A Factor-GARCH Model for High Dimensional Volatilities

Xiaoling Li[a,c] Yuan Li[a] Jiazhu Pan[b] Xingfa Zhang[a*]

[a]School of Economics and Statistics, Guangzhou Unversity, Guangzhou 510006, China

[b]Department of Mathematics & Statistics, University of Strathclyde, Glasgow G1 1XH, UK

[c]College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

Abstract

This paper proposes a method for modelling volatilities (conditional covariance matrices) of high dimensional dynamic data. We combine the ideas of approximate factor models for dimension reduction and multivariate GARCH models to establish a model to describe the dynamics of high dimensional volatilities. Sparsity condition and thresholding technique are applied to the estimation of the error covariance matrices, and quasi maximum likelihood estimation (QMLE) method is used to estimate the parameters of the common factor conditional covariance matrix. Asymptotic theories are developed for the proposed estimation. Monte Carlo simulation studies and real data examples are presented to support the methodology.

Keywords: Approximate factor models; conditional variance-covariance matrix; multivariate GARCH; sparse estimation; thresholding.

MSC2010 subject classifications:: Promary 62H25, 62M10; secondary 62F12, 62H12.

1 Introduction

In the last decade, studies on modelling high-dimensional time series and space-time data have become a hot research area in statistics. The reason is that the current big-data environment creates new challenges in modeling such big dynamic data in many scientific fields, including engineering, environmentology, psychology, economics and finance. In theory, vector autoregressive moving-average (VARMA) models are often used for multivariate time series. However, even in the moderate dimension case, people often encounter the difficulties of over-parametrization and

*Corresponding author, email address: xingfazhang@gzhu.edu.cn

identifiability problem. To overcome these issues, various methods have been developed to reduce the dimension of data and the number of parameters. One of effective methods for dimension reduction is the factor analysis for time series such as Pan and Yao[23], Lam, Yao and Bathia[16], Lam and Yao[17] and Gao and Tsay[13]. But the above results are focused on the conditional mean (the first conditional moment). There are also many papers investigating the covariance matrix estimation for high dimensional data, see e.g., Bickel and Levina[1, 2] Cai and Liu[3] and Cai and Zhou[4]. To avoid over-parametrization of covariance matrices, some restrictions on parameters are often put in the models. Sparsity is one of the commonly used assumptions in high dimensional modelling. For example, Fan, Liao and Mincheva[8, 9] considered the estimation of covariance matrices for high-dimensional time series with a factor structure. They assumed the majority of off-diagonal elements for error covariance matrix were zero or close to zero, and the number of these elements grown very slowly when the dimension became large.

However, the existing literature on high dimensional covariance matrix estimation is mainly restricted to the unconditional case. In financial application, e.g. the portfolio allocation in risk management, the conditional covariance matrices (so-called volatilities) are often used to describe the dynamical structure of risk, see e.g., Markowitz[21] and Markowitz[22]. Hence modelling high dimensional volatilities has become an important problem in statistics. Motivated by this point, Guo, Box and Zhang[14] proposed a dynamic structure and developed an estimation procedure for high-dimensional conditional covariance matrices. Their work is insightful but not applicable for the cases when the dimension p is equal to or larger than the sample size T. A feasible way to solve such problems is to use factor analysis to reduce the dimension first and then fit a lower dimensional volatility model to the common factors. Following this direction, in this paper, we propose a factor-GARCH model to study the high-dimensional volatilities which can be estimated even the dimension p is equal to or larger than the sample size T.

The rest of this paper is arranged as follows. We specify the proposed methodology in Section 2. Asymptotic properties of the proposed estimators including the necessary assumptions are given in Section 3. A portfolio allocation is constructed based on the formula for Markowitz's optimal portfolio in Section 4. Section 5 presents simulation experiments to show the finite sample behaviour of the estimation. In Section 6, we apply the methodology to a real data set consisting of 30 industry portfolios which are freely available from Kenneth French's web-site. Section 7 provides some conclusions. All proofs of theoretic results are put in the Appendix.

Through out the paper, we use A^τ, $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ to denote the transpose, the minimum and maximum eigenvalues of a matrix $\boldsymbol{A}$, respectively. Let $\boldsymbol{A}$ be a $q \times r$ matrix and denote by $\mathrm{vec}(\boldsymbol{A})$ the $qr \times 1$ vector formed by stacking the r columns of $\boldsymbol{A}$ underneath each other in the order from left to right. Notations $\|\boldsymbol{A}\|_F$, $\|\boldsymbol{A}\|$ and $\|\boldsymbol{A}\|_{\max}$ represent the Frobenius norm, spectral norm (operator norm) and elementwise norm of a matrix $\boldsymbol{A}$, defined respectively by $\|\boldsymbol{A}\|_F = \mathrm{tr}^{1/2}(\boldsymbol{A}^\tau \boldsymbol{A})$, $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}^\tau \boldsymbol{A})$ and $\|\boldsymbol{A}\|_{\max} = \max_{i,j} |a_{ij}|$. When $\boldsymbol{A}$ is a vector, both $\|\boldsymbol{A}\|_F$ and $\|\boldsymbol{A}\|$ are equal to the Euclidean norm.

2 Methodology

2.1 The factor-GARCH model

Let $\boldsymbol{y}_t = (y_{1t}, \cdots, y_{pt})^\tau$ be a p-dimensional time series. Consider the following approximate factor model

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{u}_t, \tag{1}$$

where $\boldsymbol{B} = (\boldsymbol{b}_1, \cdots \boldsymbol{b}_p)^\tau$, for $\boldsymbol{b}_i$ is an unobservable K-dimensional vector of factor loadings, $i = 1, \cdots, p$; $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{Kt})^\tau$ is an observable K-dimensional vector of common factors, and K is fixed; $\boldsymbol{u}_t = (u_{1t}, \cdots, u_{pt})^\tau$ is an unknown p-dimensional vector of idiosyncratic errors of $\boldsymbol{y}_t$, which is independent of $\{\boldsymbol{x}_{t-i}, i \geq 1\}$. We assume

$$\boldsymbol{\Sigma}_x(t) \equiv \mathrm{Var}(\boldsymbol{x}_t|\mathcal{F}_{t-1}) \equiv \boldsymbol{D}_t \boldsymbol{\Gamma} \boldsymbol{D}_t, \tag{2}$$

where $\mathcal{F}_{t-1}$ is the past information available up to time $t-1$, $\boldsymbol{D}_t = \mathrm{diag}(h_{1t}^{1/2}, \cdots, h_{Kt}^{1/2})$, and

$$\boldsymbol{\Gamma} = \left\{ \begin{array}{cccc} 1 & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & 1 & \cdots & \gamma_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{K1} & \gamma_{K2} & \cdots & 1 \end{array} \right\}. \tag{3}$$

Denote $\underline{\boldsymbol{h}}_t = (h_{1t}, \ldots, h_{Kt})^\tau$, and assume that

$$\underline{\boldsymbol{h}}_t = \boldsymbol{\varphi} + \sum_{i=1}^{r} \boldsymbol{\Phi}_i \underline{\boldsymbol{x}}_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \underline{\boldsymbol{h}}_{t-j}, \tag{4}$$

where $\underline{\boldsymbol{x}}_t^2 = (x_{1t}^2, \ldots, x_{Kt}^2)^\tau$, $\boldsymbol{\varphi}$ is a K-dimension vector, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_j$ are $K \times K$ matrices. The true parameter vector is denoted by $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}^\tau, \boldsymbol{\gamma}^\tau)^\tau$, where $\boldsymbol{\delta} = \mathrm{vec}(\boldsymbol{\varphi}, \boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_r, \boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_s)$, and

$\boldsymbol{\gamma} = (\gamma_{21}, \ldots, \gamma_{K1}, \gamma_{32}, \ldots, \gamma_{K2}, \ldots, \gamma_{KK-1})^\tau$, with $\gamma_{ij} = \gamma_{ji}$. The main feature of this model is that the conditional correlation

$$\gamma_{ij} = E(x_{it}x_{jt}|\mathcal{F}_{t-1})/\sqrt{E(x_{it}^2|\mathcal{F}_{t-1})E(x_{jt}^2|\mathcal{F}_{t-1})}$$

is constant over time, where $i \neq j$ and x_{it} is the ith element of $\boldsymbol{x}_t$. Ling and McAleer[18] has explained that it is possible to provide a straightforward explanation for the hypothesis of constant correlation. Thus we can write the common factor as follows

$$\boldsymbol{x}_t = \boldsymbol{\Sigma}_x^{1/2}(t)\boldsymbol{\eta}_t, \tag{5}$$

where $\boldsymbol{\Sigma}_x(t) = \boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t$, $\boldsymbol{\eta}_t \sim \mathrm{IID}(\mathbf{0}, \boldsymbol{I}_K)$ and $\boldsymbol{I}_K$ is an identity matrix.

Note that the conditional covariance matrix $\boldsymbol{\Sigma}_y(t)$ of $\boldsymbol{y}_t$ is given by

$$\boldsymbol{\Sigma}_y(t) = \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau + \boldsymbol{\Sigma}_u, \tag{6}$$

where $\boldsymbol{\Sigma}_u = (\sigma_{u,ij})_{p\times p}$ is the conditional covariance matrix of $\boldsymbol{u}_t$, which does not dependent on $\mathcal{F}_{t-1}$. The literature on approximate factor models typically assumes that the first K eigenvalues of $\boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau$ diverge at rate $O(p)$, whereas all the eigenvalues of $\boldsymbol{\Sigma}_u$ are bounded as $p \to \infty$. In addition, in this paper, we assume that $\boldsymbol{\Sigma}_u$ is approximately sparse, as in Bickel and Levina[1] and Rothman, Levina and Zhu[24]: for some $q \in [0, 1)$, define

$$m_p = \max_i \sum_j |\sigma_{u,ij}|^q,$$

and the sparsity assumption puts an upper bound restriction on m_p:

$$m_p^2 = o\left(\frac{T}{\log p}\right).$$

One of the distinguished features of our factor model is that $\boldsymbol{\Sigma}_y(t)$ is divided into two parts. One is dependent on t, and the other is not.

Bear in mind that the common factor is observable in our model. In the next two subsections, we are to use least square method to estimate $\boldsymbol{B}$ and a thresholding method to estimate the covariance matrix $\boldsymbol{\Sigma}_u$ in a similar way as Fan, Liao and Mincheva[8].

2.2 The least square estimation (LSE) of $\boldsymbol{B}$

To estimate the factor loading matrix $\boldsymbol{B}$ in the approximate factor model (1), denote $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)$, $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T)$ and $\boldsymbol{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_T)$. Then, model (1) can be written in a more compact form,

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X} + \boldsymbol{U}.$$

Then $\boldsymbol{B}$ can be estimated by the least square method. Denote the estimator of $\boldsymbol{B}$ by $\widehat{\boldsymbol{B}} = (\widehat{\boldsymbol{b}}_1, \widehat{\boldsymbol{b}}_2, \ldots, \widehat{\boldsymbol{b}}_p)^\tau$, where

$$\widehat{\boldsymbol{b}}_i = \arg\min_{\boldsymbol{b}_i} \frac{1}{Tp} \sum_{t=1}^{T} \sum_{i=1}^{p} (y_{it} - \boldsymbol{b}_i^\tau \boldsymbol{x}_t)^2,$$

then

$$\widehat{\boldsymbol{B}} = \boldsymbol{Y}\boldsymbol{X}^\tau (\boldsymbol{X}\boldsymbol{X}^\tau)^{-1}. \tag{7}$$

2.3 The thresholding estimation of $\boldsymbol{\Sigma}_u$

In the factor models, we do not observe the error term directly. Hence before estimating the error covariance matrix of the factor model, we need to construct a sample covariance matrix based on the residuals $\widehat{u}_{it}$ which can be obtained by the estimated factor loadings. Following the least square estimator (LSE), we have

$$\widehat{u}_{it} = y_{it} - \widehat{\boldsymbol{b}}_i^\tau \boldsymbol{x}_t.$$

Let $\widehat{\boldsymbol{u}}_t = (\widehat{u}_{1t}, \widehat{u}_{2t}, \ldots, \widehat{u}_{pt})^\tau$. We then construct the residual covariance matrix as

$$\widehat{\boldsymbol{\Sigma}}_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{\boldsymbol{u}}_t \widehat{\boldsymbol{u}}_t^\tau \equiv (\widehat{\sigma}_{u,ij})_{p \times p}.$$

Unfortunately, the above $\widehat{\boldsymbol{\Sigma}}_u$ is generally not a consistent estimator when dimension p is large than the sample size T. To obtain a consistent estimator, based on $\widehat{\boldsymbol{\Sigma}}_u$, we further apply the adaptive thresholding estimator introduced by Cai and Liu[3] to estimate the error covariance matrix, which is given by

$$\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} = (\widehat{\sigma}_{u,ij}^{\mathcal{T}}), \quad \widehat{\sigma}_{u,ij}^{\mathcal{T}} = \widehat{\sigma}_{u,ij} I(|\widehat{\sigma}_{u,ij}| \geq \sqrt{\widehat{\vartheta}_{ij}} \omega_T), \tag{8}$$

$$\widehat{\vartheta}_{ij} \equiv \frac{1}{T} \sum_{t=1}^{T} (\widehat{u}_{it}\widehat{u}_{jt} - \widehat{\sigma}_{u,ij})^2, \tag{9}$$

where ω_T is to be specified later.

2.4 The estimation of $\mathbf{\Sigma}_x(t)$

As $\boldsymbol{x}_t$ and K are known in this paper, we now give procedures to estimate the unknown parameters in the GARCH model: $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}^\tau, \boldsymbol{\gamma}^\tau)^\tau$. To capture the past information completely in the multivariate case, it is obvious that h_{kt} should contain some past information, not only from x_{kt} but also from x_{jt}. As a simple illustration, we assume, for $k = 1, 2, \ldots, K$

$$x_{kt} = h_{kt}\widetilde{\eta}_{kt}, \quad h_{kt} = \varphi_k + \sum_{i=1}^{r}\sum_{l=1}^{K}\alpha_{ikl}x_{lt-i}^2 + \sum_{j=1}^{s}\sum_{l=1}^{K}\beta_{jkl}h_{lt-i}, \tag{10}$$

where, $\varphi_k > 0$, $\alpha_{ikl}, \beta_{jkl} \geq 0$, and $\widetilde{\boldsymbol{\eta}}_t \equiv (\widetilde{\eta}_{1t}, \cdots, \widetilde{\eta}_{Kt})^\tau = \mathbf{\Gamma}^{1/2}\boldsymbol{\eta}_t$, $\boldsymbol{\eta}_t \sim \mathrm{IID}(\mathbf{0}, \boldsymbol{I}_K)$.

