\{1 - \frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial^3\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_j\partial b_k\partial b_l}\right\} \\
&+ \left\{2\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} - 1\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_j}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial^2\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_k\partial b_l}\right\} \\
&+ \left\{2\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} - 1\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_k}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial^2\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_j\partial b_l}\right\} \\
&+ \left\{2\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} - 1\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_l}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial^2\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_j\partial b_k}\right\} \\
&+ \left\{2 - 6\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_j}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_k}\right\}\left\{\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})}\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_l}\right\}.
\end{aligned} \tag{6.11}$$

To handle (6.11), we first need to find bounds for derivatives of $\widehat{\phi}_{d,i}(\boldsymbol{\theta})$. The parameters related to $\widehat{\phi}_{d,i}(\boldsymbol{\theta})$ are only $\boldsymbol{\theta}_i$ and $\mu_1,\dots,\mu_r$, and so we investigate the derivatives corresponding to them. First, we investigate the first derivatives. Since

$\widehat{\phi}_{d,i}(\boldsymbol{\theta})$ is a linear function for ω_i and $\beta_{i,1}, \ldots, \beta_{i,r}$, we obtain that

$$\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \frac{\partial \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial b_l} \leq C \text{ for } b_l \in \{\omega_i, \beta_{i,1}, \ldots, \beta_{i,r}\}.$$

For γ_i, using the fact that $x/(1+x) \leq x^s$ for all $x \geq 0$ and any $s \in [0, 1]$, we have

$$\begin{aligned}
&\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \frac{\partial \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial \gamma_i} \\
&= \frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \left\{ \frac{\widehat{\lambda}_{d-1,i}(\boldsymbol{\theta})}{2} + \frac{\gamma_i + 1}{2} \left[\sum_{k=1}^{d-2} k \gamma_i^{k-1} \left\{ p\omega_i + \sum_{l=1}^{r} \beta_{i,l} \left(\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i \right)^2 \right\} \right.\right. \\
&\qquad \left.\left. + (d-1)\gamma_i^{d-2} \lambda_{0,i}(\boldsymbol{\theta}) + \gamma_i^{d-1} \frac{\partial \lambda_{0,i}(\boldsymbol{\theta})}{\partial \gamma_i} \right] \right\} \\
&\leq C + C \sum_{k=1}^{d-2} k \frac{\gamma_i^k \left\{ \omega_i + p^{-1} \sum_{l=1}^{r} \beta_{i,l} \left(\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i \right)^2 \right\}}{\omega_i + \gamma_i^k \left\{ \omega_i + p^{-1} \sum_{l=1}^{r} \beta_{i,l} \left(\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i \right)^2 \right\}} \\
&\leq C + C \sum_{k=1}^{d-2} k \gamma_i^{ks} \left\{ \omega_i + p^{-1} \sum_{l=1}^{r} \beta_{i,l} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i)^2 \right\}^s \\
&\leq C + C \sum_{k=1}^{d-2} k \gamma_{i,u}^{ks} \left\{ \omega_{i,u} + p^{-1} \sum_{l=1}^{r} \beta_{i,l,u} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i)^2 \right\}^s \\
&\leq C + C \sum_{k=1}^{d-2} k \rho^{ks} \left(p^{-1} \sum_{l=1}^{r} \beta_{i,l,u} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i)^2 \right)^s,
\end{aligned}$$

where $\widehat{Z}_{d,l} = \widehat{\mathbf{q}}_{d,l}^{\top}(\mathbf{X}(d) - \mathbf{X}(d-1))$. Then, for any $c \geq 1$, choose $s \in [0, 1]$ such that $\left\| \left(p^{-1} \sum_{l=1}^{r} \beta_{i,l,u} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i)^2 \right)^s \right\|_{L_c} < \infty$, and by Minkowski's inequality, we obtain

$$\left\| \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \frac{\partial \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial \gamma_i} \right| \right\|_{L_c} \leq C + C \sum_{k=1}^{d-2} k \rho^{ks} \left\| \left(p^{-1} \sum_{l=1}^{r} \beta_{i,l,u} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_i)^2 \right)^s \right\|_{L_c}.$$

Since $|\rho| < 1$,

$$\sup_{1 \leq i \leq r, 1 \leq d \leq n} \left\| \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \frac{\partial \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial \gamma_i} \right| \right\|_{L_c} < \infty.$$

For μ_l, we have

$$\begin{aligned}
\left| \frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \frac{\partial \widehat{\phi}_{d,i}(\boldsymbol{\theta})}{\partial \mu_l} \right| &\leq C \frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \left[p + \beta_{i,l} \widehat{\kappa}_{d,l}(\mu_l) + \sum_{k=1}^{d-2} \gamma_i^k \beta_{i,l} (\widehat{Z}_{d-k-1,l} - \sqrt{p}\mu_l)^2 \right] \\
&\leq C.
\end{aligned}$$

Similarly, the bounds for the second and third derivatives can be found.

