Nonparametric Estimation of Large Covariance Matrices with Conditional Sparsity

Hanchao Wang[*], Bin Peng[†], Degui Li[‡], Chenlei Leng[§]

Version: September 10, 2020

Abstract

This paper studies estimation of covariance matrices with conditional sparse structure. We overcome the challenge of estimating dense matrices using a factor structure, the challenge of estimating large-dimensional matrices by postulating sparsity on covariance of random noises, and the challenge of estimating varying matrices by allowing factor loadings to smoothly change. A kernel-weighted estimation approach combined with generalised shrinkage is proposed. Under some technical conditions, we derive uniform consistency for the developed estimation method and obtain convergence rates. Numerical studies including simulation and an empirical application are presented to examine the finite-sample performance of the developed methodology.

Keywords: Approximate factor model, Kernel estimation, Large covariance matrix, Sparsity, Uniform convergence

JEL classification: C13, C23, G11

[*]Zhongtai Securities Institute for Financial Studies, Shandong University, China. Email: wanghanchao@sdu.edu.cn.
[†]Department of Econometrics and Business Statistics, Monash University, Australia. Email: bin.peng@monash.edu.
[‡]Department of Mathematics, University of York, UK. Email: degui.li@york.ac.uk, the corresponding author.
[§]Department of Statistics, University of Warwick, UK. Email: C.Leng@warwick.ac.uk.

1 Introduction

Covariance matrix estimation is one of the central problems in high-dimensional statistics and big data analytics. It has applications in a variety of fields including economics, finance, health science and social networks. The sample covariance matrix often performs very poorly in finite samples when the matrix size is large, and becomes singular when the dimension exceeds the sample size, making it difficult to use in high-dimensional statistical inference. A popular approach in the literature is to impose certain structural assumptions on the covariance matrix and then modify the sample covariance matrix. Among the assumptions typically made, the approximate sparsity assumption is probably the most commonly-used, under which many entries in the covariance matrix are assumed zero or close to zero. Consequently, various regularization techniques, such as thresholding or other generalised shrinkage methods, have been introduced; see, for example, Bickel and Levina (2008a,b), Lam and Fan (2009), Rothman, Levina and Zhu (2009) and Cai and Liu (2011). For a comprehensive survey of recent developments on estimating large sparse covariance matrices, we refer to Pourahmadi (2013), Cai, Ren and Zhou (2016) and Fan, Liao and Liu (2016).

The sparsity assumption imposed on the covariance structure is too restrictive or even unrealistic for many datasets in economics and finance where the variables are often highly correlated. For example, co-movement of many macroeconomic variables may be driven by low-dimensional latent factor processes, and many financial time series data collected in the stock market are usually determined by common market factors, both resulting in highly correlated variables. To relax the sparsity assumption for estimating meaningful covariance matrices in these applications, the following approximate factor model (e.g., Chamberlain and Rothschild, 1983) is often employed:

$$X_t = \mathbf{B}F_t + u_t = \chi_t + u_t, \quad t = 1, \cdots, T, \tag{1.1}$$

where X_t is an N-dimensional vector of stationary variables, $\mathbf{B}$ is an $N \times K$ matrix of factor loadings, F_t is a K-dimensional vector of stationary latent factors, u_t is an N-dimensional vector of idiosyncratic errors, and K is the number of latent factors. The factor model postulates that X_t is decomposed as the common component $\chi_t = \mathbf{B}F_t$ and the error component u_t. Instead of directly imposing sparsity on the covariance structure of X_t, Fan, Liao and Mincheva (2013) assume that the error covariance matrix is sparse, giving rise to the so-called conditional sparsity structure for X_t. They introduce a novel covariance matrix estimation technique by thresholding principal orthogonal complements.

The large covariance matrix in Fan, Liao and Mincheva (2013) is assumed to be *static* with constant entries. This assumption can be invalid when data is collected over a long time span. Indeed, it is not uncommon that economic or financial time series variables are often subject to abrupt structural breaks or smooth structural change over a long time period. Hence, in recent years, there has been increasing interest in estimating large *dynamic* covariance matrices by allowing

their entries to vary smoothly with certain index variable(s) or time, see, for example, Chen, Xu and Wu (2013), Chen and Leng (2016) and Chen, Li and Linton (2019). These papers work under the sparsity assumption and thus cannot handle the problem of estimating non-sparse covariance matrices. To model large, dynamic, and non-sparse matrices, we propose to allow the factor loading matrix $\mathbf{B}$ in (1.1) to be time-dependent by writing

$$X_t = \mathbf{B}_t F_t + u_t, \quad t = 1, \cdots, T, \tag{1.2}$$

where $\mathbf{B}_t$ is an $N \times K$ matrix of time-varying factor loadings, and the remaining components are the same as those in (1.1). In this paper, we mainly consider the case where each factor loading is a smooth function of a univariate stationary index variable z_t, i.e., $\mathbf{B}_t = \mathbf{B}(z_t) = (B_{ik}(z_t))_{N\times K}$. The factor model (1.2) stimulates two research directions which are closely related. One is to study structural instabilities in the factor loadings. Towards this, several estimation and detection methods have been proposed to locate break points and determine break number (e.g., Stock and Watson, 2009; Breitung and Eickmeier, 2011; Chen, Dolado and Gonzalo, 2014; Cheng, Liao and Schorfheide, 2016; Barigozzi, Cho and Fryzlewicz, 2018; Ma and Su, 2018). The other direction is to consider estimating and testing the state-varying factor model (1.2) with factor loadings relying on a state process (Pelger and Xiong, 2020) or the time-varying factor model (1.2) with $\mathbf{B}_t$ defined as functions of scaled time (Motta, Hafner and von Sachs, 2011; Su and Wang, 2017). In particular, Pelger and Xiong (2020) give a few examples, showing that the time-dependent factor model (1.2) allows for a more parsimonious representation of the data than the conventional factor model (1.1).

In practice, it is often important to study how the second-order moment structure of the variables responds to technological innovation and changes in policy, business circles and economic conditions. However, to the best of our knowledge, there is virtually no work on estimating the covariance matrix of X_t defined in (1.2), which is more challenging than only estimating $\mathbf{B}_t$ and F_t. This paper aims to fill this gap. In particular, we propose a kernel-based local estimation method to estimate the large dynamic covariance matrices with conditional sparsity, extending the Principal Orthogonal complEment Thresholding (POET) methodology in Fan, Liao and Mincheva (2013) developed for estimating static covariance matrices. In the proposed estimation procedure, we first estimate the factors and their loadings by a local version of principal component analysis as in Su and Wang (2017) and Pelger and Xiong (2020), and then apply the generalised shrinkage method to estimate the sparse error covariance structure. As the idiosyncratic errors u_t are unobservable, we use their approximation in the large covariance matrix estimation with entries varying over time. This is substantially different from that in the literature (e.g., Chen and Leng, 2016; Chen, Li and Linton, 2019), leading to difficulties in the subsequent theoretical justification.