Remark: In fact, Equation (10) is the same as Equation (4), e.g. φ_k is the k component of $\boldsymbol{\varphi}$, and α_{ikl} and β_{jkl} are the (k,l) entry of $\mathbf{\Phi}_i$ and $\mathbf{\Psi}_j$ respectively, for $1 \leq k, j \leq K$, $i = 1, \cdots, r$ and $j = 1, \cdots, s$.

The estimators of the parameters in model (10) are obtained by maximizing

$$L_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{D}_t\mathbf{\Gamma}\boldsymbol{D}_t| - \frac{1}{2}\boldsymbol{x}_t^\tau(\boldsymbol{D}_t\mathbf{\Gamma}\boldsymbol{D}_t)^{-1}\boldsymbol{x}_t, \tag{11}$$

with initial values being

$$\underline{\boldsymbol{x}}_0^2 = \cdots = \underline{\boldsymbol{x}}_{1-r}^2 = \underline{\boldsymbol{h}}_0 = \cdots = \underline{\boldsymbol{h}}_{1-s} = \boldsymbol{\varphi}.$$

Francq and Zakoïan[11] has shown that the choice of the initial values is unimportant for the asymptotic properties of the QMLE, but making sure that the conditional variance is positive. Then, a QMLE of $\boldsymbol{\theta}$ is defined by:

$$\widehat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\mathbf{\Theta}} L_T(\boldsymbol{\theta}). \tag{12}$$

Once the parameters in $\boldsymbol{\theta}$ have been estimated, by substituting them into (10), we can obtain an estimator $\widehat{h}_{kt}$ of h_{kt}, an estimated matrix $\widehat{\mathbf{\Gamma}}$ of $\mathbf{\Gamma}$, and hence an estimator $\widehat{\mathbf{\Sigma}}_x(t)$ of $\mathbf{\Sigma}_x(t)$.

3 Theoretical properties

In this section, we are going to present the asymptotic properties of the proposed estimators described in Section 2 when T, $p \to \infty$.

First, we need to impose the strong mixing conditions to conduct asymptotic analysis of the least square estimates. Let $\mathcal{F}_{u,-\infty}^0$ and $\mathcal{F}_{u,T}^\infty$ denote the σ-algebra that are generated by $\{\boldsymbol{u}_t : t \leq 0\}$

and $\{\boldsymbol{u}_t : t \geq T\}$ respectively. Define the mixing coefficient

$$\alpha_u(T) = \sup_{\boldsymbol{A}\in\mathcal{F}^0_{u,-\infty}, \boldsymbol{B}\in\mathcal{F}^\infty_{u,T}} |P(\boldsymbol{A})P(\boldsymbol{B}) - P(\boldsymbol{AB})|. \tag{13}$$

Assumption 1. $\{\boldsymbol{u}_t\}_{t\geq 1}$ is stationary and ergodic, and α-mixing with geometric rate $\alpha_u(T) \leq \exp\{-cT^{\nu_0}\}$, for $\nu_0 > 0$, and $c > 0$. In addition, $\{\boldsymbol{u}_t\}_{t\geq 1}$ and $\{\boldsymbol{x}_t\}_{t\geq 1}$ are independent.

This assumption is standard in approximate factor models literature (see Fan, Liao and Mincheva[8, 9]), allowing $\{\boldsymbol{u}_t\}_{t\geq 1}$ to be weakly dependent. Note that, under Assumption 4 defined later, $\{\boldsymbol{x}_t\}_{t\geq 1}$ is strictly stationary and ergodic, and α-mixing with geometric rate (see Lindner[19]). And in Lemma A.2, we will prove that the α-mixing coefficient defined on $\{(\boldsymbol{x}_t, \boldsymbol{u}_t)\}_{t\geq 1}$ has geometric rate.

Assumption 2.

(a). There are constants $0 < a_1 < a_2$, and $a_3 < a_4$, such that $a_1 < \lambda_{\min}(\boldsymbol{\Sigma}_u) \leq \lambda_{\max}(\boldsymbol{\Sigma}_u) < a_2$, and $a_3 < \mathrm{Var}(u_{it}u_{jt}) < a_4$ for all $1 \leq i, j \leq p$.

(b). There exists a constant $M > 0$, for all i and j, such that $|b_{ij}| < M$.

(c). There are $\nu_1 > 0$ and $b_1 > 0$ such that, for any $y > 0$ and $i \leq p$,

$$P(|u_{it}| > y) \leq \exp\{-(y/b_1)^{\nu_1}\}.$$

Condition (a) requires the non-singularity of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_u^{-1}$, and allows the idiosyncratic components to be weakly dependent. Condition (c) allows us to apply the Bernstein-type inequality for the weakly dependent data. This assumption is standard in the approximate factor models as in Fan, Liao and Mincheva[8].

Assumption 3. $\|p^{-1}\boldsymbol{B}^\tau\boldsymbol{B} - \boldsymbol{\Omega}\| = o(1)$ for some $K \times K$ symmetric positive definite matrix $\boldsymbol{\Omega}$ such that $\lambda_{\min}(\boldsymbol{\Omega})$ is bounded away from zero.

Assumption 3 allows that the factors to be pervasive, that is, impact every individual time series (e.g. Harding[15], Fan, Liao and Mincheva[8, 9]). To establish the results for estimating conditional covariance matrix of $\boldsymbol{\Sigma}_x(t)$, we introduce more assumption

Assumption 4.

Let $\mathcal{A}_\theta(z) = \sum_{i=1}^r \boldsymbol{\Phi}_i z^i$ and $\mathcal{B}_\theta(z) = \boldsymbol{I}_K - \sum_{j=1}^s \boldsymbol{\Psi}_j z^j$. By convention, $\mathcal{A}_\theta(z) = 0$ if $r = 0$, and $\mathcal{B}_\theta(z) = \boldsymbol{I}_K$ if $s = 0$. Let $\ell(\boldsymbol{\Phi}_0)$ denote the top Lyapunov coefficient of the sequence of matrices $\boldsymbol{\Phi}_0 = (\boldsymbol{\Phi}_{0t})$ as defined in (10.41) of Francq and Zakoïan[11].

(a). The parameter space $\boldsymbol{\Theta}$ is a compact subspace of Euclidean space, such that $\boldsymbol{\theta}_0$ is an interior point in $\boldsymbol{\Theta}$.

(b). $\ell(\boldsymbol{\Phi}_0) < 0$ and, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\det \mathcal{B}(z) = 0 \Rightarrow |z| > 1$.

(c). The components of $\boldsymbol{\eta}_t$ are independent and their squares have non-degenerate distributions. And there exist $\nu_3 > 0$ with $3\nu_3^{-1} + \nu_2^{-1} > 1$ and $b_3 > 0$ such that for any $y > 0$ and $i \leq K$,

$$P(|\eta_{it}| > y) \leq \exp(-(y/b_3)^{\nu_3}),$$

(d). If $s > 0$, then $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left coprime and $M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0})$ has full rank K, where $M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0}) = \begin{bmatrix} \boldsymbol{\Phi}_r & \boldsymbol{\Psi}_s \end{bmatrix}$.

(e). $\boldsymbol{\Gamma}$ is a positive definite correlation matrix for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(f). $\mathrm{E}\|\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{\tau}\|^2 < \infty$.

The above assumption is standard in CCC-GARCH models as in Francq and Zakoïan[11]. Condition (a) is a common condition for GARCH models. Under Condition (b), $\{\boldsymbol{x}_t\}_{t \geq 1}$ is strictly stationary and ergodic. We can apply the Bernstein-type inequality for the weakly dependent data based on Condition (c). Condition (d) makes that $\mathcal{A}_{\theta_0}$ and $\mathcal{B}_{\theta_0}$ have no common roots. $\boldsymbol{\Sigma}_x(t)$ is positive definite under Condition (e). If Condition (f) is satisfied, $\{\boldsymbol{x}_t\}_{t \geq 1}$ has finite κ-order comments, where κ will be defined in Theorem 1 later.

Then, we are to state asymptotic theory for model estimator. It can be seen from Equation (6) that the asymptotic properties of $\widehat{\Sigma}_y(t)$ depend on the estimators for $\Sigma_x(t)$ and Σ_u, respectively. Hence, we firstly give asymptotic results for the estimation of $\Sigma_x(t)$ and Σ_u. Following Theorem 10.8 of Francq and Zokoïan[11], we have

Lemma 1. Let $\widehat{\boldsymbol{\theta}}_T$ be the QMLE from (12). Then, under Assumption 4,

$$\widehat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0, \quad \text{almost surely when } T \to \infty$$

Lemma 1 ensures that the estimator of $\boldsymbol{\theta}_0$ is strong consistent. And the following theorem establishes the convergence rates of $\boldsymbol{\Sigma}_x(t)$.

Theorem 1. Suppose $0 < 2\epsilon < \kappa/2 - 2$, and $\kappa > 4$. Under the same assumptions as in Lemma 1, there exists $C > 0$, such that

$$P\left(\left\|\widehat{\boldsymbol{\Sigma}}_x(t) - \boldsymbol{\Sigma}_x(t)\right\|_F^2 > \frac{C \log T}{T}\right) = O\left(\frac{1}{T^{1+\epsilon}}\right).$$

In Lemma 2, we will construct the asymptotic properties of the thresholding estimator $\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}$ based on observation with estimation errors, where $\nu^{-1} = 1.5\nu_1^{-1} + 1.5\nu_3^{-1} + \nu_2^{-1}$, ν_1, ν_2 and ν_3 are defined in Assumptions 2 and 4, and Lemma A.2, respectively.

Lemma 2. Suppose Assumptions 1-4 hold, and let $(\log p)^{6/\nu - 1} = o(T)$. Then there exist $C_1 > 0$ and $C_2 > 0$ in the adaptive thresholding estimator (7) with

$$\omega_T = C_1\bigg(\sqrt{\frac{\log p}{T}}\bigg),$$

such that

$$P\bigg(\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\| \leq C_2 m_p\sqrt{\frac{\log p}{T}}\bigg) \geq 1 - O\Big(\frac{1}{p^2} + \frac{1}{T^2}\Big).$$

Further, if $\omega_T m_p = o(1)$, then with probability at least $1 - O(\frac{1}{p^2} + \frac{1}{T^2})$,

$$\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}) \geq 0.5\lambda_{\min}(\boldsymbol{\Sigma}_u),$$

and

$$\bigg\|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\bigg\| \leq C_2 m_p\sqrt{\frac{\log p}{T}}.$$

Now we are to state the asymptotic theory for $\widehat{\boldsymbol{\Sigma}}_y(t)$. Following Equations (6)-(8) and (12), it is easy to obtain

$$\widehat{\boldsymbol{\Sigma}}_y(t) = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^{\tau} + \widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}. \tag{14}$$

To evaluate the accuracy of an estimator $\widehat{\boldsymbol{A}}$ of a matrix $\boldsymbol{A}$ of size p, the entropy loss norm proposed by James and Stein[12] is often applied. Its formula is

$$\|\widehat{\boldsymbol{A}} - \boldsymbol{A}\|_{\Sigma} = p^{-1/2}\|\boldsymbol{A}^{-1/2}(\widehat{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{A}^{-1/2}\|_F.$$

Then the following theorem establishes the asymptotic properties of the estimator for $\boldsymbol{\Sigma}_y(t)$ based on several norms.

Theorem 2. Under the assumptions of Lemma 2 and Theorem 1, there exists a constant $C_3 > 0$, and we have

(i)

$$P\bigg(\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\|_{\Sigma}^2 \leq \frac{C_3 p(\log p)^2}{T^2} + \frac{C_3 m_p^2 \log p}{T}\bigg) = 1 - O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big),$$

$$P\left(\left\|\widehat{\boldsymbol{\Sigma}}_y(t)-\boldsymbol{\Sigma}_y(t)\right\|_{\max}^2 \leq \frac{C_3\log p + C_3\log T}{T}\right) = 1-O\left(\frac{1}{p^{1+\epsilon}}+\frac{1}{T^{1+\epsilon}}\right).$$

(ii) If $m_p\sqrt{\frac{\log p}{T}}=o(1)$, with probability at least $1-O(\frac{1}{p^{1+\epsilon}}+\frac{1}{T^{1+\epsilon}})$,

$$\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_y(t)) \geq 0.5\lambda_{\min}(\boldsymbol{\Sigma}_u),$$

and

$$\left\|\widehat{\boldsymbol{\Sigma}}_y^{-1}(t)-\boldsymbol{\Sigma}_y^{-1}(t)\right\| \leq C_3 m_p\sqrt{\frac{\log p}{T}}.$$

Remark: In the proposed model, the number of factors K can be large, possibly growing with T at a certain rate. However, it is typically small compared to p and T in many applications, hence it is assumed to be fixed for simplicity.

4 Portfolio allocation

In this section, we will briefly introduce the estimated portfolio allocation based on proposed model. We denote $\boldsymbol{E}(\boldsymbol{y}_t|\mathcal{F}_{t-1})$ as the conditional expectation of $\boldsymbol{y}_t$, and take conditional expectation of (1), we have

$$\boldsymbol{E}(\boldsymbol{y}_t|\mathcal{F}_{t-1}) = \boldsymbol{B}\boldsymbol{E}(\boldsymbol{x}_t|\mathcal{F}_{t-1}).$$

Therefore, we use

$$\widehat{\boldsymbol{E}}(\boldsymbol{y}_t|\mathcal{F}_{t-1}) = \widehat{\boldsymbol{B}}\widehat{\boldsymbol{E}}(\boldsymbol{x}_t|\mathcal{F}_{t-1}) \tag{15}$$

to estimate $\boldsymbol{E}(\boldsymbol{y}_t|\mathcal{F}_{t-1})$, where $\boldsymbol{E}(\boldsymbol{x}_t|\mathcal{F}_{t-1})$ is estimated by using a VAR(1) model

$$\widehat{\boldsymbol{E}}(\boldsymbol{x}_t|\mathcal{F}_{t-1}) = \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\Xi}}\boldsymbol{x}_{t-1}.$$

Based on the mean-variance optimal portfolio by Markowitz[21, 22], we construct the estimated optimal portfolio allocation similar to Guo, Box and Zhang[14]: $\boldsymbol{w}$ denoted the allocation vector of p risky assets, to be held between times $t-1$ and t, is defined as the solution to

$$\begin{aligned}&\min_{\boldsymbol{w}} \boldsymbol{w}^\tau\boldsymbol{\Sigma}_y(t)\boldsymbol{w}\\ &\text{subject to } \boldsymbol{w}^\tau\mathbf{1}_p=1 \text{ and } \boldsymbol{w}^\tau\boldsymbol{E}(\boldsymbol{y}_t|\mathcal{F}_{t-1})=\pi,\end{aligned}$$

Table 1: Mean and covariance used to generate $\boldsymbol{b}$

$\boldsymbol{\mu}_{\boldsymbol{B}}$	$\boldsymbol{\Sigma}_{\boldsymbol{B}}$		
0.9219234	0.01313649	0.02523216	0.01722212
0.7589148	0.02523216	0.09039758	0.03606229
0.4904847	0.01722212	0.03606229	0.08802161

where π is the target return imposed on the portfolio, and $\mathbf{1}_p$ is a p-dimensional column vector of 1. The solution $\widehat{\boldsymbol{w}}$ is given by

$$\widehat{\boldsymbol{w}} = \frac{d_3 - d_2\pi}{d_1 d_3 - d_2^2}\widehat{\boldsymbol{\Sigma}}_y^{-1}(t)\mathbf{1}_p + \frac{d_1\pi - d_2}{d_1 d_3 - d_2^2}\widehat{\boldsymbol{\Sigma}}_y^{-1}(t)\widehat{\boldsymbol{E}}(\boldsymbol{y}_t|\mathcal{F}_{t-1}), \tag{16}$$

where

$$d_1 = \mathbf{1}_p^\tau \widehat{\boldsymbol{\Sigma}}_y^{-1}(t)\mathbf{1}_p, \quad d_2 = \mathbf{1}_p^\tau \widehat{\boldsymbol{\Sigma}}_y^{-1}(t)\widehat{\boldsymbol{E}}(\boldsymbol{y}_t|\mathcal{F}_{t-1}),$$
$$d_3 = \widehat{\boldsymbol{E}}^\tau(\boldsymbol{y}_t|\mathcal{F}_{t-1})\widehat{\boldsymbol{\Sigma}}_y^{-1}(t)\widehat{\boldsymbol{E}}^\tau(\boldsymbol{y}_t|\mathcal{F}_{t-1}).$$

5 Simulations

In this section, we use four examples to show how well the proposed estimation procedure works. The models for our simulation study are modified versions of the Fama-French three factor models described in Fan, Fan and Lv[7] and Fan, Liao and Mincheva[8]. The Fama-French three-factor model constructed by Fama and French[5, 6] has the form

$$y_{it} = b_{i1}x_{1t} + b_{i2}x_{2t} + b_{i3}x_{3t} + u_{it},$$

where y_{it} is the excess return of the ith industry's portfolio, $i = 1, 2, \cdots, p$, x_{1t}, x_{2t}, and x_{3t} are three observable common factors: market, size and value.