Since $p^{-1}\widehat{\phi}_{d,i}(\boldsymbol{\theta})$ stays away from zero,

$$\sup_{\boldsymbol{\theta} \in \Theta} \frac{\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta})} \leq C p^{-1} \widehat{\xi}_{d,i} = O_p(1).$$

Combining these results,

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{\partial^3 \widehat{l}_{d,i}(\boldsymbol{\theta})}{\partial b_j \partial b_k \partial b_l} \right| = O_p(1). \quad \blacksquare$$

Proposition 6.1. *Under the assumptions of Theorem 3.2, we have*

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\max} = o_p(1).$$

Proof of Proposition 6.1. First, we show that there is a unique maximizer of $L_n(\boldsymbol{\theta})$. By the definition of $L_n(\boldsymbol{\theta})$, we have

$$\max_{\boldsymbol{\theta}\in\Theta} L_n(\boldsymbol{\theta}) \le -\frac{1}{2n}\sum_{i=1}^{r}\sum_{d=1}^{n}\min_{\boldsymbol{\theta}\in\Theta}\left\{\log\phi_{d,i}(\boldsymbol{\theta})+\frac{\phi_{d,i}(\boldsymbol{\theta}_0)}{\phi_{d,i}(\boldsymbol{\theta})}\right\}.$$

Thus, the maximizer $\boldsymbol{\theta}^*$ should satisfy that $\phi_{d,i}(\boldsymbol{\theta}^*)=\phi_{d,i}(\boldsymbol{\theta}_0)$ for $d=1,\ldots,n$ and $i=1,\ldots,r$. Suppose that there exists $\boldsymbol{\theta}^*\neq\boldsymbol{\theta}_0$ such that $\boldsymbol{\theta}^*$ satisfies that $\phi_{d,i}(\boldsymbol{\theta}^*)=\phi_{d,i}(\boldsymbol{\theta}_0)$ for $d=1,\ldots,n$ and $i=1,\ldots,r$.

Simple algebra shows

$$\lambda_{d,i}(\boldsymbol{\theta})=p\omega_i+\gamma_i\frac{2\{\phi_{d,i}(\boldsymbol{\theta})-\sum_{l=1}^{r}\beta_{i,l}\kappa_{d,l}(\mu_l)\}-p\omega_i}{\gamma_i+1}+\sum_{l=1}^{r}\beta_{i,l}(Z_{d,l}+\sqrt{p}\mu_{0,l}-\sqrt{p}\mu_l)^2$$

and

$$\begin{aligned}
\phi_{d+1,i}(\boldsymbol{\theta}) &= \frac{p\omega_i+(\gamma_i+1)\lambda_{d,i}(\boldsymbol{\theta})}{2}+\sum_{l=1}^{r}\beta_{i,l}\,\kappa_{d+1,l}(\mu_l)\\
&= p\omega_i+\gamma_i\phi_{d,i}(\boldsymbol{\theta})+\sum_{l=1}^{r}\frac{(\gamma_i+1)\beta_{i,l}}{2}(Z_{d,l}+\sqrt{p}\mu_{0,l})^2+p\sum_{l=1}^{r}\frac{(\gamma_i+1)\beta_{i,l}\mu_l^2}{2}\\
&\quad -2\sqrt{p}\sum_{l=1}^{r}\frac{(\gamma_i+1)\beta_{i,l}\mu_l}{2}(Z_{d,l}+\sqrt{p}\mu_{0,l})\\
&\quad -\sum_{l=1}^{r}\gamma_i\beta_{i,l}\kappa_{d,l,1}-\sum_{l=1}^{r}\mu_l\gamma_i\beta_{i,l}\kappa_{d,l,2}-\frac{p}{3}\sum_{l=1}^{r}\mu_l^2\gamma_i\beta_{i,l}\\
&\quad +\sum_{l=1}^{r}\beta_{i,l}\,\kappa_{d+1,l,1}+\sum_{l=1}^{r}\beta_{i,l}\mu_l\,\kappa_{d+1,l,2}+\frac{p}{3}\sum_{l=1}^{r}\beta_{i,l}\mu_l^2\\
&= \gamma_i\phi_{d,i}(\boldsymbol{\theta})+\sum_{l=1}^{r}\beta_{i,l}\left\{(Z_{d,l}+\sqrt{p}\mu_{0,l})^2/2+\kappa_{d+1,l,1}\right\}\\
&\quad +\sum_{l=1}^{r}\gamma_i\beta_{i,l}\left\{(Z_{d,l}+\sqrt{p}\mu_{0,l})^2/2-\kappa_{d,l,1}\right\}+p\left\{\omega_i+\sum_{l=1}^{r}(\gamma_i+5)\beta_{i,l}\mu_l^2/6\right\}\\
&\quad -\sum_{l=1}^{r}\beta_{i,l}\mu_l(\sqrt{p}Z_{d,l}+p\mu_{0,l}-\kappa_{d+1,l,2})\\
&\quad -\sum_{l=1}^{r}\gamma_i\beta_{i,l}\mu_l(\sqrt{p}Z_{d,l}+p\mu_{0,l}+\kappa_{d,l,2}),
\end{aligned}$$