To measure the uniform distance between the estimated covariance matrices and the true ones, we derive uniform consistency results for the developed matrix estimators under appropriate matrix norms and regularity conditions, and obtain uniform convergence rates that are comparable

to those in the existing literature (e.g., Fan, Liao and Mincheva, 2013; Chen and Leng, 2016). In particular, the uniform consistency results are derived over an expanding set whose size is divergent to infinity as the sample size increases. As a consequence, the commonly-used compact support restriction on the index variable (e.g., Chen and Leng, 2016; Chen, Li and Linton, 2019) is removed in the technical assumptions. This relaxation enhances the applicability of the developed asymptotic results, but makes their technical proofs much more complicated than those in the literature. In addition, we extend the ratio criterion introduced by Lam and Yao (2012) and Ahn and Horenstein (2013) to estimate the number of factors in (1.2), and prove its consistency. An easy-to-implement method is proposed to select the tuning parameter in the generalised shrinkage technique, allowing temporal dependence for X_t.

The present paper is partly motivated by the empirical study in constructing minimum variance portfolio for vast financial time series, where large dynamic covariance matrices and their inverse (i.e., precision matrices) are often preferred to the static ones (e.g., Guo, Box and Zhang, 2017; Chen, Li and Linton, 2019; Engle, Ledoit and Wolf, 2019). In this paper, we provide further empirical evidence to this subject, applying the developed model framework and covariance matrix estimation methodology to analyse daily returns of 319 companies listed in the S&P 500 index and estimate their dynamic covariance structure conditioning on the CBOE volatility index. Subsequently, we construct the out-of-sample minimum variance portfolio making use of the estimated covariance matrix, and compute the standard deviation, the information ratio as well as the Sharpe ratio, from which we find that our method outperforms that using Fan, Liao and Mincheva (2013)'s POET method.

The rest of the paper is organised as follows. Section 2 introduces the model setting and kernel-weighted least squares method to estimate the covariance matrix of X_t. Section 3 lists some regularity conditions and states the main asymptotic results. Section 4 discusses selection of the factor number and the tuning parameter for shrinkage. Section 5 reports both simulation and empirical studies. Section 6 concludes the paper. Proofs of the main theorems are given in Appendix A. The supplemental document contains proofs of some technical lemmas as well as additional numerical results. Throughout the paper, for a square matrix $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote its maximum and minimum eigenvalues, respectively, and $\mathsf{trace}\{\mathbf{A}\}$ denotes its trace. Define $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_F = \mathsf{trace}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_{\max} = \max_{ij}|a_{ij}|$ for $\mathbf{A} = (a_{ij})_{N\times N}$, and $\|\mathbf{A}\|_{\mathbf{\Sigma}} = \frac{1}{\sqrt{N}}\|\mathbf{\Sigma}^{-1/2}\mathbf{A}\mathbf{\Sigma}^{-1/2}\|_F$, where $\mathbf{\Sigma}$ is a given $N \times N$ positive definite matrix. Let $\wedge$ and $\vee$ denote minimum and maximum, respectively; and let $a_n \propto b_n$ denote that $a_n = cb_n$ for a positive constant c.

2 Model and estimation methodology

In this section, we first introduce the model setting and discuss some identification issues, and then combine kernel-weighted least squares with generalised shrinkage to estimate large dynamic covariance matrices.

2.1 Model setting

As introduced in Section 1, we let $\mathbf{B}_t = \mathbf{B}(z_t)$ and rewrite the time-dependent factor model (1.2) as

$$X_t = \mathbf{B}(z_t)F_t + u_t, \quad t = 1, \cdots, T, \tag{2.1}$$

where we allow the univariate index variable z_t to be supported on an expanding set. From (2.1), we readily have

$$\mathbf{\Sigma}_X(z) = \mathbf{B}(z)\mathbf{\Sigma}_F\mathbf{B}(z)' + \mathbf{\Sigma}_u(z) \tag{2.2}$$

for $z_t = z$, where $\mathbf{\Sigma}_X(z)$ and $\mathbf{\Sigma}_u(z)$ denote the conditional covariance matrices of X_t and u_t, respectively, given $z_t = z$, and

$$\mathbf{\Sigma}_F(z) := \mathsf{E}\left(F_tF_t'|z_t = z\right) = \mathsf{E}\left(F_tF_t'\right) =: \mathbf{\Sigma}_F \tag{2.3}$$

is assumed to be time-invariant and positive definite. For (2.2), F_t and u_t are assumed to be conditionally uncorrelated (given z_t), to be consistent with the literature. In fact, when the conditional covariance matrix $\mathbf{\Sigma}_F(z) = \mathsf{E}\left(F_tF_t'|z_t = z\right)$ varies with z, letting $F_t^\star = \mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_F^{-1/2}(z_t)F_t$ and $\mathbf{B}^\star(z_t) = \mathbf{B}(z_t)\mathbf{\Sigma}_F^{1/2}(z_t)\mathbf{\Sigma}_F^{-1/2}$, we may show that the conditional covariance matrix $\mathsf{E}\left(F_t^*F_t^{*\prime}|z_t = z\right)$ is time-invariant, and (2.1) and (2.2) still hold with $\mathbf{B}^\star(\cdot)$ and $F_t^\star$ replacing $\mathbf{B}(\cdot)$ and F_t, respectively. In practical implementation, without loss of generality, we may let $\mathbf{\Sigma}_F = \mathbf{I}_K$, a $K \times K$ identity matrix, as one identification restriction in the following kernel-weighted principal component analysis, see (2.6). In addition, we assume that F_t and u_t have zero-mean conditioning on z_t, indicating that $\mathbf{\Sigma}_X(z)$ and $\mathbf{\Sigma}_u(z)$ in (2.2) equal to their respective conditional second moments.

As in the classic factor model (1.1), the time-varying factor loadings and common factors in (2.1) are only identified up to a $K \times K$ invertible matrix $\mathbf{Q}(z_t)$ depending on z_t, indicating that K^2 restrictions are needed for model identification given $z_t = z$. Similarly to Assumption A.1(ii) and (iii) in Su and Wang (2017) and Assumptions 3 and 4 in Pelger and Xiong (2020), we impose full-rank conditional second moment conditions on F_t and $\mathbf{B}(z_t)$ in order to develop sensible asymptotic results, see (3.1) and (3.2) in Assumption 2 below. Meanwhile, another identification issue arises due to the presence of two time series processes z_t and F_t in the common component of model (2.1)[1]. From (2.3), the conditional covariance matrix $\mathsf{E}\left(F_tF_t'|z_t = z\right)$ does not rely on z_t.

[1]We thank a referee for pointing out this issue.

Hence, Assumption 1 in Pelger and Xiong (2020) is satisfied, ensuring that the index variable z_t and the factor F_t are separable and the model is identifiable. Section 2.3 of Pelger and Xiong (2020) provides some further examples to illustrate this identification condition and the relevant factor model representation.