We adopt the over two-year daily data $(\check{\boldsymbol{y}}_t, \check{\boldsymbol{x}}_t)$ from Jun 1st, 2018 to Jun 26th, 2020 (T=524) of 30 industry portfolios to generate the factor loadings $\boldsymbol{B}$: obtain the LSE of $\boldsymbol{B}$ from $\check{\boldsymbol{y}}_t = \boldsymbol{B}\check{\boldsymbol{x}}_t + \boldsymbol{u}_t$, denoted $\check{\boldsymbol{B}} = (\check{\boldsymbol{b}}_1, \cdots, \check{\boldsymbol{b}}_{30})^\tau$, where $\check{\boldsymbol{b}}_i = (\check{b}_{i1}, \check{b}_{i2}, \check{b}_{i3})^\tau, i = 1, \ldots, 30$, and calculate the sample mean vector $\boldsymbol{\mu}_{\check{\boldsymbol{B}}}$ and sample covariance matrix $\boldsymbol{\Sigma}_{\check{\boldsymbol{B}}}$. The results are depicted in Table 1. Then, the factor loadings $\{\boldsymbol{b}_i\}_{i=1}^T$ can be drawn from normal distribution $N_3(\boldsymbol{\mu}_{\check{\boldsymbol{B}}}, \boldsymbol{\Sigma}_{\check{\boldsymbol{B}}})$.

For each $p \in [1, 30]$, we create the sparse matrix $\boldsymbol{\Sigma}_u = \boldsymbol{Q} + \boldsymbol{s}\boldsymbol{s}^\tau - \mathrm{diag}\{s_1^2, \cdots, s_p^2\}$ in the following way. Suppose $\boldsymbol{Q} = \mathrm{diag}\{\sigma_1^2, \cdots, \sigma_p^2\}$, where $\sigma_1^2, \cdots, \sigma_p^2$ are generated independently

Table 2: Mean and SD of the estimate errors in Example 1

$T = 500$	$\Vert\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\Vert_{\Sigma}$	$\Vert\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\Vert_{\max}$	$\Vert\widehat{\boldsymbol{\Sigma}}_y^{-1}(t) - \boldsymbol{\Sigma}_y^{-1}(t)\Vert$
$p = 40$	**0.168**(0.006)	**1.341**(0.639)	**2.200**(0.357)
$p = 100$	**0.136**(0.014)	**1.614**(0.939)	**1.932**(0.217)
$p = 300$	**0.157**(0.040)	**1.806**(1.225)	**1.897**(0.245)
$p = 540$	**0.192**(0.053)	**1.9014**(1.102)	**2.094**(0.179)

from the Gamma distribution $G(\zeta, \xi)$ with $\zeta = 4.204$, $\xi = 0.2227$. Create $\boldsymbol{s} = (s_1, \cdots, s_p)^\tau$ to be a sparse vector by setting that: each $s_i \sim N(0, 1)$ with probability $\frac{0.2}{\sqrt{p}\log p}$, and $s_i = 0$ otherwise. This leads to an average of $\frac{0.2\sqrt{p}}{\log p}$ nonzero elements per each row of the error covariance matrix. Create a loop that generates $\boldsymbol{\Sigma}_u$ multiple times until it is positive definite. Note that, we will let $\boldsymbol{\Sigma}_u = \boldsymbol{Q}$ in Examples 1 and 3, in other words, u_{it} is independent of u_{jt}, for $i \neq j$.

Now, for each fixed p, the following steps give detailed simulation procedures:

(1) Generate $\{\boldsymbol{b}_i\}_{i=1}^p$ independently from $N_3(\boldsymbol{\mu}_{\breve{\boldsymbol{B}}}, \boldsymbol{\Sigma}_{\breve{\boldsymbol{B}}})$, and set $\boldsymbol{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_p)^\tau$;

(2) Generate $\{\boldsymbol{u}_t\}_{t=1}^T$ independently from $N_p(\mathbf{0}, \boldsymbol{\Sigma}_u)$;

(3) Generate $\{\boldsymbol{x}_t\}_{t=1}^T$ from $N_K(\mathbf{0}, \boldsymbol{\Sigma}_x(t))$, where $\boldsymbol{\Sigma}_x(t)$ is given in different GARCH form in the corresponding example;

(4) Calculate $\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{u}_t$ for $t = 1, \cdots, T$;

In the following examples, the sample size is $T = 500$, the replication times is $N = 200$ and the dimensions are $p = 40$, 100, 300, 540, respectively. In addition, $\omega_T = 0.3\sqrt{\log p/T}$ is set to obtain the thresholding estimator $\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}$ in (8).

Example 1. $\boldsymbol{\Sigma}_u$ is a diagonal matrix, and $\boldsymbol{\Sigma}_x(t) = \mathrm{diag}(h_{1t}, h_{2t}, h_{3t})$, where

$$\begin{cases} h_{1t} = 0.02 + 0.3x_{1t-1}^2 + 0.3h_{1t-1}, \\ h_{2t} = 0.05 + 0.1x_{2t-1}^2 + 0.3x_{2t-2}^2 + 0.1h_{2t-1}, \\ h_{3t} = 0.001 + 0.2x_{3t-1}^2 + 0.1h_{3t-1} + 0.15h_{3t-2}. \end{cases}$$

Example 2. $\boldsymbol{\Sigma}_u$ is sparse, and $\boldsymbol{\Sigma}_x(t) = \mathrm{diag}(h_{1t}, h_{2t}, h_{3t})$, where

$$\begin{cases} h_{1t} = 0.002 + 0.3x_{1t-1}^2 + 0.1h_{1t-1}, \\ h_{2t} = 0.0007 + 0.1x_{2t-1}^2 + 0.3x_{2t-2}^2 + 0.1h_{2t-1}, \\ h_{3t} = 0.001 + 0.1x_{3t-1}^2 + 0.1h_{3t-1} + 0.2h_{3t-2}. \end{cases}$$

Table 3: Mean and SD of the estimate errors in Example 2

$T = 500$	$\|\widehat{\mathbf{\Sigma}}_y(t) - \mathbf{\Sigma}_y(t)\|_\Sigma$	$\|\widehat{\mathbf{\Sigma}}_y(t) - \mathbf{\Sigma}_y(t)\|_{\max}$	$\|\widehat{\mathbf{\Sigma}}_y^{-1}(t) - \mathbf{\Sigma}_y^{-1}(t)\|$
$p = 40$	**0.090**(0.201)	**0.354**(0.165)	**0.681**(0.962)
$p = 100$	**0.087**(0.042)	**0.475**(0.156)	**0.944**(0.548)
$p = 300$	**0.114**(0.032)	**0.583**(0.132)	**1.487**(0.413)
$p = 540$	**0.141**(0.046)	**0.669**(0.155)	**1.687**(0.222)

Example 3. Covariance matrix $\mathbf{\Sigma}_u$ is diagonal, and $\mathbf{\Sigma}_x(t) = \boldsymbol{D}_t\mathbf{\Gamma}\boldsymbol{D}_t$, where

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & 0 & 0.13 \\ 0 & 1 & 0.3 \\ 0.13 & 0.3 & 1 \end{pmatrix}, \quad \boldsymbol{D}_t = \begin{pmatrix} h_{1t}^{1/2} & 0 & 0 \\ 0 & h_{2t}^{1/2} & 0 \\ 0 & 0 & h_{3t}^{1/2} \end{pmatrix},$$

and

$$\begin{cases} h_{1t} = 0.008 + 0.2x_{1t-1}^2, \\ h_{2t} = 0.02 + 0.02x_{2t-1}^2 + 0.4h_{2t-1}, \\ h_{3t} = 0.01 + 0.4x_{3t-1}^2 + 0.1h_{3t-1}. \end{cases}$$

Example 4. $\mathbf{\Sigma}_u$ is a sparse matrix, and $\mathbf{\Sigma}_x(t) = \boldsymbol{D}_t\mathbf{\Gamma}\boldsymbol{D}_t$, where

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & 0.1 & 0 \\ 0.1 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{pmatrix}, \boldsymbol{D}_t = \begin{pmatrix} h_{1t}^{1/2} & 0 & 0 \\ 0 & h_{2t}^{1/2} & 0 \\ 0 & 0 & h_{3t}^{1/2} \end{pmatrix},$$

and

$$\begin{cases} h_{1t} = 0.008 + 0.01x_{1t-1}^2 + 0.4h_{1t-1}, \\ h_{2t} = 0.002 + 0.2x_{2t-1}^2 + 0.4h_{2t-1}, \\ h_{3t} = 0.001 + 0.1x_{3t-1}^2 + 0.2h_{3t-1}. \end{cases}$$

The above examples are different mainly in two aspects: In Examples 1 and 2, $\mathbf{\Sigma}_x(t)$ is diagonal, i.e. $\{\boldsymbol{x}_t\}_{t=1}^T$ are combined by K conditional independent sub-series, and each sub-series has univariate GARCH structure; Whereas $\{\boldsymbol{x}_t\}_{t=1}^T$ is a CCC-GARCH process in Examples 3 and 4, i.e. $\mathbf{\Sigma}_x(t) = \boldsymbol{D}_t\mathbf{\Gamma}\boldsymbol{D}_t$. On the other hand, the components of $\boldsymbol{u}_t$ are independent corresponding to different j in Examples 1 and 3, but they have cross-sectional correlation in Examples 2 and 4.

To describe the convergence of $\widehat{\mathbf{\Sigma}}_y(t)$ to $\mathbf{\Sigma}_y(t)$, we give the results on averages and standard deviations of $\sup_{1\leq t\leq T} \widehat{\mathbf{\Sigma}}_y(t) - \mathbf{\Sigma}_y(t)$ under the entropy-loss norm $\|\cdot\|_\Sigma$ and the element-wise norm $\|\cdot\|_{\max}$, and over $N = 200$ replications. We also give the averages and standard deviations of $\sup_{1\leq t\leq T} \widehat{\mathbf{\Sigma}}_y^{-1}(t) - \mathbf{\Sigma}_y^{-1}(t)$ under the operator norm.

Results. Tables 2-5 present the averages (in boldface) and standard deviations (in parentheses) of different errors of Examples 1-4. Based on the simulation results, we obtain the following observations:

(1). The standard deviations of the norms are relatively small when compared to their corresponding averages.

(2). Under the entropy-norm, the results of all examples are perfect. It can be seen that Examples 1 and 2 outperform Examples 3 and 4 respectively, and this is reasonable because the former two examples have simpler GARCH structures. The results are in line with Theorem 2, showing our method works well.

(3). Under the infinity norm, the means and standard deviations are generally larger than those based on entropy norm. This is due to the fact that thresholding affects mainly the elements of the conditional covariance matrix that are closest to 0, and the infinity norm depicts the magnitude of the largest element-wise absolute error. Such similar results are also obtained in Fan, Liao and Mincheva[8]. Moreover, Examples 2 and 4 perform better than Examples 1 and 3 respectively, implying the performances under infinity norm are also influenced by the complexity of $\mathbf{\Sigma}_x(t)$.

(4). Under the operator norm, it is seen that the means of error are slightly increased but the standard deviations are decreased in all examples. It is acceptable, as we are considering the estimate errors of $\widehat{\mathbf{\Sigma}}_y^{-1}(t)$, where extra approximation or computation is needed. These conclusions are in consistent with Theorem 2, and similar conclusions are also given in Fan, Liao and Mincheva[8]. It is also found in Example 4 that the performance under operator norm is influenced by the complexity of $\mathbf{\Sigma}_x(t)$.

In conclusion, the simulation results are in line with the asymptotic results obtained in Section 2, which indicates that our approach works well.

Table 4: Mean and SD of the estimate errors in Example 3

$T = 500$	$\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\|_{\Sigma}$	$\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\|_{\max}$	$\|\widehat{\boldsymbol{\Sigma}}_y^{-1}(t) - \boldsymbol{\Sigma}_y^{-1}(t)\|$
$p = 40$	**0.475**(0.188)	**0.830**(0.377)	**0.785**(0.216)
$p = 100$	**0.525**(0.162)	**1.100**(0.407)	**1.058**(0.191)
$p = 300$	**0.583**(0.196)	**1.306**(0.500)	**1.255**(0.113)
$p = 540$	**0.697**(0.304)	**1.527**(0.712)	**1.405**(0.112)

Table 5: Mean and SD of the estimate errors in Example 4

$T = 500$	$\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\|_{\Sigma}$	$\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\|_{\max}$	$\|\widehat{\boldsymbol{\Sigma}}_y^{-1}(t) - \boldsymbol{\Sigma}_y^{-1}(t)\|$
$p = 40$	**0.237**(0.045)	**0.425**(0.136)	**0.975**(0.205)
$p = 100$	**0.320**(0.050)	**0.514**(0.135)	**1.668**(0.238)
$p = 300$	**0.480**(0.049)	**0.677**(0.181)	**3.589**(0.262)
$p = 540$	**0.607**(0.048)	**0.746**(0.166)	**6.450**(0.671)

6 A real data example

In this section, we apply our model to analyze the real data in American stock market. We consider the data of daily returns of 30 Industrial Portfolios, which are available at Data-set (2021) (*http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*). The data series span from January 3rd, 2007 to December 31st, 2010 with a total of 1008 observations. Figure 1 depicts the time plots of the three factors (Rm-Rf, SMB and HML), denoted by $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, respectively. And let $\{\boldsymbol{x}_t\}_{t=1}^T$ be the vector time series consisted of these three factors. The plots in Figure 1 show clearly that there exist periods of large volatility during the 2008-2009 financial crisis. First, we use ADF (augmented Dickey Fuller) test method to test whether the time series is stationary. And the results are reported in Table 6. It is seen that all the series pass the stationary test. Then, we use the four statistics in Tsay[25] to test the conditional heteroscedasticity of $\{\boldsymbol{x}_t\}_{t=1}^T$, and the results are presented in Table 7. It is shown that all the tests suggest the presence of conditional heteroscedasticity in the data.