where $\kappa_{d,i}(\mu_i)=\kappa_{d,i,1}+\kappa_{d,i,2}\mu_i+\frac{p}{3}\mu_i^2$,

$$\kappa_{d,i,1}(\mu_i)=\int_{d-1}^{d}\left[\mathbf{q}_{d,i}^{\top}\{\mathbf{X}(t)-\mathbf{X}(d-1)\}-\mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t)-\mathbf{u}(d-1)\}\right]^2dt,$$

$$\kappa_{d,i,2}(\mu_i)=-2\sqrt{p}\int_{d-1}^{d}(t-d+1)\left[\mathbf{q}_{d,i}^{\top}\{\mathbf{X}(t)-\mathbf{X}(d-1)\}-\mathbf{q}_{d,i}^{\top}\{\mathbf{u}(t)-\mathbf{u}(d-1)\}\right]dt.$$

Then, similar to the proofs of Theorem 1 in Kim (2016), for $n\ge 2+4r$, we can show that $\boldsymbol{\theta}^*$ should be the same as $\boldsymbol{\theta}_0$ to satisfy $\phi_{d,i}(\boldsymbol{\theta}^*)=\phi_{d,i}(\boldsymbol{\theta}_0)$ for $d=1,\ldots,n$ and $i=1,\ldots,r$. Therefore, $\boldsymbol{\theta}_0$ is the unique maximizer of $L_n(\boldsymbol{\theta})$. Then, since $L_n(\boldsymbol{\theta})$ is the continuous function with respect to $\boldsymbol{\theta}$, for any given $\varepsilon>0$, there is a constant $c>0$ such that

$$L_n(\boldsymbol{\theta}_0)-\max_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_{\max}\ge\varepsilon}L_n(\boldsymbol{\theta})>c \text{ a.s.} \tag{6.12}$$

Then the statement is the consequence of Theorem 1 in Xiu (2010), Lemma 6.4, and (6.12). ■

Proof of Theorem 3.2. By Taylor expansion and the mean value theorem, there exist $\widetilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$ such that

$$\widehat{\psi}_{n,m}(\widehat{\boldsymbol{\theta}})-\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)=-\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)=\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0).$$

If $-\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})\xrightarrow{p}-\triangledown\psi_n(\boldsymbol{\theta}_0)$ which is a positive definite matrix by Lemma 6.5(a), then we can conclude that the convergence rate of $(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ is the same as that of $\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)$. First we show $-\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})\xrightarrow{p}-\triangledown\psi_n(\boldsymbol{\theta}_0)$. Define

$$\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta})=(\widehat{H}_{i,j}(\boldsymbol{\theta}))_{i,j=1,\ldots,3r+r^2}\quad\text{and}\quad\triangledown\psi_n(\boldsymbol{\theta})=(H_{i,j}(\boldsymbol{\theta}))_{i,j=1,\ldots,3r+r^2}.$$

We have

$$\left\|\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})-\triangledown\psi_n(\boldsymbol{\theta}_0)\right\|_{\max} \le \left\|\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})-\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)\right\|_{\max} + \left\|\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)-\triangledown\psi_n(\boldsymbol{\theta}_0)\right\|_{\max}.$$