2.2 Kernel-weighted covariance matrix estimation

Let $K(\cdot)$ be a kernel function and b be a bandwidth which tends to zero as T goes to infinity. Multiplying both sides of (2.1) by

$$K_{t,b}(z) = K_b\left(z_t - z\right) / \left[\frac{1}{T}\sum_{s=1}^{T} K_b\left(z_s - z\right)\right]$$

with $K_b(\cdot) = \frac{1}{b}K(\cdot/b)$, we immediately have the following local approximation:

$$X_t K_{t,b}^{1/2}(z) \approx \mathbf{B}(z)F_t K_{t,b}^{1/2}(z) + u_t K_{t,b}^{1/2}(z) \ \text{ when } \ z_t \approx z. \tag{2.4}$$

Let $X_t(z) = X_t K_{t,b}^{1/2}(z)$ and define the kernel-weighted sample covariance matrix:

$$\overline{\mathbf{\Sigma}}_X(z) = \frac{1}{T}\sum_{t=1}^{T} X_t(z)X_t(z)' = \left[\frac{1}{T}\sum_{t=1}^{T} X_t X_t' K_b\left(z_t - z\right)\right] / \left[\frac{1}{T}\sum_{t=1}^{T} K_b\left(z_t - z\right)\right]. \tag{2.5}$$

However, when the dimension N is large, the above kernel-weighted sample covariance matrix will be ill-conditioned, resulting in poor estimation for $\mathbf{\Sigma}_X(z)$. To address this, we next combine a local principal component analysis with a generalised shrinkage technique to construct meaningful covariance matrix estimation.

The number of common factors, K, is assumed to be known for the time being and will be determined by a ratio criterion in Section 4 when it is unknown. Let $\mathbf{F}(z) = \left[F_1 K_{1,b}^{1/2}(z), \cdots, F_T K_{T,b}^{1/2}(z)\right]'$, $\mathbf{B}(z) = [B_1(z), \cdots, B_N(z)]'$ and $\mathbf{X}(z) = [X_1(z), \cdots, X_T(z)]$, where $B_i(z) = [B_{i1}(z), \cdots, B_{iK}(z)]'$. Consider the following identification condition:

$$\frac{1}{T}\mathbf{F}(z)'\mathbf{F}(z) = \mathbf{I}_K, \ \ \frac{1}{N}\mathbf{B}(z)'\mathbf{B}(z) \ \text{ is diagonal.} \tag{2.6}$$

Motivated by (2.4), we define the kernel-weighted local least squares objective function:

$$\sum_{t=1}^{T}\left(X_t - \mathbf{B}_* F_{t*}\right)'\left(X_t - \mathbf{B}_* F_{t*}\right) K_{t,b}\left(z\right) = \|\mathbf{X}(z) - \mathbf{B}_*\mathbf{F}_*'\|_F^2, \tag{2.7}$$

where $\mathbf{B}_*$ and $\mathbf{F}_*$ are general notation for $N \times K$ and $T \times K$ matrices, respectively. Making use of

the identification condition (2.6), given $\mathbf{F}_*$, we estimate $\mathbf{B}(z)$ by $\frac{1}{T}\mathbf{X}(z)\mathbf{F}_*$, and consequently the objective function in (2.7) becomes

$$\left\|\mathbf{X}(z) - \frac{1}{T}\mathbf{X}(z)\mathbf{F}_*\mathbf{F}_*'\right\|_F^2 = \text{trace}\{\mathbf{X}(z)'\mathbf{X}(z)\} - \frac{1}{T}\text{trace}\{\mathbf{F}_*'\mathbf{X}(z)'\mathbf{X}(z)\mathbf{F}_*\}. \tag{2.8}$$

From (2.8), minimisation of (2.7) subject to the restriction (2.6) is equivalent to the maximisation of $\text{trace}\{\mathbf{F}_*'\mathbf{X}(z)'\mathbf{X}(z)\mathbf{F}_*\}$ subject to $\frac{1}{T}\mathbf{F}_*'\mathbf{F}_* = \mathbf{I}_K$. Hence, consider eigen-analysis of the $T \times T$ kernel-weighted matrix $\mathbf{X}(z)'\mathbf{X}(z)$ as in Su and Wang (2017) and Pelger and Xiong (2020) and let

$$\widetilde{\mathbf{F}}(z) = \left[\widetilde{F}_1(z), \cdots, \widetilde{F}_T(z)\right]'$$

be the K eigenvectors (multiplied by $\sqrt{T}$) of the matrix $\mathbf{X}(z)'\mathbf{X}(z)$ corresponding to the K largest eigenvalues. The factor loading matrix can be estimated as

$$\widetilde{\mathbf{B}}(z) = \frac{1}{T}\mathbf{X}(z)\widetilde{\mathbf{F}}(z) = \left[\widetilde{B}_1(z), \cdots, \widetilde{B}_N(z)\right]'.$$

Then, we can approximate the local residuals $u_t K_{t,b}^{1/2}(z)$ by

$$\widetilde{u}_t(z) = [\widetilde{u}_{1t}(z), \cdots, \widetilde{u}_{Nt}(z)]' \quad \text{with} \quad \widetilde{u}_{it}(z) = X_{it}(z) - \widetilde{B}_i(z)'\widetilde{F}_t(z),$$

where $X_{it}(z)$ is the i-th element in $X_t(z)$. A naive method of estimating $\mathbf{\Sigma}_u(z)$ is to directly calculate the sample covariance matrix of $\widetilde{u}_t(z)$:

$$\check{\mathbf{\Sigma}}_u(z) = [\check{\sigma}_{u,ij}(z)]_{N\times N} = \frac{1}{T}\sum_{t=1}^{T}\widetilde{u}_t(z)\widetilde{u}_t(z)', \tag{2.9}$$

which is usually unstable when the number N is large. To address this problem, we apply the generalised shrinkage technique to the off-diagonal elements in $\check{\mathbf{\Sigma}}_u(z)$ (c.f., Chen and Leng, 2016) and estimate $\mathbf{\Sigma}_u(z)$ by

$$\widetilde{\mathbf{\Sigma}}_u(z) = [\widetilde{\sigma}_{u,ij}(z)]_{N\times N}, \quad \widetilde{\sigma}_{u,ij}(z) = \begin{cases} \check{\sigma}_{u,ii}(z), & i = j, \\ s_{\rho(z)}\left(\check{\sigma}_{u,ij}(z)\right), & i \neq j, \end{cases} \tag{2.10}$$

where $s_\rho(\cdot)$ is a shrinkage function satisfying the following three restrictions: (i) $|s_\rho(w)| \leqslant |w|$ for $w \in \mathbb{R}$; (ii) $s_\rho(w) = 0$ if $|w| \leqslant \rho$; (iii) $|s_\rho(w) - w| \leqslant \rho$, with ρ being a tuning parameter. The shrinkage function satisfying these three restrictions covers the hard thresholding, soft thresholding and SCAD function, all of which are commonly used in the literature. Note that the tuning parameter $\rho(z)$ in (2.10) is allowed to change with the index variable, which is needed in the context of dynamic covariance matrix estimation. Combining the above estimates, we finally estimate $\mathbf{\Sigma}_X(z)$

by

$$\widetilde{\mathbf{\Sigma}}_X(z) = \widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' + \widetilde{\mathbf{\Sigma}}_u(z). \tag{2.11}$$

Proposition 1 of Appendix A.1 shows the equivalence of the kernel-weighted estimators defined in (2.10) and (2.11) and the local POET estimators to be defined in Appendix A.1. Hence, the method developed above can be seen as an extension of Fan, Liao and Mincheva (2013)'s POET method to the more general nonparametric model setting. We focus on the kernel-weighted estimators in the main text of this paper.