Hence, we use the CCC-GARCH model to fit the common factor $\boldsymbol{x}_t$, and the lag order in the

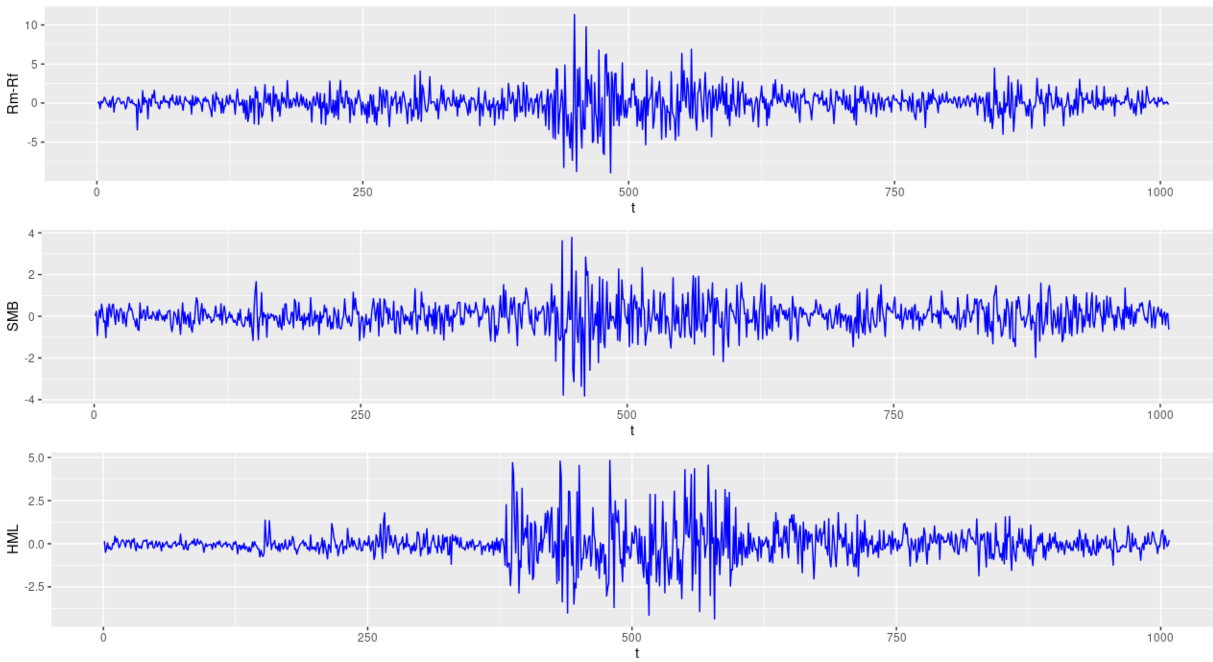

Figure 1: The time plot of the three factors (Rm-Rf, SMB and HML).

Table 6: Stationary test results

Variable	x_1	x_2	x_3
t-statistic	-9.8586***	-9.3702***	-9.6581***
p-value	0	0	0

Table 7: Test the conditional heteroscedasticity

Method	Test statistic	p-value
$Q(m)$ of squared series(LM test)	1688.444	0.000
$Q_k(m)$ of squared series	2453.326	0.000
Robust Test (5% trimming)	1476.106	0.000
Rank-based Test	1778.974	0.000

GARCH model is decided by BIC rule. The estimated model is given as followed:

$$\widehat{\boldsymbol{\Sigma}}_x(t) = \widehat{\boldsymbol{D}}_t \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{D}}_t,$$

where

$$\widehat{\boldsymbol{\Gamma}} = \begin{pmatrix} 1 & 0.1798 & 0.4736 \\ 0.1798 & 1 & -0.0280 \\ 0.4736 & -0.0280 & 1 \end{pmatrix},$$

and $\widehat{\boldsymbol{D}}_t = \mathrm{diag}^{1/2}\{\widehat{h}_{1t}, \widehat{h}_{2t}, \widehat{h}_{3t}\}$,

$$\begin{cases} \widehat{h}_{1t} = 0.0522 + \underset{(0.0002)}{0.1060} x_{1t-1}^2 + \underset{(0.0000)}{0.8693} \widehat{h}_{1t-1}, \\ \widehat{h}_{2t} = 0.0159 + \underset{(0.0000)}{0.1440} x_{2t-1}^2 + \underset{(0.0000)}{0.8230} \widehat{h}_{2t-1}, \\ \widehat{h}_{3t} = 0.0140 + \underset{(0.0000)}{0.1921} x_{3t-1}^2 + \underset{(0.0000)}{0.7847} \widehat{h}_{3t-1}. \end{cases} \tag{17}$$

For each sub-series of $\{\boldsymbol{x}_t\}_{t=1}^T$, we use the Wald test introduced in Francq and Zakoïan[11] to test the significance of the coefficients, and the values in the parentheses in (17) are the p-values for related coefficients. It is found that all the parameter estimates in (17) are significant at 1% level, implying the estimated model (17) is suitable for the sub-series of $\boldsymbol{x}_t$.

Then, we apply the LSE method to estimate $\boldsymbol{B}$. Further, we obtain the residuals vector $\widehat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{B}}\boldsymbol{x}_t$ and the estimated residual covariance matrix $\widehat{\boldsymbol{\Sigma}}_u = \frac{1}{T}\sum_{t=1}^T \widehat{\boldsymbol{u}}_t \widehat{\boldsymbol{u}}_t^\tau$. We also give the heat map of the residual covariance matrix $\widehat{\boldsymbol{\Sigma}}_u$ in Figure 2 and it can be seen that a large number of $\widehat{\boldsymbol{\Sigma}}_u$'s off-diagonal entries are zeros or close to zero. Similarly, the heat map of the error correlation matrix of $\widehat{\boldsymbol{u}}_t$ is depicted in Figure 3, showing that many pairs of the cross-sectional units become weakly correlated. Hence, it can be concluded that $\widehat{\boldsymbol{\Sigma}}_u$ exists sparsity. And it is reasonable to use the thresholding method to estimate $\mathbf{\Sigma}_u$ and denote the corresponding estimator as $\widehat{\mathbf{\Sigma}}_u^{\mathcal{T}}$. Finally, the conditional volatility of $\boldsymbol{y}_t$ can be estimated by

$$\widehat{\mathbf{\Sigma}}_y(t) = \widehat{\boldsymbol{B}}\widehat{\mathbf{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^{\tau} + \widehat{\mathbf{\Sigma}}_u^{\mathcal{T}}.$$

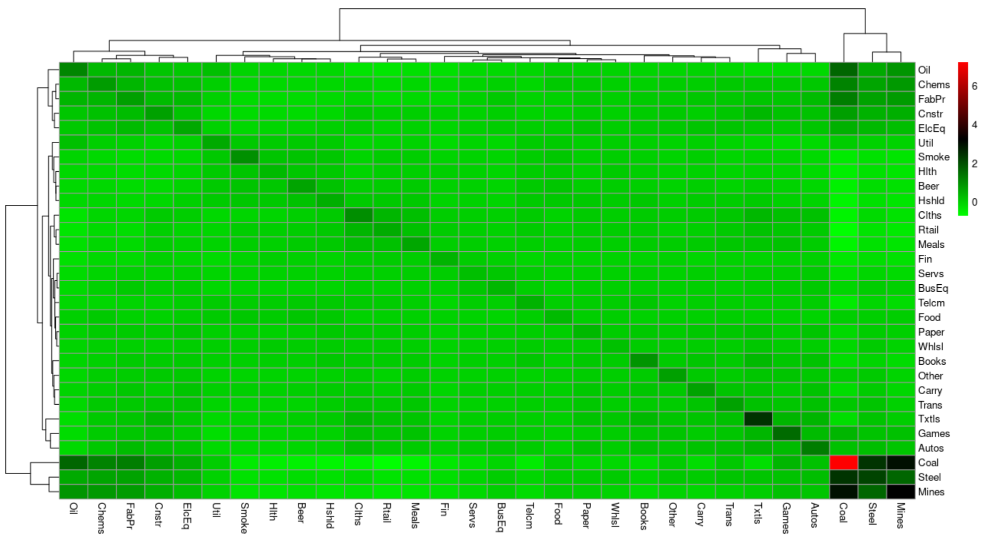

Figure 2: Heat map of residual covariance matrix $\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}_t^{\tau}$.

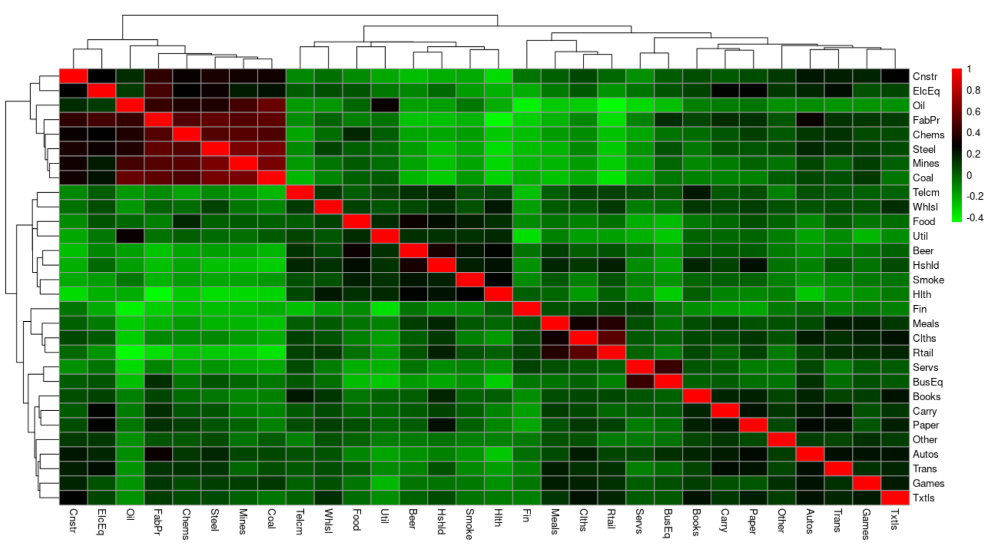

Figure 3: Heat map of the error correlation matrix of $\boldsymbol{u}_t$.

In addition, in order to find how well the proposed procedure works, we compare the three portfolio allocations based on our method, Fan, Fan and Lv[7] and Fan, Liao and Mincheva[8], denoted by *FGM* (Factor-GARCH model), *Fan2008* and *Fan2011*, respectively, and the portfolio allocations are constructed month by month from January 2010 to December 2010. We trade on approximately $T = 21$ trading days per month. Assume that all possible portfolio allocations are attainable and no transaction costs, and allow for short selling. In this trading strategy, we form a portfolio allocation $\widehat{\boldsymbol{w}}$ at the end of the trading day and hold it until the end of the next trading day. Between day $t-1$ and t, we calculate the realized return as follows:

$$R_t(\widehat{\boldsymbol{w}}) = \widehat{\boldsymbol{w}}^{\tau}\boldsymbol{y}_t,$$

where $\widehat{\boldsymbol{w}}$ is calculated based on $(\boldsymbol{x}_{t-i}, \boldsymbol{y}_{t-i})$, $i = 1, \cdots, n$, for some look-back integer n, and $t = 1, \cdots, T$. Then, we can obtain the monthly Sharpe ratio

$$\mathrm{SR}(\widehat{\boldsymbol{w}}) = \frac{\bar{R}(\widehat{\boldsymbol{w}})}{\mathrm{sd}(\widehat{\boldsymbol{w}})}\sqrt{T},$$

where

$$\bar{R}(\widehat{\boldsymbol{w}}) = \frac{1}{T}\sum_{t=1}^{T}[R_t(\widehat{\boldsymbol{w}}) - R_{ft}], \quad \mathrm{sd}(R) = \left\{\frac{1}{T}\sum_{t=1}^{T}[R_t(\widehat{\boldsymbol{w}}) - R_{ft} - \bar{R}(\widehat{\boldsymbol{w}})]^2\right\},$$

and R_{ft} is the risk-free rate on day t. We calculate the monthly Sharpe ratios at the end of the final trading day of each month for each trading strategy, and repeat this by using $n = 100$ and $n = 500$. From the monthly Sharpe ratios depicted in Figures 4 and 5, it is clear that our model outperforms the other two models in the majority of the observations in the period of post-financial crisis, i.e. in 2010. However, our model is not overwhelming, as the loading factor matrix estimated in the proposed paper is in the same way as Fan, Fan and Lv[7] and Fan, Liao and Mincheva[8]. But it provides us a potential direction to improve.