For $\left\|\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})-\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)\right\|_{\max}$, we have, for some $\bar{\boldsymbol{\theta}}$ between $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$,

$$\begin{aligned}\left|\widehat{H}_{i,j}(\widetilde{\boldsymbol{\theta}})-\widehat{H}_{i,j}(\boldsymbol{\theta}_0)\right| &= \left|\frac{\partial\widehat{H}_{i,j}(\bar{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right| \\ &\le C\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial\widehat{H}_{i,j}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|_{\max}\left\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right\|_{\max} \\ &\le o_p(1),\end{aligned}$$

where the first equality is due to the mean value theorem and Taylor expansion, and the last inequality is from Proposition 6.1 and Lemma 6.5(b).

For $\left\|\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)-\triangledown\psi_n(\boldsymbol{\theta}_0)\right\|_{\max}$, by Lemmas 6.1–6.3, Theorem 3.1, and Proposition 6.1, we can show

$$\left\|\triangledown\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)-\triangledown\psi_n(\boldsymbol{\theta}_0)\right\|_{\max} = o_p(1).$$

Thus, $\left\|\triangledown\widehat{\psi}_{n,m}(\widetilde{\boldsymbol{\theta}})-\triangledown\psi_n(\boldsymbol{\theta}_0)\right\|_{\max} \xrightarrow{p} 0$.

Now, we consider $\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)$. We have

$$\begin{aligned}\|\widehat{\psi}_{n,m}(\boldsymbol{\theta}_0)\|_{\max} &\le \frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\left\|\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\left(\frac{1}{\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)}-\frac{\widehat{\xi}_{d,i}}{\widehat{\phi}^2_{d,i}(\boldsymbol{\theta}_0)}\right)\right\|_{\max} \\ &\le \frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\left\|\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)^{-1}\right\|_{\max}\left|\frac{\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)-\widehat{\xi}_{d,i}}{\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)}\right| \\ &\le O_p\left(m^{-1/4}+(s_p/p)^{1/2}\right)\frac{1}{2n}\sum_{d=1}^{n}\sum_{i=1}^{r}\left\|\frac{\partial\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\widehat{\phi}_{d,i}(\boldsymbol{\theta}_0)^{-1}\right\|_{\max} \\ &\le O_p\left(m^{-1/4}+(s_p/p)^{1/2}\right),\end{aligned}$$

where the third inequality is due to Theorem 3.1 and Lemma 6.3, and the last inequality can be derived similar to the proofs of Lemma 6.5(b). ■

6.4. Proof of Theorem 4.1

Proof of Theorem 4.1. The statements (4.5) and (4.6) are immediately showed by Theorem 4.1 (Fan and Kim, 2018).

We consider (4.7). We have

$$\begin{aligned}\|\widetilde{\boldsymbol{\Gamma}}_{n+1}-\mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)\|_{\Gamma^*} &\le \|\widehat{\boldsymbol{\Sigma}}_n-\boldsymbol{\Sigma}_n\|_{\Gamma^*}+\|\widehat{\boldsymbol{\Psi}}_{n+1}-\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^{\top}\|_{\Gamma^*} \\ &\le \|\widehat{\boldsymbol{\Psi}}_{n+1}-\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^{\top}\|_{\Gamma^*}+O_p\left(s_p\tau_m^{1-\delta}\right),\end{aligned}$$

where the last inequality is due to (4.5). Now, it is enough to show

$$\|\widehat{\boldsymbol{\Psi}}_{n+1}-\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^{\top}\|_{\Gamma^*} = O_p(p^{1/2}m^{-1/2}+s_p/p^{1/2}+s_p\tau_m^{1-\delta}).$$