3 Large-sample theory

In this section, we first give some technical assumptions and then state the uniform consistency results for the kernel-weighted covariance matrix estimator defined in Section 2.

3.1 Technical assumptions

We allow temporal dependence on the high-dimensional data by assuming the stationary process $\{(F_t, z_t, u_t) : t \geqslant 0\}$ to be α-mixing with the mixing coefficient $\alpha(n) \to 0$ as $n \to \infty$, where

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

with $\mathcal{F}_s^t$ denoting a σ-algebra generated by $\{(F_i, z_i, u_i) : s \leqslant i \leqslant t\}$. The α-mixing dependence condition is mild for a stationary and weakly dependent process and is satisfied by commonly-used time series models such as the vector ARMA process. We next give some regularity conditions, which are needed to derive the main asymptotic results.

Assumption 1. *(a) The α-mixing coefficient decays to 0 at a geometric rate, i.e., $\alpha(n) \leqslant C_0\gamma^n$ with $0 < \gamma < 1$ and C_0 being a positive constant.*

(b) There exists a density function $f(z)$ for the index variable z_t, which is twice continuously differentiable for $z \in \mathbb{R}$ and satisfies that $\inf_{|z| \leqslant L_T} f(z) \geqslant a_T(f) > 0$, where L_T may diverge to infinity and $a_T(f)$ may converge to zero as $T \to \infty$. In addition, the joint density function for (z_1, z_t) exists and is bounded uniformly over $t \geqslant 2$.

Assumption 2. *(a) The factor loading functions $B_i(z)$ are Lipschitz continuous and bounded uniformly over $i = 1, \cdots, N$ and $z \in [-L_T, L_T]$ with L_T defined in Assumption 1(b). There exists a $K \times K$ positive definite matrix $\mathbf{\Sigma}_B(z)$ with uniformly bounded eigenvalues, such that*

$$\sup_{|z| \leqslant L_T} \left\| \frac{1}{N} \sum_{i=1}^{N} B_i(z) B_i(z)' - \mathbf{\Sigma}_B(z) \right\| = o(1). \tag{3.1}$$

(b) The kernel-weighted common factors $F_t(z) := F_t K_{t,b}^{1/2}(z)$ *satisfy that*

$$\sup_{|z|\leqslant L_T} \left\| \frac{1}{T}\sum_{t=1}^{T} F_t(z)F_t(z)' - \mathbf{\Sigma}_F \right\| = O_P\left(a_T^{-1}(f)\left(\sqrt{\frac{\log T}{Tb}} + b^2\right)\right), \tag{3.2}$$

and $\frac{1}{T}\sum_{t=1}^{T}\|F_t(z)\|^2 = O_P(1)$ *uniformly over* $-L_T \leqslant z \leqslant L_T$*, where* $\mathbf{\Sigma}_F$ *is defined in (2.3).*

(c) The error covariance matrix $\mathbf{\Sigma}_u(z)$ *is positive definite, satisfying that*

$$0 < \underline{c} \leqslant \inf_{|z|\leqslant L_T} \lambda_{\min}(\mathbf{\Sigma}_u(z)) \leqslant \sup_{|z|\leqslant L_T} \lambda_{\max}(\mathbf{\Sigma}_u(z)) \leqslant \overline{c} < \infty,$$

where $\underline{c}$ *and* $\overline{c}$ *are two positive constants. The* (i,j)*-entry of* $\mathbf{\Sigma}_u(z)$*,* $\sigma_{u,ij}(z)$*, is bounded and twice continuously differentiable uniformly over* $z \in \mathbb{R}$ *and* $1 \leqslant i,j \leqslant N$*.*

Assumption 3. *(a) There exist constants* $C_1 > 0$ *and* $\theta_1 > 0$ *such that*

$$\max_{1\leqslant i\leqslant N} E\left[e^{\theta u_{it}^2}\right] \leqslant C_1 \text{ for } 0 < \theta \leqslant \theta_1, \tag{3.3}$$

$$\max_{1\leqslant j\leqslant K} E\left[e^{\theta F_{jt}^2}\right] \leqslant C_1 \text{ for } 0 < \theta \leqslant \theta_1. \tag{3.4}$$

(b) Let $E(u_{it}|z_t, F_t) = 0$ *a.s. for any* i*, and there exists a constant* $C_2 > 0$ *such that*

$$\max_{1\leqslant t,s\leqslant T} E\left[(u_s'u_t - E[u_s'u_t])^4\right] \leqslant C_2 N^2. \tag{3.5}$$

In addition,

$$\max_{1\leqslant t,s\leqslant T} \|\mathbf{B}(z_s)'u_t\|^2 = O_P(N\log T). \tag{3.6}$$

Assumption 4. *(a) The kernel function* $K(\cdot)$ *is symmetric and Lipschitz continuous, and has a compact support.*

(b) Let the bandwidth b *satisfy that* $b \propto T^{-\theta_2}$ *and* $T^{1-2\iota}b/\{\log^3(N \vee T)\} \to \infty$*, where* θ_2 *and* ι *are two positive constants. In addition,* $NT\exp\{-\theta T^{\iota}\} = o(1)$ *for some* $0 < \theta \leqslant \theta_1$ *with* θ_1 *defined in Assumption 3(a).*

(c) Let the tuning parameter in the generalised shrinkage method be chosen as

$$\rho(z) = M(z)a_T^{-1}(f)\omega(T,N,b) \text{ with } \omega(T,N,b) = \sqrt{\frac{\log T}{Nb} + \frac{\log(N\vee T)}{Tb} + b^2},$$

where $M(z)$ *is a positive function satisfying that* $0 < \underline{M} \leqslant \inf_{|z|\leqslant L_T} M(z) \leqslant \sup_{|z|\leqslant L_T} M(z) \leqslant \overline{M} < \infty$*, and* $\underline{M}$ *is sufficiently large. In addition,* $a_T^{-1}(f)\omega(T,N,b) = o(1)$*.*

Remark 1. (a) The above technical assumptions are mild and justifiable, although some of them can

be weakened at the cost of more lengthy proofs. The exponential convergence rate for the α-mixing coefficient in Assumption 1(a) can be replaced by a polynomial rate if the dimension N diverges at a polynomial rate of the time series length T. Assumption 1(b) imposes some smoothness conditions on the density function. In particular, we remove the compact support restriction on the stationary index variable, which is rather restrictive but commonly used in proving the uniform consistency of the kernel-weighted large covariance matrix estimators (e.g., Chen and Leng, 2016; Chen, Li and Linton, 2019). Conditions similar to Assumption 1(b) can be found in Hansen (2008) and Li, Lu and Linton (2012), both of which consider uniform consistency of the kernel-based nonparametric estimators over an expanding set.