7 Conclusions

In this paper, we investigate the problem of estimating conditional volatilities of high dimensional data. In order to achieve this goal, a factor-GARCH model is proposed, where the approximate factor models is adopted for dimension reduction and multivariate GARCH process is used to describe the dynamics of the conditional volatility. The novelty of our article is: under the framework of the proposed model, the high dimensional conditional volatility is divided into two parts which can be computed based on the estimable factor loading, low dimensional conditional volatility for factor and the covariance matrix of the idiosyncratic error, by using LSE method

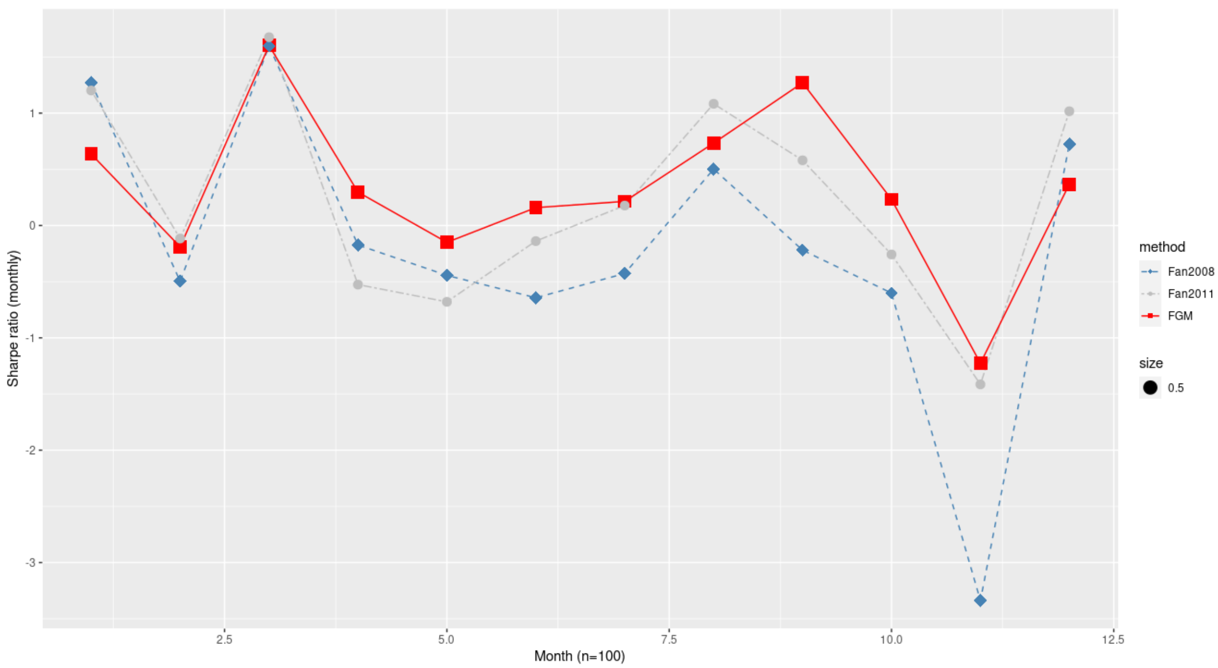

Figure 4: Monthly Sharpe ratio ($n = 100$).

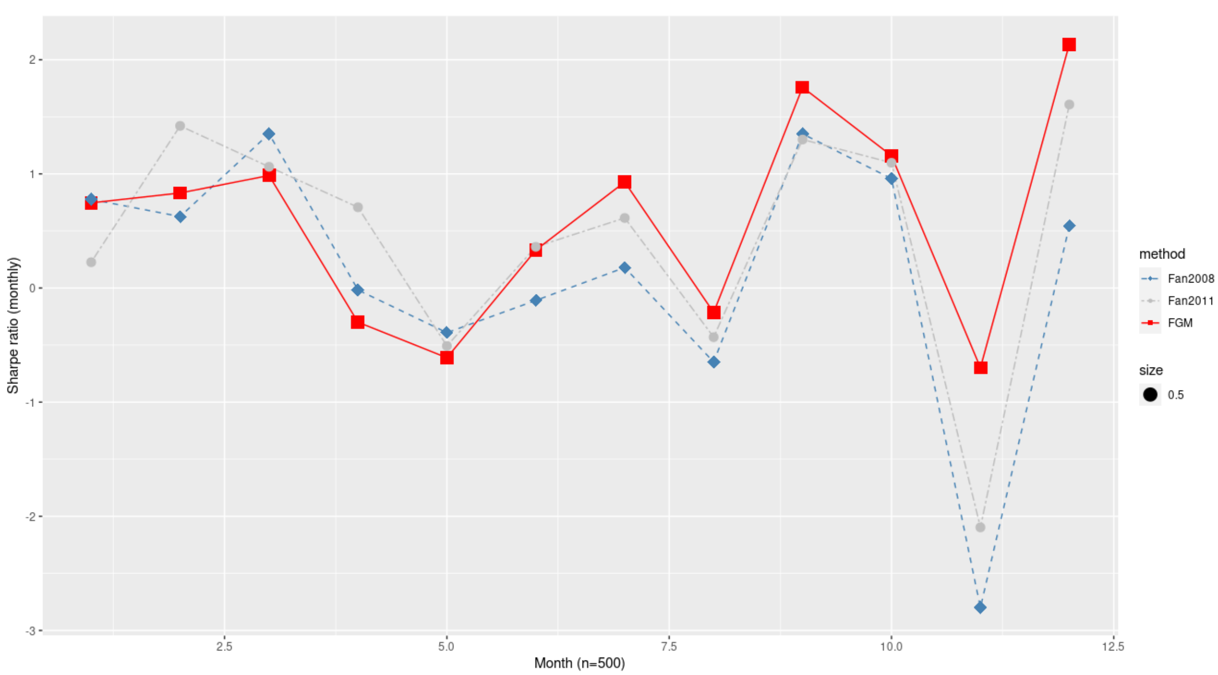

Figure 5: Monthly Sharpe ratio ($n = 500$).

for factor loading, QMLE method for multivariate GARCH and thresholding method for residual covariance matrix. Large sample properties of the estimator are proved under certain assumptions and their performances are examined through simulation studies. Given empirical example presents our model works well with real data.

On the basis of this paper, several questions are worth of further studying. In our study, we just consider the case that the common factors are observable and K is known. However, the common factors may be unobservable and K could be unknown in practice. Hence it makes sense to extend the proposed model to more general cases and we leave these for future studies.

Appendix: Proofs for Section 3

Lemma A.1. *Under Assumption 4(c), the components of* $\{\boldsymbol{x}_t\}_{t=1}^T$ *satisfy the exponential tail condition, i.e. there exist* $\nu_3 > 0$ *with* $3\nu_3^{-1} + \nu_2^{-1} > 1$*, and* $b_3 > 0$ *such that for any* $y > 0$ *and* $i \le K$*,*

$$P\big(|x_{it}| > y\big) \le \exp(-(y/b_3)^{\nu_3}).$$

Proof: By the proof of Theorem 10.8 of Francq and Zakoïan[11], and under Assumption 4, there exists a constant M_1 such that $\mathrm{E}\|\boldsymbol{\Sigma}_x(t)\|^{\varsigma} \le M_1 < \infty$ for some $\varsigma > 0$. And

$$\begin{aligned}
\|\boldsymbol{x}_t\| = \big\|\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t)\boldsymbol{\eta}_t\big\| &= \left\|\big(\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t) - \mathrm{E}\{\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t)\} + \mathrm{E}\{\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t)\}\big)\boldsymbol{\eta}_t\right\| \\
&\le \left(\big\|\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t) - \mathrm{E}\{\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t)\}\big\| + \big\|\mathrm{E}\{\boldsymbol{\Sigma}_x^{\frac{1}{2}}(t)\}\big\|\right)\|\boldsymbol{\eta}_t\| \\
&\le M\|\boldsymbol{\eta}_t\|,
\end{aligned}$$

where M is a constant, and there exists a small constant ν, such that

$$\begin{aligned}
P\big\{\|\boldsymbol{x}_t\| > y\big\} &\le P\big\{M\|\boldsymbol{\eta}_t\| > y\big\} \\
&\le K \max_{1\le k\le K} P\left\{\eta_{kt}^2 > \frac{1}{K}\big(\frac{y}{M}\big)^2\right\} \\
&\le K \exp\left(-\big(\frac{y}{\sqrt{K}Mb}\big)^{\nu}\right),
\end{aligned}$$

as K is fixed and small, and we can choose $\nu_3 \in (0, \nu)$ satisfying $3\nu_3^{-1} + \nu_2^{-1} > 1$, and $b_3 > \sqrt{K}Mb$, such that

$$K \exp\left(-\big(\frac{y}{\sqrt{K}Mb}\big)^{\nu}\right) \le \exp\big(-(y/b_3)^{\nu_3}\big).$$

Then, we obtain the result. $\square$

Lemma A.2. *Under Assumptions of Theorem 2, there exists $C' > 0$ such that:*

(i)

$$P\bigg(\max_{i,j\le K}\Big|\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{jt}-Ex_{it}x_{jt}\Big|>C'\sqrt{\frac{\log T}{T}}\bigg)=O\Big(\frac{1}{T^2}\Big),$$

(ii)

$$P\bigg(\max_{k\le K,i\le p}\Big|\frac{1}{T}\sum_{t=1}^{T}x_{kt}u_{it}\Big|>C'\sqrt{\frac{\log p}{T}}\bigg)=O\Big(\frac{1}{T^2}+\frac{1}{p^2}\Big).$$

Proof:

Let $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{T}^{\infty}$ denote the σ-algebras generated by $\{(\boldsymbol{x}_t,\boldsymbol{u}_t):-\infty\le t\le 0\}$ and $\{(\boldsymbol{x}_t,\boldsymbol{u}_t):T\le t\le\infty\}$, respectively. We define the mixing coefficient

$$\alpha(T)=\sup_{A\in\mathcal{F}_{-\infty}^{0},B\in\mathcal{F}_{T}^{\infty}}|P(A)(B)-P(AB)|.$$

Under Assumption 4, $\{x_{kt}\}_{t=1}^{T}$ are strictly stationary and α-mixing with geometric rate (see Lindner[19]); Meanwhile, by Assumption 1, $\boldsymbol{x}_t$ and $\boldsymbol{u}_t$ are independent, and $\boldsymbol{u}_t$ are α-mixing with geometric rate, so there exist positive constants ν_2 and c, such that for all $t\in\mathcal{Z}^{+}$

$$\alpha(t)\le\exp(-ct^{\nu_2}),$$

and by Lemma A.1 and Assumption 2, $\boldsymbol{x}_t$ and $\boldsymbol{u}_t$ both have exponential tails. Therefore, the conditions of Lemma B.1 of Fan et al.[8] are satisfied. By using similar arguments of this Lemma, we obtain the results of Lemma A.2. $\square$

Lemma A.3. *Under the assumptions of Theorem 1 and Lemma A.2, there exist $C_1' > 0$ and $C_2' > 0$, such that:*

(i)

$$P\bigg(\max_{k\le K}\|\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\|>C_1'\sqrt{\frac{\log T}{T}}\bigg)=O\Big(\frac{1}{T^{1+\epsilon}}\Big),$$

(ii)

$$P\bigg(\|\widehat{\boldsymbol{B}}-\boldsymbol{B}\|_F^2>\frac{C_2'p\log p}{T}\bigg)=O\Big(\frac{1}{T^2}+\frac{1}{p^2}\Big).$$

Proof:

(i) Define the negative quasi log-likelihood function

$$\widetilde{L}_T(\boldsymbol{\theta})=\frac{1}{T}\sum_{t=1}^{T}\widetilde{l}_t(\boldsymbol{\theta}),\quad \widetilde{l}_t=\boldsymbol{x}_t^{\tau}(\widetilde{\boldsymbol{D}}_t\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{D}}_t)^{-1}\boldsymbol{x}_t+\log|\widetilde{\boldsymbol{D}}_t\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{D}}_t|.$$

First, we consider the consistency of $\widehat{\boldsymbol{\theta}}_T$. Recall that the observed negative quasi log-likelihood function

$$L_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\boldsymbol{\theta}), \quad l_t = \boldsymbol{x}_t^\tau(\boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t)^{-1}\boldsymbol{x}_t + \log|\boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t|.$$

By the proof of Theorem 10.7 in Francq and Zakoïan[11],

$$\widehat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0, \quad \text{almost surely as } T \to \infty.$$

Now, we consider the convergence rate of $\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\|$.

The proof of this part is based on a standard Taylor expansion of $L_T(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$. Since $\widehat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_0$, which lies in the interior of the parameter space, we thus have

$$\begin{aligned} 0 &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\widehat{\boldsymbol{\theta}}_T)}{\partial\boldsymbol{\theta}} \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\Big)(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0), \end{aligned}$$

where $\boldsymbol{\theta}^*$ is between $\widehat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$. Suppose we have shown that there exist two positive constants c_1 and c_2 such that

$$P\Big\{\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\| > c_1\varrho_T\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big), \tag{A.1}$$

where $\varrho_T = \sqrt{\frac{\log T}{T}}$, and

$$P\Big\{\inf_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\Big) \le c_2\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big), \tag{A.2}$$

where $V(\boldsymbol{\theta}_0)$ is a neighbourhood of $\boldsymbol{\theta}_0$. Denote

$$\mathcal{A}_T = \Big\{\inf_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\Big) > c_2\Big\},$$

where c_2 is defined in (A.2). Then, for each $x > 0$,

$$P\Big\{\Big\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\Big\| > x\Big\} \le P\Big\{\Big\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big\| > c_2 x\Big\} + P\Big(\mathcal{A}_T^C\Big),$$

Take $x = c_1\varrho_T/c_2$ and the proof of part (i) follows immediately from (A.1) and (A.2).

Now we prove (A.1) and (A.2). To establish (A.1) and (A.2), it suffices to prove the following four parts:

(b1) There exists a constant $c > 0$ such that

$$P\Big\{\Big\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\| > c\varrho_T\Big\} = o(1),$$

(b2) There exists a constant $c > 0$ such that

$$P\Big\{\Big\|\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \sum_{t=1}^{T}\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\| > cT\varrho_T\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big),$$

(b3) There exists a constant $c > 0$ such that

$$P\Big\{\lambda_{\min}\Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\tau}\Big) \le c\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big),$$

(b4) For any $c > 0$, we have

$$P\Big\{\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\Big\|\sum_{t=1}^{T}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\tau} - \sum_{t=1}^{T}\frac{\partial^2 \widetilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\tau}\Big\| > Tc\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big).$$

It can be seen that (A.1) can be proved from (b1) and (b2), and (A.2) follows from (b3) and (b4). Now we prove them separately.

Note that, for simplicity, we denote $\boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t$ by $\boldsymbol{H}$, and $\widetilde{\boldsymbol{H}}$ is its counterpart.