We have

$$\begin{aligned}&\|\widehat{\boldsymbol{\Psi}}_{n+1}-\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_{n,i}\mathbf{q}_{n,i}^{\top}\|_F \\ &\le \|\sum_{i=1}^{r}\{\widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}})-h_{n+1,i}(\boldsymbol{\theta}_0)\}\widehat{\mathbf{q}}_{n,i}\widehat{\mathbf{q}}_{n,i}^{\top}\|_F+\|\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\left(\mathbf{q}_{n,i}\mathbf{q}_{n,i}^{\top}-\widehat{\mathbf{q}}_{n,i}\widehat{\mathbf{q}}_{n,i}^{\top}\right)\|_F \\ &\le \sum_{i=1}^{r}|\widehat{h}_{n+1,i}(\widehat{\boldsymbol{\theta}})-h_{n+1,i}(\boldsymbol{\theta}_0)|\|\widehat{\mathbf{q}}_{n,i}\widehat{\mathbf{q}}_{n,i}^{\top}\|_F+\sum_{i=1}^{r}h_{n+1,i}(\boldsymbol{\theta}_0)\|\mathbf{q}_{n,i}\mathbf{q}_{n,i}^{\top}-\widehat{\mathbf{q}}_{n,i}\widehat{\mathbf{q}}_{n,i}^{\top}\|_F \\ &\le O_p\left(p\{m^{-1/4}+(s_p/p)^{1/2}\}\right),\end{aligned} \tag{6.13}$$

where the last inequality is due to Theorem 3.1 and (6.14). We have

$$\begin{aligned}\|e^{\widehat{\beta}} - e^{\beta_0}\|_2 &= \|e^{(\widehat{\beta}-\beta_0)+\beta_0} - e^{\beta_0}\|_2 \\ &\le \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 e^{\|\boldsymbol{\beta}_0\|_2} e^{\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2} \\ &\le O_p\left(m^{-1/4} + (s_p/p)^{1/2}\right),\end{aligned}$$

where the last inequality is due to Theorem 3.2, and by Theorem 3.2 and Lemma 6.2, we can show

$$\|\widehat{\boldsymbol{\lambda}}_{n-1}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{\lambda}_{n-1}(\boldsymbol{\theta}_0)\|_2 = O_p\left(p\{m^{-1/4} + (s_p/p)^{1/2}\}\right),$$

which imply, together with Theorem 3.2,

$$\|\widehat{\mathbf{h}}_{n+1}(\widehat{\boldsymbol{\theta}}) - \mathbf{h}_{n+1}(\boldsymbol{\theta}_0)\|_2 \le O_p\left(p\{m^{-1/4} + (s_p/p)^{1/2}\}\right). \tag{6.14}$$

Similar to the proofs of Theorem 4.1 (Fan and Kim, 2018), we can show

$$\begin{aligned}\|\widehat{\boldsymbol{\Psi}}_{n+1} - \sum_{i=1}^{r} h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^\top\|_{\boldsymbol{\Gamma}^*} &\le C\Big[\{p^{-3/2} + p^{-2}\widehat{h}_{n+1,1}(\widehat{\boldsymbol{\theta}})\}\|\widehat{\boldsymbol{\Psi}}_{n+1} - \sum_{i=1}^{r} h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^\top\|_F \\ &\qquad + p^{-5/2}\widehat{h}_{n+1,1}(\widehat{\boldsymbol{\theta}})\|\widehat{\boldsymbol{\Psi}}_{n+1} - \sum_{i=1}^{r} h_{n+1,i}(\boldsymbol{\theta}_0)\mathbf{q}_i\mathbf{q}_i^\top\|_F^2\Big] \\ &\le O_p\left(m^{-1/4} + p^{1/2}m^{-1/2} + s_p/p^{1/2}\right),\end{aligned}$$

where the last inequality is due to (6.13) and (6.14). Therefore, we have

$$\|\widetilde{\boldsymbol{\Gamma}}_{n+1} - \mathrm{E}\left(\boldsymbol{\Gamma}_{n+1}|\mathcal{F}_n\right)\|_{\boldsymbol{\Gamma}^*} = O_p\left(p^{1/2}m^{-1/2} + s_p/p^{1/2} + s_p\tau_m^{1-\delta}\right). \quad \blacksquare$$

Acknowledgments

The research of Donggyu Kim was supported in part by KAIST Settlement/Research Subsidies for Newly-hired Faculty, South Korea grant G04170049 and KAIST Basic Research Funds by Faculty, South Korea (A0601003029). The research of Jianqing Fan was supported in part by National Science Foundation, United States grant DMS-1406266 and DMS-1712591 and a Princeton Engineering Innovation Fund, United States. The bulk of the work was conducted while Donggyu Kim was a postdoctoral fellow at Department of Operations Research and Financial Engineering, Princeton University.