(b) Assumption 2(a) imposes some smoothness conditions on the factor loading functions, which are not uncommon when the kernel smoothing technique is applied. They are similar to some smoothness conditions assumed in the literature on time-varying or state-varying factor models (e.g., Su and Wang, 2017; Pelger and Xiong, 2020). When $\lambda_{\min}(\mathbf{\Sigma}_B(z))$ is bounded away from zero uniformly over z, from (3.1), we readily have that $\lambda_{\min}(\mathbf{B}(z)'\mathbf{B}(z))$ diverges at a rate N, indicating that all the factors are pervasive. Assumption 2(b) can be verified by using (2.3) and applying some classic uniform consistency result to the kernel-weighted factor process under the α-mixing dependence assumption (e.g., Theorem 8 in Hansen, 2008). The asymptotic properties in Section 3.2 remain valid if we replace $\mathbf{\Sigma}_F$ by $\mathbf{\Sigma}_F(z)$ in (3.2). When the main interest lies in estimation and identification of factors and factor loadings, we need to impose an additional restriction that the K eigenvalues of $\mathbf{\Sigma}_B(z)\mathbf{\Sigma}_F$ are distinct for any z. Assumption 2(c) is a natural extension of Assumption 3.2(ii) in Fan, Liao and Mincheva (2013), ensuring that $\mathbf{\Sigma}_u(z)$ is well conditioned uniformly over z.

(c) The moment conditions in (3.3) and (3.4) are similar to those in Bickel and Levina (2008a,b) and Chen and Leng (2016), and may be replaced by the following weaker conditions:

$$\max_{1\leqslant i\leqslant N} \mathsf{E}\left[|u_{it}|^{2\theta}\right] < C_1^* < \infty, \quad \max_{1\leqslant j\leqslant K} \mathsf{E}\left[|F_{jt}|^{2\theta}\right] < C_1^* < \infty$$

for $\theta > 0$ sufficiently large if the dimension N diverges at a polynomial rate of T. Appendix C in the supplemental document discusses the relevant asymptotic theorems and sketches their proofs under these weaker moment conditions. The conditions (3.5) and (3.6) in Assumption 3(b) are sensible and similar to those in the literature (e.g., Bai and Ng, 2002; Fan, Liao and Mincheva, 2013; Su and Wang, 2017).

(d) Assumptions 4(a)(b) impose some commonly-used conditions on the kernel function and bandwidth, and the choice of the tuning parameter in Assumption 4(c) ensures the validity of the shrinkage method in large covariance matrix estimation. In particular, the condition $NT\exp\{-\theta T^{\iota}\} = o(1)$ in Assumption 4(b) indicates that the number of variables N is allowed to diverge at an exponential rate of T. When N diverges at a polynomial rate of T, we readily have $NT\exp\{-\theta T^{\iota}\} = o(1)$ by choosing ι and θ as any positive constants. Meanwhile, the restriction

$\frac{T^{1-2\iota}b}{\log^3(N\vee T)} \to \infty$ is satisfied if $2\iota + \theta_2 < 1$. This indicates that θ_2 can be chosen from $(0,\ 1)$ (by letting ι be sufficiently small), covering the order of the optimal bandwidth in conventional kernel-based estimation with univariate index variable.

3.2 Uniform consistency

As introduced in Section 1, we assume that the error covariance matrix is approximately sparse, i.e., $\mathbf{\Sigma}_u(\cdot) \in \mathcal{U}(q, m_N, M_0)$, where $\mathcal{U}(q, m_N, M_0)$ is defined by

$$\mathcal{U}(q, m_N, M_0) = \left\{ \mathbf{\Sigma}(\cdot) = [\sigma_{ij}(\cdot)]_{N\times N} \ \Big| \ \sup_{z\in\mathbb{R}} \sigma_{ii}(z) \leqslant M_0,\ \sup_{z\in\mathbb{R}} \left(\sum_{j=1}^{N} |\sigma_{ij}(z)|^q \right) \leqslant m_N\ \forall i \right\},$$

where $0 \leqslant q < 1$ and M_0 is a positive constant. For the special case of $q = 0$,

$$\mathcal{U}(0, m_N, M_0) = \left\{ \mathbf{\Sigma}(\cdot) = [\sigma_{ij}(\cdot)]_{N\times N} \ \Big| \sup_{z\in\mathbb{R}} \sigma_{ii}(z) \leqslant M_0,\ \sup_{z\in\mathbb{R}} \sum_{j=1}^{N} I(\sigma_{ij}(z) \neq 0) \leqslant m_N\ \forall i \right\},$$

and consequently $\mathbf{\Sigma}_u(\cdot) \in \mathcal{U}(0, m_N, M_0)$, which is called the exact sparsity assumption uniformly over z. The above definition is similar to that in Chen and Leng (2016) and Chen, Li and Linton (2019), a natural extension of the classic sparsity assumption used by Bickel and Levina (2008a), Rothman, Levina and Zhu (2009) and Cai and Liu (2011). The following theorem gives the uniform consistency (in the operator norm) of $\widetilde{\mathbf{\Sigma}}_u(z)$ defined in (2.10).

Theorem 1. *Suppose that Assumptions 1–4 are satisfied and* $\mathbf{\Sigma}_u(\cdot) \in \mathcal{U}(q, m_N, L)$. *Then, as both* N *and* T *tend to infinity jointly, we have*

$$\sup_{|z|\leqslant L_T} \left\| \widetilde{\mathbf{\Sigma}}_u(z) - \mathbf{\Sigma}_u(z) \right\| = O_P\left(m_N \left[\omega(T, N, b)/a_T(f)\right]^{1-q} \right), \tag{3.7}$$

where $a_T(f)$ *is defined in Assumption 1(b) and* $\omega(T, N, b)$ *is defined in Assumption 4(c).*

Remark 2. (a) The uniform consistency for $\widetilde{\mathbf{\Sigma}}_u(z)$ is achieved by assuming that

$$m_N \left[\omega(T, N, b)/a_T(f)\right]^{1-q} = o(1)$$

and letting both N and T tend to infinity simultaneously. The latter is a typical setting in large panel data analysis (e.g., Bai and Ng, 2002; Fan, Liao and Mincheva, 2013), ensuring that kernel-based estimates of the latent factors and factor loading functions are consistent (up to an appropriate rotation). The uniform consistency result in (3.7) is derived on an expanding set and the lower bound of the density function $a_T(f)$ thus affects the convergence rate. If L_T diverges to infinity (as T increases), $a_T(f)$ converges slowly to zero at an appropriate rate, slowing down the convergence

rate. This is similar to the uniform consistency results developed by Hansen (2008) and Li, Lu and Linton (2012) for the kernel-based estimation in low-dimensional setting.

(b) The order b^2 in $\omega(T, N, b)$ is due to the asymptotic bias of the kernel estimation, and can be removed if we assume that $Tb^3/\log(N \vee T) \to 0$ (which is similar to the condition $Tb^3 \to 0$ in Motta, Hafner and von Sachs, 2011). Furthermore, assuming that $m_N < \infty$, $q = 0$, L_T is a positive constant (independent of T) and $a_T(f) = a_0 > 0$, the uniform convergence rate in (3.7) becomes $\left(\frac{\log T}{Nb}\right)^{1/2}$ when $N/T \to 0$, and $\left(\frac{\log T}{Tb}\right)^{1/2}$ (a typical uniform convergence rate in the kernel-based nonparametric estimation) when $T/N \to 0$ and N diverges at a polynomial rate of T.