For **(b1)**, following the proof of Theorem 10.7 in Francq and Zakoïan[11], it is easy to show that

$$\begin{aligned}\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_i} &= \mathrm{Tr}\Big\{-\widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}_0)\boldsymbol{x}_t\boldsymbol{x}_t^\tau\widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta}_0)}{\partial \theta_i} + \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}_0)\frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta}_0)}{\partial \theta_i}\Big\} \\ &= \frac{\partial \mathrm{vec}^\tau \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta}_0)}{\partial \theta_i}\Big\{\widetilde{\boldsymbol{H}}_t^{-\frac{1}{2}'}(\boldsymbol{\theta}_0)\otimes\widetilde{\boldsymbol{H}}_t^{-\frac{1}{2}'}(\boldsymbol{\theta}_0)\mathrm{vec}\Big\{\boldsymbol{I}_K - \boldsymbol{\eta}_t\boldsymbol{\eta}_t^\tau\Big\}\Big\},\end{aligned} \tag{A.3}$$

where $\boldsymbol{H}_t^{-\frac{1}{2}'} = (\boldsymbol{H}_t^{\frac{1}{2}})^{-1}$, $\boldsymbol{A}\otimes\boldsymbol{B}$ is the Kronecker product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and

$$\mathrm{E}\Big\|\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \theta_i}\Big\|^\kappa < \infty.$$

Note that $\{\boldsymbol{x}_t\}_{t=1}^T$ are strictly stationary and α-mixing with geometric rate (Also see Lindner[19]). It follows from Theorem 2(ii) of Liu, Xiao and Wu[20] that, there exist positive constants c_1', c_2' and c_3' such that for all $x > 0$

$$P\Big\{\Big\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\| > x\Big\} \le \frac{c_1'T}{(Tx)^\kappa} + c_2'\exp\Big(-c_3'Tx^2\Big).$$

Hence, by taking $x = c\varrho_T$, for a large constant $c > 0$ and $\kappa > 4$. We obtain that

$$\begin{aligned}P\Big\{\Big\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\Big\| > c\varrho_T\Big\} &\le \frac{c_1'T^{1-\kappa/2}}{c^\kappa(\log T)^{\kappa/2}} + c_2'\exp\Big(-c_3'c^2\log T\Big) \\ &\le O\Big(\frac{1}{T^{1+\epsilon}}\Big).\end{aligned}$$

For **(b2)**, by the proof of Theorem 10.9 of Francq and Zakoïan[11],

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big\|\frac{\partial\widetilde{\boldsymbol{H}}_t}{\partial\theta_i}-\frac{\partial\boldsymbol{H}_t}{\partial\theta_i}\Big\|\le K\rho^t\Big\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\widetilde{\boldsymbol{D}}_t\|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\Big\|\frac{\partial\widetilde{\boldsymbol{D}}_t}{\partial\theta_i}\Big\|+1\Big\}.$$

Note that (A.3) continues to hold when $\widetilde{l}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})$ are replaced by $l_t(\boldsymbol{\theta})$ and $\boldsymbol{H}_t(\boldsymbol{\theta})$, respectively. Therefore,

$$\frac{\partial\widetilde{l}_t(\boldsymbol{\theta})}{\partial\theta_i}-\frac{\partial l_t(\boldsymbol{\theta})}{\partial\theta_i}=\mathrm{Tr}(I+II),$$

where

$$I=\Big(\boldsymbol{I}_K-\widetilde{\boldsymbol{H}}_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t^{\tau}\Big)\Big(\widetilde{\boldsymbol{H}}_t^{-1}-\boldsymbol{H}_t^{-1}\Big)\frac{\partial\widetilde{\boldsymbol{H}}_t}{\partial\theta_i}-\Big(\widetilde{\boldsymbol{H}}_t^{-1}-\boldsymbol{H}_t^{-1}\Big)\boldsymbol{x}_t\boldsymbol{x}_t^{\tau}\boldsymbol{H}_t^{-1}\frac{\partial\boldsymbol{H}_t}{\partial\theta_i},$$
$$II=\Big(\boldsymbol{I}_K-\widetilde{\boldsymbol{H}}_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t^{\tau}\Big)\boldsymbol{H}_t^{-1}\Big(\frac{\partial\widetilde{\boldsymbol{H}}_t}{\partial\theta_i}-\frac{\partial\boldsymbol{H}_t}{\partial\theta_i}\Big).$$

In Theorem 10.9 of Francq and Zakoïan[11], it has been shown that $\mathrm{Tr}(I+II)\le K\rho^t z_t$, where z_t is a random variable such that $\sup_t \mathrm{E}|z_t|^{\varsigma}<\infty$ for some small $\varsigma\in(0,1)$. Then it follows that,

$$\sum_{t=1}^T\Big|\frac{\partial l_t(\boldsymbol{\theta})}{\partial\theta_i}-\frac{\partial\widetilde{l}_t(\boldsymbol{\theta})}{\partial\theta_i}\Big|\le\sum_{t=1}^T K\rho^t z_t.$$

By Markov inequality for martingale, we claim that there exists a constant $c'>0$ such that

$$P\Big\{\sum_{t=1}^T K\rho^t|z_t|>c'T^{1/2}\Big\}=O\Big(\frac{1}{T^{1+\epsilon}}\Big).$$

Hence, if follows that there exists a constant $c>0$ such that

$$P\Big\{\frac{1}{T}\sum_{t=1}^T\Big|\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\theta_i}-\frac{\partial\widetilde{l}_t(\boldsymbol{\theta}_0)}{\partial\theta_i}\Big|\ge c\varrho_T\Big\}=O\Big(\frac{1}{T^{1+\epsilon}}\Big),$$

and the conclusion of this part follows.

For **(b3)**, $T^{-1}\sum_{t=1}^T\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}$ can be expressed as

$$\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}=\frac{1}{T}\sum_{t=1}^T\Big\{\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}-\mathrm{E}\Big(\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}\Big)\Big\}+\mathrm{E}\Big\{\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}\Big\}.$$

Note that $\mathrm{E}\Big\{\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}\Big\}$ is positive definite, and there exists a constant $c>0$, such that $\lambda_{\min}\mathrm{E}\Big\{\frac{\partial^2\widetilde{l}_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\tau}}\Big\}>c$. Differentiating (A.3), we have

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\widetilde{l}_t(\boldsymbol{\theta})=\sum_{t=1}^5\widetilde{c}_i(\boldsymbol{\theta}),\tag{A.4}$$

with

$$\begin{aligned}
\widetilde{c}_1(\boldsymbol{\theta}) &= \boldsymbol{x}_t^\tau \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_i} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_j} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{x}_t, \\
\widetilde{c}_2(\boldsymbol{\theta}) &= \boldsymbol{x}_t^\tau \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_j} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_i} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{x}_t, \\
\widetilde{c}_3(\boldsymbol{\theta}) &= -\boldsymbol{x}_t^\tau \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \boldsymbol{x}_t, \\
\widetilde{c}_4(\boldsymbol{\theta}) &= -\mathrm{Tr}\bigg(\frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_i} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_j} \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \bigg), \\
\widetilde{c}_5(\boldsymbol{\theta}) &= \mathrm{Tr}\bigg(\widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \bigg).
\end{aligned}$$

Similar to (b1), we claim that there exist three positive constants c_1', c_2' and c_3' such that

$$P\bigg\{ \frac{1}{T} \bigg\| \sum_{t=1}^T \Big\{ \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} - \mathrm{E}\big(\frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \big) \Big\} \bigg\| > c \bigg\} \le c_1' \frac{T}{(Tc)^\kappa} + c_2' \exp\Big(-c_3' T c^2 \Big) = O\Big(\frac{1}{T^{1+\epsilon}} \Big),$$

so

$$P\bigg\{ \lambda_{\min}\Big(\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \Big) < c \bigg\} = O\Big(\frac{1}{T^{1+\epsilon}} \Big),$$

then, part (b3) follows.

For **(b4)**, note that (A.4) continues to hold when $\widetilde{l}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{H}}_t(\boldsymbol{\theta})$ are replaced by $l_t(\boldsymbol{\theta})$ and $\boldsymbol{H}_t(\boldsymbol{\theta})$. Therefore, similar to (b2), we can prove that there exists a constant $c > 0$ such that

$$P\bigg\{ \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{1}{T} \bigg\| \sum_{t=1}^T \Big\{ \frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} - \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \Big\} \bigg\| > c \bigg\} = O\Big(\frac{1}{T^{1+\epsilon}} \Big).$$

Moreover,

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} &= \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{1}{T} \sum_{t=1}^T \Big\{ \frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} - \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \Big\} \\
&\quad + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{1}{T} \sum_{t=1}^T \Big\{ \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} - \mathrm{E}\Big(\frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \Big) \Big\} + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \mathrm{E}\Big\{ \frac{\partial^2 \widetilde{l}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \Big\}.
\end{aligned}$$

Together with the proof of (b3), (A.2) follows.

(ii) Let $\boldsymbol{C}_T \equiv \widehat{\boldsymbol{B}} - \boldsymbol{B} = \boldsymbol{U}\boldsymbol{X}^\tau(\boldsymbol{X}\boldsymbol{X}^\tau)^{-1}$. By Lemma A.2(ii), there exists $C_2' > 0$ such that

$$P\bigg(\max_{k,i} \Big| \frac{1}{T} \sum_{t=1}^T x_{kt} u_{it} \Big| > C_2' \sqrt{\frac{\log p}{T}} \bigg) = O\Big(\frac{1}{T^2} + \frac{1}{p^2} \Big).$$

Under the event

$$A = \bigg\{ \max_{k,i} \Big| \frac{1}{T} \sum_{t=1}^T x_{kt} u_{it} \Big| \le C_2' \sqrt{\frac{\log p}{T}} \bigg\} \cap \Big\{ \lambda_{\min}(T^{-1} \boldsymbol{X}\boldsymbol{X}^\tau) \ge 0.5 \lambda_{\min}(\mathrm{E}\boldsymbol{x}_t \boldsymbol{x}_t^\tau) \Big\},$$

$\|\boldsymbol{C}_T^{\tau}\|_F^2 \leq 4\lambda_{\min}^{-2}(\mathrm{E}\boldsymbol{x}_t\boldsymbol{x}_t^{\tau})C_2'^2 pK\log p/T$, which proves the result since $\lambda_{\min}(\mathrm{E}\boldsymbol{x}_t\boldsymbol{x}_t^{\tau})$ is bounded away from zero and $P(A) \geq 1 - O\Big(\frac{1}{T^2} + \frac{1}{p^2}\Big)$. $\square$

Proof of Theorem 1.

For

$$\Big\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\Big\| \leq K^2 \max_{i,j\leq K}\Big|\widehat{\gamma}_{ij} - \gamma_{ij}\Big|,$$

following Lemma A.3(i) and K is a constant, it is easy to prove that

$$P\Big\{\Big\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\Big\| \geq C\sqrt{\frac{\log T}{T}}\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big).$$

And for

$$\Big\|\widehat{\boldsymbol{D}_t} - \boldsymbol{D}_t\Big\| \leq K \max_{k\leq K}\Big|\widehat{h}_{kt} - h_{kt}\Big|,$$

as h_{kt} and $\widehat{h}_{kt}$ are measurable functions of $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_T$, respectively, then

$$\Big|\widehat{h}_{kt} - h_{kt}\Big| = O_p\Big(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\Big),$$

so following Lemma A.3(i), we obtain

$$P\Big\{\Big\|\widehat{\boldsymbol{D}_t} - \boldsymbol{D}_t\Big\| > C\sqrt{\frac{\log T}{T}}\Big\} = O\Big(\frac{1}{T^{1+\epsilon}}\Big).$$

Denote

$$\begin{aligned}
\boldsymbol{G}_T =& \widehat{\boldsymbol{D}}_t\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t\boldsymbol{\Gamma}\boldsymbol{D}_t\\
=&(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{D}}_t + \boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\widehat{\boldsymbol{D}}_t + \boldsymbol{D}_t\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\\
=&(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\widehat{\boldsymbol{D}}_t + (\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t) + (\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\boldsymbol{\Gamma}\boldsymbol{D}_t\\
&+ \boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t) + \boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{D}_t + \boldsymbol{D}_t\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\\
=&(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t) + (\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{D}_t + (\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\\
&+ (\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\boldsymbol{\Gamma}\boldsymbol{D}_t + \boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t) + \boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{D}_t + \boldsymbol{D}_t\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t),
\end{aligned}$$

and we have

$$\begin{aligned}
\Big\|\boldsymbol{G}_T\Big\|_F \leq& \Big\|(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\Big\|_F + 2\Big\|(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{D}_t\Big\|_F\\
&+ \Big\|(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\Big\|_F + 2\Big\|\boldsymbol{D}_t\boldsymbol{\Gamma}(\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t)\Big\|_F + \Big\|\boldsymbol{D}_t(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{D}_t\Big\|_F\\
=&O_p\Big(\|\widehat{\boldsymbol{D}}_t - \boldsymbol{D}_t\|_F\Big) + O_p\Big(\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_F\Big).
\end{aligned}$$

We shall repeatedly use the fact that, for a $K \times K$ matrix $\boldsymbol{A}$

$$\|\boldsymbol{A}\| \le \|\boldsymbol{A}\|_F \le \sqrt{K}\|\boldsymbol{A}\|,$$

then, we have

$$P\Big(\Big\|\boldsymbol{G}_T\Big\|_F^2 > \frac{C\log T}{T}\Big) = O\Big(\frac{1}{T^{1+\epsilon}}\Big).$$

This completes the proof of Theorem 1. □

In order to prove Theorem 2, we need the following extra Lemmas A.4 - A.7. First, define

$$\boldsymbol{C}_T = \widehat{\boldsymbol{B}} - \boldsymbol{B}.$$

Lemma A.4. *Under the same assumptions of Lemmas A.2 and A.3, there exists $C_3' > 0$ such that:*

(i)

$$P\Big(\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\|_\Sigma^2 + \|\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\|_\Sigma^2 > \frac{C_3'\log T}{Tp} + \frac{C_3'\log p}{T}\Big) = O\Big(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\Big);$$

(ii)

$$P\Big(\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\|_\Sigma^2 > \frac{C_3' p(\log p)^2}{T^2}\Big) = O\Big(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\Big);$$

Proof:

(i) Like the argument in proof of Theorem 2 in Fan, Fan and Lv[7], by the Sherman-Morrison-Woodbury formula:

$$\boldsymbol{\Sigma}_y^{-1}(t) = \boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1},$$

further,

$$\begin{aligned}
\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B} &= \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B} - \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\\
&= \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\big\{[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}] - \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\big\}\\
&= \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{\Sigma}_x^{-1}(t)\\
&= [\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B} - \boldsymbol{\Sigma}_x^{-1}(t)][\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{\Sigma}_x^{-1}(t)\\
&= \boldsymbol{\Sigma}_x^{-1}(t) - \boldsymbol{\Sigma}_x^{-1}(t)[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{\Sigma}_x^{-1}(t).
\end{aligned}$$

As $\boldsymbol{\Sigma}_x^{-1}(t)$ is positive definite, and $\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}$ is positive semi-definite. Then, we have

$$\boldsymbol{\Sigma}_x^{-1}(t)+\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\geq\boldsymbol{\Sigma}_x^{-1}(t),$$

so

$$[\boldsymbol{\Sigma}_x^{-1}(t)+\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\leq\boldsymbol{\Sigma}_x(t).$$

It means that

$$\boldsymbol{\Sigma}_x^{-1}(t)[\boldsymbol{\Sigma}_x^{-1}(t)+\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{\Sigma}_x^{-1}(t)\leq\boldsymbol{\Sigma}_x^{-1}(t)\boldsymbol{\Sigma}_x(t)\boldsymbol{\Sigma}_x^{-1}(t)=\boldsymbol{\Sigma}_x^{-1}(t),$$

and

$$\big\|\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\big\|\leq\big\|\boldsymbol{\Sigma}_x^{-1}(t)\big\|+\big\|\boldsymbol{\Sigma}_x^{-1}(t)[\boldsymbol{\Sigma}_x^{-1}(t)+\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\boldsymbol{\Sigma}_x^{-1}(t)\big\|\leq2\big\|\boldsymbol{\Sigma}_x^{-1}(t)\big\|=O_p(1).$$