References

Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous financial data. J. Amer. Statist. Assoc. 105 (492), 1504–1517.
Aït-Sahalia, Y., Xiu, D., 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. J. Econometrics 201 (2), 384–399.
Barigozzi, M., Hallin, M., 2016. Generalized dynamic factor models and volatilities: recovering the market volatility shocks. Econom. J. 19 (1).
Barigozzi, M., Hallin, M., 2017. Generalized dynamic factor models and volatilities: estimation and forecasting. J. Econometrics 201 (2), 307–321.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. Econometrica 76 (6), 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 (2), 149–169.
Bibinger, M., Hautsch, N., Malec, P., Reiß, M., et al., 2014. Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. Ann. Statist. 42 (4), 1312–1346.
Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? J. Econometrics 132 (1), 169–194.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 (3), 307–327.
Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. Rev. Econom. Stat. 498–505.
Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. J. Polit. Econ. 96 (1), 116–131.
Candès, E.J., Li, X., Ma, Y., Wright, J., 2011. Robust principal component analysis? J. ACM 58 (3), 11.
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 (1), 116–133.
Connor, G., Korajczyk, R.A., Linton, O., 2006. The common and specific components of dynamic volatility. J. Econometrics 132 (1), 231–255.
Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: a multivariate latent factor arch model. J. Appl. Econom. 4 (1), 1–21.
Engle, R., 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econom. Statist. 20 (3), 339–350.
Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized arch. Econometric Theory 11 (1), 122–150.
Fan, J., Furger, A., Xiu, D., 2016a. 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.
Fan, J., Kim, D., 2018. Robust high-dimensional volatility matrix estimation for high-frequency factor model. J. Am. Stat. Assoc. 113 (523), 1268–1283.
Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 603–680.
Fan, J., Wang, W., Zhong, Y., 2016b, An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. arXiv preprint arXiv:1603.03516.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: identification and estimation. Rev. Econom. Stat. 82 (4), 540–554.

Forni, M., Hallin, M., Lippi, M., Zaffaroni, P., 2015. Dynamic factor models with infinite-dimensional factor spaces: one-sided representations. J. Econometrics 185 (2), 359–371.

Harvey, A., Ruiz, E., Sentana, E., 1992. Unobserved component time series models with arch disturbances. J. Econometrics 52 (1–2), 129–157.

Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: quantitative asset pricing implications. J. Financ. Econ. 119 (2), 249–283.

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 (7), 2249–2276.

Jacod, J., Protter, P.E., 2011. Discretization of Processes, vol. 67. Springer Science & Business Media.

Kim, D., 2016. Statistical inference for unified garch–itô Models with High-Frequency Financial Data. J. Time Series Anal. 37 (4), 513–532.

Kim, D., Liu, Y., Wang, Y., 2018. Large volatility matrix estimation with factor-based diffusion model for high-frequency financial data. Bernoulli 24 (4B), 3657–3682.

Kim, D., Wang, Y., 2016a. Sparse pca based on high-dimensional itô processes with measurement errors. J. Multivariate Anal. 152, 172–189.

Kim, D., Wang, Y., 2016b. Unified discrete-time and continuous-time models and statistical inferences for merged low-frequency and high-frequency financial data. J. Econometrics 194 (2), 220–230.

Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. Stochastic Process. Appl. 126, 3527–3577.

Mandelbrot, B., 1963. The variation of certain speculative prices. J. Bus. 36 (4), 394–419.

Ng, V., Engle, R.F., Rothschild, M., 1992. A multi-dynamic-factor model for stock returns. J. Econometrics 52 (1–2), 245–266.

Rangel, J.G., Engle, R.F., 2012. The factor–spline–garch model for high and low frequency correlations. J. Bus. Econom. Statist. 30 (1), 109–124.

Sentana, E., Calzolari, G., Fiorentini, G., 2008. Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks. J. Econometrics 146 (1), 10–25.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Am. Stat. Assoc. 97 (460), 1167–1179.

Tao, M., Wang, Y., Chen, X., 2013a. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. Econometric Theory 29 (04), 838–856.

Tao, M., Wang, Y., Zhou, H.H., et al., 2013b. Optimal sparse volatility matrix estimation for high-dimensional itô processes with measurement errors. Ann. Statist. 41 (4), 1816–1864.

Van der Weide, R., 2002. GO-GARCH: a multivariate generalized orthogonal garch model. J. Appl. Econometrics 17 (5), 549–564.

Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12 (6), 1019–1043.

Zhang, L., 2011. Estimating covariation: epps effect, microstructure noise. J. Econometrics 160 (1), 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 (472), 1394–1411.