(c) The uniform rate of convergence in (3.7) is slower than those in Chen and Leng (2016) and Chen, Li and Linton (2019). The additional order $\frac{\log T}{Nb}$ in $\omega(T, N, b)$ is mainly contributed by the uniform estimation errors for $\widetilde{F}_t(z)$ and $\widetilde{B}_i(z)$, see the proof of Lemma 2 in Appendix B. Meanwhile, we remove the compact support restriction on the index variable z_t, and derive the uniform consistency result over a wider region than that in Chen and Leng (2016) and Chen, Li and Linton (2019). This further slows down the uniform convergence rate as discussed in Remark 2(a) above.

We next state the uniform consistency of $\widetilde{\mathbf{\Sigma}}_X(z)$ defined in (2.11). As the first K eigenvalues are very spiked (diverging at a rate N), the large covariance matrix $\mathbf{\Sigma}_X(z)$ cannot be consistently estimated in the absolute term. Motivated by Fan, Fan and Lv (2008) and Fan, Liao and Mincheva (2013), we measure the covariance matrix estimate in the relative error and consider

$$\left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z)\right\|_{\mathbf{\Sigma}_X(z)} = \frac{1}{\sqrt{N}} \left\|\mathbf{\Sigma}_X^{-1/2}(z)\widetilde{\mathbf{\Sigma}}_X(z)\mathbf{\Sigma}_X^{-1/2}(z) - \mathbf{I}_N\right\|_F .$$

The following theorem gives the uniform rates of convergence for $\widetilde{\mathbf{\Sigma}}_X(z)$ in both the relative error and max norm $\|\cdot\|_{\max}$.

Theorem 2. *Suppose that the assumptions of Theorem 1 are satisfied. Then, as both* N *and* T *tend to infinity jointly, we have*

$$\sup_{|z|\leqslant L_T} \left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z)\right\|_{\mathbf{\Sigma}_X(z)} = O_P\left(N^{1/2}\left[\omega(T, N, b)/a_T(f)\right]^2 + m_N\left[\omega(T, N, b)/a_T(f)\right]^{1-q}\right), \quad (3.8)$$

and

$$\sup_{|z|\leqslant L_T} \left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z)\right\|_{\max} = O_P\left(\omega(T, N, b)/a_T(f)\right), \quad (3.9)$$

where $a_T(f)$ *is defined in Assumption 1(b) and* $\omega(T, N, b)$ *is defined in Assumption 4(c).*

Remark 3. (a) As in Theorem 1, both N and T diverge to infinity jointly in the above theorem. Furthermore, to ensure uniform consistency with a sensible convergence rate in (3.8), we need to assume that $m_N\left[\omega(T, N, b)/a_T(f)\right]^{1-q} = o(1)$ and $N^{1/2}\left[\omega(T, N, b)/a_T(f)\right]^2 = o(1)$. The latter holds

if

$$\frac{\log T}{N^{1/2} b a_T^2(f)} = o(1), \quad \frac{\sqrt{N}\log(N \vee T)}{T b a_T^2(f)} = o(1) \text{ and } \frac{N b^4}{a_T^2(f)} = o(1), \tag{3.10}$$

indicating that N cannot diverge too fast to infinity. Letting the index variable $z_t \sim N(0,1)$ and choosing $L_T = \sqrt{\log\log T}$, we may show that $a_T(f) \propto (\log T)^{-1}$. Furthermore, when $N = T$, the conditions in (3.10) can be simplified to

$$\frac{\log^3 T}{T^{1/2} b} = o(1) \text{ and } T b^4 \log^2 T = o(1),$$

which are satisfied if $b \propto T^{-\theta_2}$ as in Assumption 4(b) with $1/4 < \theta_2 < 1/2$.

(b) Assuming that the density function $f(z)$ has a compact support $[-L, L]$ and is strictly larger than a positive constant, using Theorems 1 and 2 above, we may show that

$$\sup_{|z| \leqslant L-\epsilon} \left\| \widetilde{\mathbf{\Sigma}}_u(z) - \mathbf{\Sigma}_u(z) \right\| = O_P\left(m_N \left[\omega(T,N,b)\right]^{1-q} \right), \tag{3.11}$$

$$\sup_{|z| \leqslant L-\epsilon} \left\| \widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z) \right\|_{\mathbf{\Sigma}_X(z)} = O_P\left(N^{1/2} \left[\omega(T,N,b)\right]^2 + m_N \left[\omega(T,N,b)\right]^{1-q} \right), \tag{3.12}$$

$$\sup_{|z| \leqslant L-\epsilon} \left\| \widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z) \right\|_{\max} = O_P\left(\omega(T,N,b) \right) \tag{3.13}$$

for any small $\epsilon > 0$. It is well known that the local constant kernel estimation may perform poorly in the boundary region of the index variable z_t, due to the so-called boundary effect. This is the main reason for us to state the uniform consistency results (3.11)–(3.13) over $|z| \leqslant L - \epsilon$ rather than $|z| \leqslant L$. Following Su and Wang (2017), we may replace the conventional kernel weight $K_b(\cdot)$ in the local principal component analysis by a boundary-adjusted kernel defined as in Li and Racine (2007) to remove the boundary effect. Other boundary correction techniques include the reflection and transformation methods (e.g., Fan and Yao, 2003).

In some practical applications such as the dynamic optimal portfolio allocation, we need to take inverse of the estimated large covariance matrix and study its asymptotic property. The following theorem tackles this issue, showing the asymptotic invertibility of $\widetilde{\mathbf{\Sigma}}_X(z)$ and providing the uniform convergence rate for its inverse.

Theorem 3. *Suppose that the assumptions of Theorem 1 are satisfied and* $m_N \left[\omega(T,N,b)/a_T(f)\right]^{1-q} = o(1)$. *Then,* $\widetilde{\mathbf{\Sigma}}_X(z)$ *is non-singular with probability approaching* 1, *and*

$$\sup_{|z| \leqslant L_T} \left\| \widetilde{\mathbf{\Sigma}}_X^{-1}(z) - \mathbf{\Sigma}_X^{-1}(z) \right\| = O_P\left(m_N \left[\omega(T,N,b)/a_T(f)\right]^{1-q} \right). \tag{3.14}$$

as both N *and* T *tend to infinity jointly.*

Remark 4. (a) Although we mainly derive the limit results for the case when z_t is random in the

present paper, similar uniform consistency results hold for the fixed design case when $z_t = t/T$ (e.g., Robinson, 1989). In fact, the latter is analogous to the random design setting with z_t uniformly distributed over $[0,\ 1]$. Some regularity conditions need to be slightly modified. For example, the smoothness condition on density function in Assumption 1(b) can be removed; and the condition $\mathsf{E}(u_{it}|z_t, F_t) = 0$ a.s. and (3.6) in Assumption 3(b) should be replaced by $\mathsf{E}(u_{it}|F_t) = 0$ a.s., and

$$\max_{1\leqslant t\leqslant T} \sup_{0\leqslant z\leqslant 1} \|\mathbf{B}(z)'u_t\|^2 = O_P\left(N \log T\right),$$

respectively. By modifying the proofs in Appendices A.3 and B accordingly, we can get the uniform consistency results (3.11)–(3.13) by replacing $|z| \leqslant L - \epsilon$ by $\epsilon \leqslant z \leqslant 1 - \epsilon$.