Hence

$$\begin{aligned}\Big\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\Big\|_\Sigma^2&=p^{-1}tr\Big(\boldsymbol{\Sigma}_y^{-1/2}(t)\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1/2}(t)\Big)\\&=p^{-1}tr\Big(\boldsymbol{G}_T\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\Big)\\&\leq p^{-1}\|\boldsymbol{G}_T\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\|_F^2\\&\leq O(p^{-1})\|\boldsymbol{G}_T\|_F^2\\&=O_p\Big(\frac{C_3'\log T}{pT}\Big).\end{aligned}$$

Under Assumption 4, $\|\boldsymbol{\Sigma}_x(t)\|=O_p(1)$. And by Theorem 1, $P(\|\widehat{\boldsymbol{\Sigma}}_x(t)\|>C_3')=O(\frac{1}{T^{1+\epsilon}})$, for some $C_3'>0$. Hence, Lemma A.3(ii) implies

$$\begin{aligned}\Big\|\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|_\Sigma^2&=p^{-1}\Big\|\boldsymbol{\Sigma}_y^{-\frac{1}{2}}(t)\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_y^{-\frac{1}{2}}(t)\Big\|_F^2\\&=p^{-1}tr\Big\{\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\Big\}\\&\leq p^{-1}\|\widehat{\boldsymbol{\Sigma}}_x(t)\|_F^2\|\boldsymbol{C}_T\|_F^2\|\boldsymbol{\Sigma}_y^{-1}(t)\|_F\|\boldsymbol{B}^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{B}\|_F^2\\&=O_p\Big(\frac{C_3'\log p}{T}\Big).\end{aligned}$$

(ii). Straightforward calculation yields

$$\begin{aligned}p\Big\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|_\Sigma^2&=tr\Big\{\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_y^{-1}(t)\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_y^{-1}(t)\Big\}\\&\leq\Big\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_y^{-1}(t)\Big\|_F^2\\&\leq\lambda_{\max}^2(\boldsymbol{\Sigma}_y^{-1}(t))\lambda_{\max}^2(\widehat{\boldsymbol{\Sigma}}_x(t))\|\boldsymbol{C}_T\|_F^4\\&=O_p\Big\{\Big(\frac{C_3'p\log p}{T}\Big)^2\Big\},\end{aligned}$$

since $\|\boldsymbol{\Sigma}_x(t)\| = O_p(1)$, and by Theorem 1, $\lambda^2_{\max}(\widehat{\Sigma}_x(t))$ is bounded with probability at least $1 - O\Big(\frac{1}{T^{1+\epsilon}}\Big)$. The result again follows from Lemma A.3(ii). $\square$

Lemma A.5. *Suppose* $(\log p)^2 = o(T)$ *and* $(\log p)^{2/\nu - 1} = o(T)$. *Then under Assumptions 1-4, there exists* $C_4' > 0$, *such that:*

(i)

$$P\bigg(\max_{i \le p} \Big\|\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i\Big\| > C_4' \sqrt{\frac{\log p}{T}}\bigg) = O\Big(\frac{1}{T^2} + \frac{1}{p^2}\Big).$$

Proof: This part can be proved in a similar fashion to part (i) of Lemma 3.1 in Fan, Liao and Mincheva[8]. $\square$

Proof of Theorem 2, part (i).

We have

$$\begin{aligned}
\Big\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\Big\|^2_\Sigma &\le \Big\|\Big(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big) + \Big(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big)\Big\|^2_\Sigma \\
&\le \Big(\Big\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}} - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|_\Sigma + \Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|_\Sigma\Big)^2 \\
&\le 2\Big\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|^2_\Sigma + 2\Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|^2_\Sigma,
\end{aligned}$$

and

$$\begin{aligned}
\Big\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|^2_\Sigma =& \Big\|(\widehat{\boldsymbol{B}} - \boldsymbol{B})\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau + \boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|^2_\Sigma \\
=& \Big\|(\widehat{\boldsymbol{B}} - \boldsymbol{B})\widehat{\boldsymbol{\Sigma}}_x(t)(\widehat{\boldsymbol{B}} - \boldsymbol{B})^\tau + (\widehat{\boldsymbol{B}} - \boldsymbol{B})\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{B}^\tau \\
&+ \boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)(\widehat{\boldsymbol{B}} - \boldsymbol{B})^\tau + \boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{B}^\tau - \boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|^2_\Sigma \\
=& \Big\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau + \boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau + \boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T + \boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{B}^\tau\Big\|^2_\Sigma \\
\le& 4\Big\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|^2_\Sigma + 4\Big\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\Big\|^2_\Sigma + 8\Big\|\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|^2_\Sigma.
\end{aligned}$$

So

$$\begin{aligned}
\Big\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\Big\|^2_\Sigma \le& 8\Big\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|^2_\Sigma + 8\Big\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\Big\|^2_\Sigma \\
&+ 16\Big\|\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\Big\|^2_\Sigma + 2\Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|^2_\Sigma.
\end{aligned}$$

(a) We have

$$\begin{aligned}
\Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|_\Sigma &= p^{-1/2}\Big\|\boldsymbol{\Sigma}_y^{-1/2}(t)(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u)\boldsymbol{\Sigma}_y^{-1/2}(t)\Big\|_F \\
&\le \Big\|\boldsymbol{\Sigma}_y^{-1/2}(t)(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u)\boldsymbol{\Sigma}_y^{-1/2}(t)\Big\| \\
&\le \Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\| \cdot \lambda_{\max}(\boldsymbol{\Sigma}_y^{-1}(t)).
\end{aligned}$$

Therefore, Lemma 2 and Lemma A.4 yield the following result.

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}}_y(t)-\boldsymbol{\Sigma}_y(t)\right\|_{\Sigma}^2 \leq & 8\left\|\boldsymbol{C}_T\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\right\|_{\Sigma}^2+8\left\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\right\|_{\Sigma}^2 \\ & +16\left\|\boldsymbol{B}\widehat{\boldsymbol{\Sigma}}_x(t)\boldsymbol{C}_T^\tau\right\|_{\Sigma}^2+2\left\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}-\boldsymbol{\Sigma}_u\right\|_{\Sigma}^2 \\ = & O_p\Big(\frac{p(\log p)^2}{T^2}\Big)+O_p\Big(\frac{\log p}{T}\Big)+O_p\Big(\frac{\log T}{Tp}\Big)+O_p\Big(\frac{m_p^2\log p}{T}\Big) \\ = & O_p\Big(\frac{p(\log p)^2}{T^2}+\frac{m_p^2\log p}{T}\Big).\end{aligned}$$

(b). For the infinity norm, it is straightforward to find that

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}}_y(t)-\boldsymbol{\Sigma}_y(t)\right\|_{\max} = & \left\|\Big(\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau-\boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big)+(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}-\boldsymbol{\Sigma}_u)\right\|_{\max} \\ \leq & \left\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Sigma}}_x(t)\widehat{\boldsymbol{B}}^\tau-\boldsymbol{B}\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\right\|_{\max}+\left\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}-\boldsymbol{\Sigma}_u\right\|_{\max} \\ \leq & 2\left\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\right\|_{\max}+\left\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\right\|_{\max}+\left\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{C}_T^\tau\right\|_{\max} \\ & +2\left\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{C}_T^\tau\right\|_{\max}+\left\|\boldsymbol{C}_T\boldsymbol{G}_T\boldsymbol{C}_T^\tau\right\|_{\max}+\left\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}}-\boldsymbol{\Sigma}_u\right\|_{\max}.\end{aligned}$$

Under Assumptions 3 and 4, both $\|\boldsymbol{B}\|_{\max}$ and $\|\boldsymbol{\Sigma}_x(t)\|_{\max}$ are bounded uniformly in (p,T). And by Lemma A.5, $P\Big(\max_{i\leq p}\|\widehat{\boldsymbol{b}}_i-\boldsymbol{b}_i\|>C_4'\sqrt{\frac{\log p}{T}}\Big)=O\Big(\frac{1}{p^2}+\frac{1}{T^2}\Big)$. In addition, let $\boldsymbol{e}_i$ be a p-dimensional column vector whose i^{th} component is one with the remaining components being zeros. Then under the events $\|\boldsymbol{G}_T\|_{\max}\leq c_4'\sqrt{\frac{\log T}{T}}$, $\|\widehat{\boldsymbol{\Sigma}}_x(t)\|\leq c_4'$, and $\max_{j\leq p}\|\widehat{\boldsymbol{b}}_i-\boldsymbol{b}_i\|\leq c_4'\sqrt{\log p/T}$, for some $c_4'>0$, we have several results as follows:

$$\begin{aligned}2\left\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\right\|_{\max} & \leq 2\max_{i,j\leq p}\left\|\boldsymbol{e}_i^\tau\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\boldsymbol{e}_j\right\| \\ & \leq 2\max_{i\leq p}\left\|\widehat{\boldsymbol{b}}_i-\boldsymbol{b}_i\right\|\cdot\left\|\boldsymbol{\Sigma}_x(t)\right\|\cdot\max_{j\leq p}\left\|\boldsymbol{b}_j\right\| \\ & \leq O_p\Big(\sqrt{\frac{\log p}{T}}\Big)\cdot O_p(1)\cdot\max_j\sqrt{\sum_{i=1}^K b_{ji}^2} \\ & = O_p\Big(\sqrt{\frac{\log p}{T}}\Big),\end{aligned}\tag{A.5}$$

$$\begin{aligned}\left\|\boldsymbol{C}_T\right\|_{\max} & \leq \max_{i,j\leq p}\left|\boldsymbol{e}_i^\tau\frac{1}{T}\boldsymbol{U}\boldsymbol{X}^\tau(\frac{1}{T}\boldsymbol{X}\boldsymbol{X}^\tau)^{-1}\boldsymbol{e}_j\right| \\ & \leq \max_{i\leq p}\left\|\boldsymbol{e}_i^\tau\frac{1}{T}\boldsymbol{U}\boldsymbol{X}^\tau\right\|\cdot\left\|(\frac{1}{T}\boldsymbol{X}\boldsymbol{X}^\tau)^{-1}\right\| \\ & \leq \sqrt{K}\max_{i\leq K,j\leq p}\left|\frac{1}{T}\sum_{t=1}^T x_{it}u_{jt}\right|\cdot\left\|(\frac{1}{T}\boldsymbol{X}\boldsymbol{X}^\tau)^{-1}\right\| \\ & = O_p\Big(\sqrt{\frac{\log p}{T}}\Big),\end{aligned}\tag{A.6}$$

$$\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\|_{\max} \leq K^2\|\boldsymbol{B}\|^2_{\max}\|\boldsymbol{G}_T\|_{\max} = O_p\Big(\sqrt{\frac{\log T}{T}}\Big), \tag{A.7}$$

$$\begin{aligned}\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{C}_T^\tau\|_{\max} &\leq \max_{i,j}\|\boldsymbol{e}_i^\tau\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{C}_T^\tau\boldsymbol{e}_j\| \\ &\leq \max_{i\leq p}\|\boldsymbol{e}_i^\tau\boldsymbol{C}_T\|^2\|\boldsymbol{\Sigma}_x(t)\| \\ &\leq \max_{i\leq p}\|\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i\|^2\cdot\|\boldsymbol{\Sigma}_x(t)\| \\ &= O_p\Big(\frac{\log p}{T}\Big),\end{aligned} \tag{A.8}$$

$$\begin{aligned}\|2\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{C}_T^\tau\|_{\max} &\leq 2K^2\|\boldsymbol{B}\|_{\max}\|\boldsymbol{G}_T\|_{\max}\|\boldsymbol{C}_T\|_{\max} \\ &\leq C'K^2O_p\Big(\sqrt{\frac{\log T}{T}}\Big)O_p\Big(\sqrt{K\frac{\log p}{T}}\Big) \\ &= o_p\Big(\sqrt{\frac{\log T}{T}}\Big),\end{aligned} \tag{A.9}$$

and

$$\big\|\boldsymbol{C}_T\boldsymbol{G}_T\boldsymbol{C}_T^\tau\big\|_{\max} \leq K^2\big\|\boldsymbol{G}_T\big\|_{\max}\big\|\boldsymbol{C}_T^\tau\big\|^2_{\max} = o_p\Big(\sqrt{\frac{\log T}{T}}\Big). \tag{A.10}$$

Moreover, the (i,j)th entry of $\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u$ is given by

$$\widehat{\sigma}_{u,ij}I\Big(|\widehat{\sigma}_{u,ij}| \geq \omega_T\sqrt{\widehat{\vartheta}_{u,ij}}\Big) - \sigma_{u,ij} = \begin{cases} -\sigma_{u,ij}, & \text{if } |\widehat{\sigma}_{u,ij}| < \omega_T\sqrt{\widehat{\vartheta}_{u,ij}}, \\ \widehat{\sigma}_{u,ij} - \sigma_{u,ij}, & \text{o.w.}\end{cases}.$$

Hence,

$$\Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|_{\max} \leq \max_{i,j\leq p}\Big|\widehat{\sigma}_{u,ij} - \sigma_{u,ij}\Big| + \omega_T\max_{i,j\leq p}\sqrt{\vartheta_{u,ij}},$$

where $\omega_T = C_1\sqrt{\frac{\log p}{T}}$, and by Lemma A.3 and Lemma A.4 of Fan, Liao and Mincheva[8], we can obtain the following inequality with probability at least $1 - O(p^{-2} + T^{-2})$

$$\Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|_{\max} \leq C_2\sqrt{\frac{\log p}{T}}.$$

Then, combining Equations (A.5)-(A.10), we have

$$\begin{aligned}\Big\|\widehat{\boldsymbol{\Sigma}}_y(t) - \boldsymbol{\Sigma}_y(t)\Big\|_{\max} \leq& 2\Big\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{B}^\tau\Big\|_{\max} + \Big\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{B}^\tau\Big\|_{\max} + \Big\|\boldsymbol{C}_T\boldsymbol{\Sigma}_x(t)\boldsymbol{C}_T^\tau\Big\|_{\max} \\ &+ 2\Big\|\boldsymbol{B}\boldsymbol{G}_T\boldsymbol{C}_T^\tau\Big\|_{\max} + \Big\|\boldsymbol{C}_T\boldsymbol{G}_T\boldsymbol{C}_T^\tau\Big\|_{\max} + \Big\|\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}} - \boldsymbol{\Sigma}_u\Big\|_{\max} \\ =& O_p\Big(\sqrt{\frac{\log p}{T}}\Big) + O_p\Big(\sqrt{\frac{\log T}{T}}\Big) + O_p\Big(\frac{\log p}{T}\Big) \\ &+ o_p\Big(\sqrt{\frac{\log T}{T}}\Big) + o_p\Big(\sqrt{\frac{\log T}{T}}\Big) + O_p\Big(\sqrt{\frac{\log p}{T}}\Big) \\ =& O_p\Big(\sqrt{\frac{\log p}{T}}\Big) + O_p\Big(\sqrt{\frac{\log T}{T}}\Big).\end{aligned}$$

Hence, we obtain the second equation of part (i) in Theorem 2. □

To complete the proof of part (ii) in Theorem 2, we first prove two technical lemmas.

Lemma A.6. *Under the Assumptions 2-3,*

(i) $\lambda_{\min}(\boldsymbol{B}^\tau \boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}) \geq c''p$ *for some* $c'' > 0$.

(ii) $\|[\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau \boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}]^{-1}\| = O_p(p^{-1})$.

Proof:

(i) The proof is the same as the Lemma B.4(i) of Fan, Liao and Mincheva[8].

(ii) Note that $\boldsymbol{\Sigma}_x(t)$ is positive definite. Then, it follows immediately from

$$\lambda_{\min}(\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau \boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{B}^\tau \boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}).$$

□

Lemma A.7. *Under Assumptions 1-4, there exists* $C_5' > 0$ *such that*

(i)

$$P\Big(\Big\|\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}} - \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\Big\| > C_5' p m_p\sqrt{\frac{\log p}{T}}\Big) = O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big);$$

(ii)

$$P\Big(\Big\|[\widehat{\boldsymbol{\Sigma}}_x^{-1}(t) + \widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}}]^{-1}\Big\| > \frac{C_5'}{p}\Big) = O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big);$$

(iii) for $\widehat{\boldsymbol{A}}^{-1} = [\widehat{\boldsymbol{\Sigma}}_x(t) + \widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}}]^{-1}$,

$$P\Big(\Big\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\Big\| > C_5'\Big) = O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big).$$

Proof:

(i) Let $\mathcal{L} = \Big\|\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}} - \boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\Big\|$. Then,

$$\begin{aligned}\mathcal{L} \leq & 2\Big\|\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\Big\| + 2\Big\|\boldsymbol{C}_T^\tau((\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1})\boldsymbol{B}\Big\| \\ & + \Big\|\boldsymbol{B}^\tau((\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1})\boldsymbol{B}\Big\| + \Big\|\boldsymbol{C}_T^\tau\boldsymbol{\Sigma}_u^{-1}\boldsymbol{C}_T\Big\| \\ & + \Big\|\boldsymbol{C}_T^\tau((\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1})\boldsymbol{C}_T\Big\|.\end{aligned}$$

Therefore, by Assumption 2(b), Lemma 2 and Lemma A.3(ii), it is straightforward to verify the result.

(ii) Since $\|\boldsymbol{G}_T\|_F \geq \|\boldsymbol{G}_T\|$, according to Theorem 1, there exists $C > 0$ such that with probability at least $1-O(\frac{1}{T^{1+\epsilon}})$, $\|\boldsymbol{G}_T\| \leq C\sqrt{\frac{\log T}{T}}$. Thus by Lemma A.1 of Fan, Liao and Mincheva[8], for some $c''>0$

$$P\Big(\Big\|\widehat{\boldsymbol{\Sigma}}_x^{-1}(t)-\boldsymbol{\Sigma}_x^{-1}(t)\Big\| < c''\|\boldsymbol{G}_T\|\Big) \geq P\Big(\|\boldsymbol{G}_T\| < C\sqrt{\frac{\log T}{T}}\Big) \geq 1-O\Big(\frac{1}{T^{1+\epsilon}}\Big),$$

which implies

$$P\Big(\Big\|\widehat{\boldsymbol{\Sigma}}_x^{-1}(t)-\boldsymbol{\Sigma}_x^{-1}(t)\Big\| < Cc''\sqrt{\frac{\log T}{T}}\Big) \geq 1-O\Big(\frac{1}{T^{1+\epsilon}}\Big). \tag{A.11}$$

Now let $\widehat{\boldsymbol{A}}=\widehat{\boldsymbol{\Sigma}}_x^{-1}(t)+\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}}$, and $\boldsymbol{A}=\boldsymbol{\Sigma}_x^{-1}(t)+\boldsymbol{B}^\tau(\boldsymbol{\Sigma}_u^{\mathcal{T}})^{-1}\boldsymbol{B}$. Then part(i) and (A.11) imply that

$$P\Big(\Big\|\widehat{\boldsymbol{A}}-\boldsymbol{A}\Big\| < Cc''\sqrt{\frac{\log T}{T}}+C_3pm_p\sqrt{\frac{\log p}{T}}\Big) \geq 1-O\Big(\frac{1}{p^{1+\epsilon}}+\frac{1}{T^{1+\epsilon}}\Big).$$

In addition, $m_p\sqrt{\log p/T}=o(1)$. Hence by Lemma A.1 in Fan, Liao and Mincheva[8] and A.6(ii), for some $C_5'>0$,

$$P\Big(\lambda_{\min}(\widehat{\boldsymbol{A}}) \geq C_5'p\Big) \geq P\Big(\|\widehat{\boldsymbol{A}}-\boldsymbol{A}\| < C_5'p\Big) \geq 1-O\Big(\frac{1}{p^{1+\epsilon}}+\frac{1}{T^{1+\epsilon}}\Big),$$

which implies the desired result.

(iii) By the triangular inequality, $\|\widehat{\boldsymbol{B}}\|_F \leq \|\boldsymbol{C}_T\|+O_p(\sqrt{p})$. Hence Lemma A.3(ii) implies, for some $c_5'>0$,

$$P\Big(\Big\|\widehat{\boldsymbol{B}}\Big\|_F \leq \sqrt{p}\Big) \geq 1-O\Big(\frac{1}{T^{1+\epsilon}}+\frac{1}{p^{1+\epsilon}}\Big).$$

In addition, since $\|\boldsymbol{\Sigma}_u^{-1}\|$ is bounded, it then follows from Lemma 2 that $\|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\|$ is bounded with probability at least $1-O(p^{-2}+T^{-2})$. The result then follows from the fact that

$$P\Big(\Big\|\widehat{\boldsymbol{A}}^{-1}\Big\| > c_5'p^{-1}\Big) = O\Big(\frac{1}{p^{1+\epsilon}}+\frac{1}{T^{1+\epsilon}}\Big),$$

which is shown in part (ii).

Proof of Theorem 2, part (ii). We follow similar arguments as in Fan, Fan and Lv[7]. Using the Sherman-Morrison-Woodbury formula, and denote $\widehat{\boldsymbol{A}}^{-1}=[\widehat{\boldsymbol{\Sigma}}_x^{-1}(t)+\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\widehat{\boldsymbol{B}}]^{-1}$,

$\boldsymbol{A}^{-1} = [\boldsymbol{\Sigma}_x^{-1}(t) + \boldsymbol{B}^\tau \boldsymbol{\Sigma}_u^{-1} \boldsymbol{B}]^{-1}$. We then have

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}}_y^{-1}(t) - \boldsymbol{\Sigma}_y^{-1}(t)\right\| \leq & \left\|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\right\| + \left\|[(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}]\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\right\| \\ & + \left\|\boldsymbol{\Sigma}_u^{-1}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau[(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}]\right\| + \left\|\boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau\boldsymbol{\Sigma}_u^{-1}|\right\| \\ & + \left\|\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}\widehat{\boldsymbol{A}}^{-1}(\widehat{\boldsymbol{B}}^\tau - \boldsymbol{B})\boldsymbol{\Sigma}_u^{-1}\right\| + \left\|\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}(\widehat{\boldsymbol{A}}^{-1} - \boldsymbol{A}^{-1})\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\right\| \\ \equiv & l_1 + l_2 + l_3 + l_4 + l_5 + l_6.\end{aligned}$$

The bound of l_1 is given in Lemma 2. And

$$l_2 \leq \left\|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\right\| \cdot \left\|\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1}\right\|.$$

It follows from Lemma 2, and Lemma A.7(iii) that

$$P\Big(l_2 \leq c_6' m_p \sqrt{\frac{\log p}{T}}\Big) \geq 1 - O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big),$$

and

$$l_3 = \left\|\Big[(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\Big]\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau\boldsymbol{\Sigma}_u^{-1}\right\| \leq \|(\widehat{\boldsymbol{\Sigma}}_u^{\mathcal{T}})^{-1} - \boldsymbol{\Sigma}_u^{-1}\| \cdot \|\widehat{\boldsymbol{B}}\|^2 \cdot \|\widehat{\boldsymbol{A}}^{-1}\| \cdot \|\boldsymbol{\Sigma}_u^{-1}\|.$$

Then

$$P\Big(l_3 \leq c_6' m_p \sqrt{\frac{\log p}{T}}\Big) \geq 1 - O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big),$$

and

$$l_4 = \|\boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{B}} - \boldsymbol{B})\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}^\tau\boldsymbol{\Sigma}_u^{-1}\| \leq \|\boldsymbol{\Sigma}_u^{-1}\| \cdot \|\widehat{\boldsymbol{B}} - \boldsymbol{B}\| \cdot \|\widehat{\boldsymbol{A}}^{-1}\| \cdot \|\widehat{\boldsymbol{B}}^\tau\| \cdot \|\boldsymbol{\Sigma}_u^{-1}\|.$$

Consequently,

$$P\Big(l_4 \leq c_6' \sqrt{\frac{\log p}{T}}\Big) \geq 1 - O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big).$$

The same bound also applies to l_5. Finally,

$$\begin{aligned} l_6 &= \left\|\boldsymbol{\Sigma}_u^{-1}\boldsymbol{B}(\widehat{\boldsymbol{A}}^{-1} - \boldsymbol{A}^{-1})\boldsymbol{B}^\tau\boldsymbol{\Sigma}_u^{-1}\right\| \\ &\leq \|\boldsymbol{\Sigma}_u^{-1}\| \cdot \|\boldsymbol{B}\|^2 \cdot \|\widehat{\boldsymbol{A}}^{-1} - \boldsymbol{A}^{-1}\| \cdot \|\boldsymbol{\Sigma}_u^{-1}\| \\ &\leq \|\boldsymbol{\Sigma}_u^{-1}\|^2 \cdot \|\boldsymbol{B}\|^2 \cdot \|\widehat{\boldsymbol{A}} - \boldsymbol{A}\| \cdot \|\widehat{\boldsymbol{A}}^{-1}\| \cdot \|\boldsymbol{A}^{-1}\|.\end{aligned}$$

Lemma A.7(ii) implies $P\Big(\|\widehat{\boldsymbol{A}}^{-1}\| > c_6' p^{-1}\Big) = O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big)$. We obtain

$$P\Big(l_6 \leq c_6' m_p \sqrt{\frac{\log p}{T}}\Big) \geq 1 - O\Big(\frac{1}{p^{1+\epsilon}} + \frac{1}{T^{1+\epsilon}}\Big).$$

The proof is completed by combining $l_1, \ldots, l_6$:

$$P\left(\left\|\widehat{\boldsymbol{\Sigma}}_y^{-1}(t) - \boldsymbol{\Sigma}_y(t)\right\| \le c_6' m_p \sqrt{\frac{\log p}{T}}\right) \ge 1 - O\Big(\frac{1}{T^{1+\epsilon}} + \frac{1}{p^{1+\epsilon}}\Big).$$

$\square$

Acknowledgments

The work was partially supported by National Natural Science Foundation of China (11731015, 11701116), Innovative Team Project of Ordinary Universities in Guangdong Province (2020WCXTD018) (Yuan Li and Xingfa Zhang) and Guangzhou University Research Fund (YG2020029)(Xingfa Zhang).

References

[1] Bickel, P.J., Levina, E. Covariance regularization by thresholding. *Annals of Statistics*, **36**, 2577-2604 (2008)

[2] Bickel, P.J., Levina, E. Regularized estimation of large covariance matrices. *Annals of Statistics*, **36**, 199-227 (2008)

[3] Cai, T., Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, **106**, 672-684 (2011)

[4] Cai, T., Zhou, H. Optimal rates of convergence for sparse covariance matrix estimation. *Annals of Statistics*, **40**, 2389-2420 (2012)

[5] Fama, E., French, K. The cross-section of expected stock returns. *Journal of Finance*, **47**, 427-465 (1992)

[6] Fama, E., French, K. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**, 3-56 (1993)

[7] Fan, J., Fan, Y., Lv, J. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, **147**, 186-197 (2008)

[8] Fan, J., Liao, Y., Mincheva, M. High-dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, **39**(6), 3320-3356 (2011)

[9] Fan, J., Liao, Y., Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society, Series B*, **75**(4), 603-680 (2013)

[10] Francq, C., Zakoïan, J-M. Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Assosiation*, **104**, 313-324 (2009)

[11] Francq, C., Zakoïan, J-M. *GARCH modes: structure, statistical inference and financial applications*. John Wiley & Sons, New Jersey, 2019

[12] James, W., Stein, C. Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Mathematical Statistics and Probability, Vol. I*, 361-379. University of California Press, Berkeley, 1961

[13] Gao, Z., Tsay, R. A structural-factor approach to modeling high-dimensional time series and space-time data. *Journal of Time Series Analysis*, **40**(3), 343-362 (2019)

[14] Guo S., Box J.L., Zhang W. A Dynamic Structure for High-Dimensional Covariance Matrices and Its Application in Portfolio Allocation. *Journal of the American Statistical Association*, **112**, 235-253 (2017)

[15] Harding, M. Structural estimation of high-dimensional factor models. Unpublished manuscript, Stanford University.

[16] Lam, C., Yao, Q., Bathia,N. Estimation for latent factor models for high-dimensional time series. *Biometrika*, **98**, 901-918 (2011)

[17] Lam, C. and Yao, Q. Factor modeling for high-dimensional time series: Inference for the number of factors. *Annals of Statistics*, **40**, 694-726 (2012)

[18] Ling, S. , and Mcaleer, M. Asymptotic theory for a vector arma-garch model. *Econometric Theory*, 280-310 (2003)

[19] Lindner, A.M. Stationarity, mixing, distributional properties and moments of GARCH(p, q) processes. In *Handbook of financial time series*, 43-69. Springer Berlin Heidelberg, 2009

[20] Liu, W., Xiao, H., Wu, W.B. Probability and moment inequalities under dependence. *Statistica Sinica*, **23(3)**, 1257-1272 (2013)

[21] Markowitz, H.M. Portfolio selection. *Journal of Finance*, **7**, 77-91 (1952)

[22] Markowitz, H.M. *Portfolio Selection: Efficient Diversification of Investments.* John Wiley & Sons, New Jersey, 1959

[23] Pan, J., Yao, Q. Modelling multiple time series via common factors. *Biometrika*, **95**, 356-379 (2008)

[24] Rothman, A.J., Levina, E., Zhu, J. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, **104**, 177-186 (2009)

[25] Tsay, R.S. *Multivariate Time Series Analysis.* John Wiley & Sons, New Jersey, 2014