(b) A key step to derive the uniform consistency results stated in Theorems 1–3 above is to prove the uniform convergence for the estimated factors and factor loadings. In fact, in the proof of Lemma 2 available in the supplemental document, we show that

$$\sup_{|z|\leqslant L_T} \frac{1}{T}\sum_{t=1}^{T} \left\|\widetilde{F}_t(z) - \mathbf{H}(z)F_t(z)\right\|^2 = O_P\left(a_T^{-2}(f)\left(\frac{\log T}{Nb} + \frac{1}{T} + b^2\right)\right) \tag{3.15}$$

and

$$\max_{1\leqslant i\leqslant N} \sup_{|z|\leqslant L_T} \left\|\widetilde{B}_i(z) - \left[\mathbf{H}^{-1}(z)\right]' B_i(z)\right\| = O_P\left(\omega(T, N, b)/a_T(f)\right), \tag{3.16}$$

where $\widetilde{F}_t(z)$ and $\widetilde{B}_i(z)$ are defined in Section 2.2 and $\mathbf{H}(z)$ is a z-dependent $K \times K$ rotation matrix to be defined in the proof of Lemma 2. These results are of independent interest, complementing those developed by Su and Wang (2017) and Pelger and Xiong (2020).

4 Practical issues in estimation

In this section we discuss two practical issues for implementing the developed kernel-based estimation methodology: selection of the factor number K and choice of the tuning parameter in the generalised shrinkage.

4.1 Selection of the factor number

The number of unobservable factors is often un-specified in practice, and needs to be estimated before implementing the estimation methodology introduced in Section 2.2. Towards this, we propose a simple modification of the ratio criterion in Lam and Yao (2012) and Ahn and Horenstein (2013). Other selection criteria such as the information criterion (c.f., Bai and Ng, 2002; Fan, Liao and Mincheva, 2013) may also be applicable (with some modifications). Let $\overline{\lambda}_{k,z}$ be the k-th largest eigenvalue of the kernel-weighted sample covariance matrix $\overline{\mathbf{\Sigma}}_X(z)$ defined in (2.5). At a given

point z, we estimate the number K by

$$\widehat{K}(z) = \underset{1 \leqslant k \leqslant \bar{K}}{\operatorname{argmin}} \frac{\overline{\lambda}_{k+1,z}}{\overline{\lambda}_{k,z}}, \tag{4.1}$$

where $\bar{K}$ is a pre-specified positive integer (which is independent of z) and $0/0 = 1$. In practical implementation, we set $\overline{\lambda}_{k,z}/N$ as 0 if its absolute value is smaller than ϵ_1 which is a pre-specified small positive number (say, 0.01). Consequently, we have

$$\frac{\overline{\lambda}_{k+1,z}}{\overline{\lambda}_{k,z}} = \frac{\overline{\lambda}_{k+1,z}/N}{\overline{\lambda}_{k,z}/N} = 0/0 = 1, \tag{4.2}$$

when neither $\left|\overline{\lambda}_{k+1,z}/N\right|$ nor $\left|\overline{\lambda}_{k,z}/N\right|$ exceeds ϵ_1. Then, we take maximum of $\widehat{K}(z)$ over $z \in \mathcal{Z}$ with $\mathcal{Z}$ being a subset of $\{z : |z| \leqslant L_T\}$, i.e.,

$$\widehat{K} = \max_{z \in \mathcal{Z}} \widehat{K}(z). \tag{4.3}$$

In practice, when the index variable has a compact support, we may choose $\mathcal{Z}$ as an equidistant grid of points which lie in the interior of the support (to circumvent the boundary effect in kernel estimation). Note that slight over-identification of the factor number usually does not affect consistency or convergence rates of the subsequent estimation (e.g., Fan, Liao and Mincheva, 2013; Moon and Weidner, 2015). The following theorem shows that $\widehat{K}(z)$ converges to the true value K uniformly over $|z| \leqslant L_T$, indicating that $\widehat{K}$ is a consistent estimation.

Theorem 4. *Suppose that the assumptions of Theorem 1 are satisfied,* $m_N \left[\omega(T, N, b)/a_T(f)\right]^{1-q} = o(1)$ *and* $K \geqslant 1$. *Then, as both* N *and* T *tend to infinity jointly,* $\widehat{K}(z) = K$ *with probability approaching one uniformly over* $|z| \leqslant L_T$, *and thus* $P\left(\widehat{K} = K\right) \to 1$.

The simulation studies in Section 5.1 below show that the ratio criterion performs well in finite samples. However, a disadvantage of the ratio criterion is that it would select at least one common factor and could not work in the setting with zero factor. To address this problem, we may replace "$1 \leqslant k \leqslant \bar{K}$" in (4.1) by "$0 \leqslant k \leqslant \bar{K}$" and define a mock eigenvalue $\overline{\lambda}_{0,z} = \lambda(N, T) \cdot N$ with $\lambda(N, T)$ satisfying that

$$\lambda(N, T) \to 0 \text{ and } \frac{\lambda(N, T) a_T(f)}{\nu(T, N, b) + b^2} \to \infty,$$

where $\nu(T, N, b) = \sqrt{\frac{\log(N \vee T)}{Tb}}$. Following Ahn and Horenstein (2013)'s suggestion, we can choose $\lambda(N, T) = 1/\log(N \wedge T)$ when $a_T(f)$ is bounded away from zero.

4.2 Choice of the variable tuning parameter

It is well known that shrinkage estimation of the large covariance matrix in finite samples is sensitive to the choice of the tuning parameter. In the proposed estimation procedure, we allow the tuning parameter to vary with the index variable, as in Chen and Leng (2016). As the high-dimensional data in our paper may be serially correlated over time, we cannot directly adopt the tuning parameter selection rule in Chen and Leng (2016) where only independent data are considered. Instead, we use the following selection criterion proposed by Chen, Li and Linton (2019) which accounts for temporal dependence.

1. For each given z, we divide the full sample into $\lfloor T/(2M_0) \rfloor$ groups, where $\lfloor \cdot \rfloor$ denotes the floor function. Specifically, the m-th group contains the observations indexed by $t = (m-1) \cdot M_0 + 1, (m-1) \cdot M_0 + 2, \cdots, m \cdot M_0 + \lfloor T/2 \rfloor$, where $m = 1, \cdots, \lfloor T/(2M_0) \rfloor$, and the sample size of each group is $\lfloor T/2 \rfloor + M_0$. For each group, we further split the data into two sub-samples of size $T_1 = \lfloor \frac{T}{2}(1 - 1/\log(T/2)) \rfloor$ and $T_2 = \lfloor T/2 \rfloor - T_1$ by leaving out M_0 observations in-between them.

2. For the m-th group, we obtain $\widetilde{\mathbf{\Sigma}}_{u,m}(z|\rho)$ using (2.10) from the first sub-sample with the tuning parameter set as ρ, and $\check{\mathbf{\Sigma}}_{u,m}(z)$ using the naive covariance matrix estimation (without applying the shrinkage technique) from the second sub-sample. We then choose the variable tuning parameter to minimise

$$\sum_{m=1}^{\lfloor T/(2M_0) \rfloor} \left\| \widetilde{\mathbf{\Sigma}}_{u,m}(z|\rho) - \check{\mathbf{\Sigma}}_{u,m}(z) \right\|_F^2$$

with respect to $\rho \in [\rho_1, \rho_2]$, where

$$\rho_1 = \epsilon_2 + \inf \left\{ \rho_* > 0 \mid \lambda_{\min}\left(\widetilde{\mathbf{\Sigma}}_{u,m}(z|\rho) \right) > 0, \ \ \forall \rho > \rho_* \right\},$$

ϵ_2 is a sufficiently small positive constant and ρ_2 is a sufficiently large positive constant.

The tuning parameter chosen above is allowed to vary with z. The motivation of leaving out M_0 observations between the first and second sub-samples in each group is to make the correlation between these two sub-samples weak or negligible. As in Chen, Li and Linton (2019), we choose $M_0 = 10$ in the numerical studies. The definition of ρ_1 in Step 2 is similar to that in (4.1) of Fan, Liao and Mincheva (2013), ensuring that $\widetilde{\mathbf{\Sigma}}_{u,m}(z|\rho)$ is positive definite in finite samples as long as $\rho \geqslant \rho_1$. The numerical studies in Section 5 show that the proposed tuning parameter selection has reliable performance when T is as small as 200.

5 Numerical studies

In this section, we provide numerical studies including simulation and an empirical example to examine the performance of the proposed covariance matrix estimation method in finite samples. The supplemental document contains additional numerical results.

5.1 Monte-Carlo simulation

We use the time-dependent factor model (2.1) to generate data in simulation. The factor loading functions are defined by $\mathbf{B}(z) = [B_1(z), \cdots, B_N(z)]'$, where $B_i(z) = [B_{i1}(z), \cdots, B_{iK}(z)]'$ with $B_{ij}(z) = \zeta_{ij}\mathcal{H}_{j-1}(z)$, $\zeta_{ij} \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\mathcal{H}_j(z)$ takes the polynomial form, i.e., $\mathcal{H}_j(z) = z^j$. Note that ζ_{ij}'s are generated once only using "set.seed(42)" in R and are then fixed over replications in the simulation study. Hence, each entry of $\mathbf{B}(z)$ can be regarded as a deterministic function of z, consistent with our model design and Assumption 2(a). The index observations z_t are independently generated from the uniform distribution $U(0,1)$. For the factors $F_t = (F_{t1}, F_{t2}, \cdots, F_{tK})'$, we let $F_{t1} \equiv 1$ and generate $\overline{F}_t = (F_{t2}, \cdots, F_{tK})'$ by

$$\overline{F}_t = 0.3\overline{F}_{t-1} + \nu_t, \quad \nu_t \overset{\text{i.i.d.}}{\sim} N\left(0_{K-1}, \mathbf{I}_{K-1}\right),$$

where 0_k is a k-dimensional null vector. The number of factors is set as $K = 6$.

The idiosyncratic errors u_t are independently generated from $N\left(0_N, \mathbf{\Sigma}_u(z_t)\right)$. To save the space, we only consider the following structure: $\mathbf{\Sigma}_u(z) = [\sigma_{u,ij}(z)]_{N\times N}$ with

$$\sigma_{u,ij}(z) = \exp(z/2)\{I(i=j) + [\phi(z) + 0.1] \cdot I(|i-j| = 1) + \phi(z) \cdot I(|i-j| = 2)\},$$

where $\phi(\cdot)$ is the density of the standard normal distribution and $I(\mathcal{A})$ is the indicator function for the event $\mathcal{A}$. The online supplementary appendix reports the simulation results for alternative forms of $\mathbf{\Sigma}_u(z)$.

Four shrinkage functions are considered when we apply the generalised shrinkage to the large covariance matrix estimation in the simulation: hard thresholding, soft thresholding, SCAD and adaptive LASSO. The sample size is $T = 200, 300$, and the dimension of X_t is $N = 200, 300, 400$. We repeat each simulation setting 100 times[2]. The bandwidth b in the kernel-based estimation is chosen via the rule of thumb (i.e., $b = 1.059T^{-1/5}$)[3]. As in Chen, Li and Linton (2019), we use $K(w) = \frac{1}{\sqrt{2\pi}}\exp(-w^2/2)$ as the kernel function. For each dataset generated, we first select the

[2] As mentioned in Fan, Liao and Mincheva (2013) and Chen, Li and Linton (2019), simulation for the large covariance matrix estimation is extremely time consuming. In view of the number of replications in Fan, Liao and Mincheva (2013) and Chen, Li and Linton (2019), we adopt 100 replications in this study.

[3] Additional simulation results with different bandwidth values are reported in Appendix D.2 of the supplemental document. Overall, the developed estimation method performs stably over these different bandwidth values.

number of factors using the ratio criterion proposed in Section 4.1, and then choose the tuning parameter of each shrinkage function as in Section 4.2.

As the index variable z_t has a compact support $[0, 1]$, when implementing the ratio criterion, we estimate $\widehat{K}(z_i)$ at nine grid points $z_i = 0.1, 0.2, \cdots, 0.9$, select the tuning parameter $\epsilon_1 = 0.01$, and then take maximum of $\widehat{K}(z_i)$ over the grid points. Define the frequency of accurately selecting the true factor number over 100 replications as $\mathcal{R}(K) = \frac{1}{100}\sum_{r=1}^{100} I(\widehat{K}_r = K)$, where $\widehat{K}_r$ stands for the estimated factor number in the r-th replication. Table 1 below shows that the developed ratio criterion can very accurately estimate the number of factor in finite samples.

Table 1: The values of $\mathcal{R}(K)$.

$T \setminus N$	200	300	400
200	0.96	0.97	1
300	1	1	1

To measure the accuracy of covariance matrix estimation, we compute the medians of spectral and Frobenius losses defined as follows:

$$\begin{aligned}
\mathrm{MSL}_u &= \mathrm{Median}\{\nabla_{S_u}(z_i);\ i = 1, \cdots, 9\}, \quad \nabla_{S_u}(z) = \left\|\widetilde{\boldsymbol{\Sigma}}_u(z) - \boldsymbol{\Sigma}_u(z)\right\|, \\
\mathrm{MFL}_u &= \mathrm{Median}\{\nabla_{F_u}(z_i);\ i = 1, \cdots, 9\}, \quad \nabla_{F_u}(z) = \left\|\widetilde{\boldsymbol{\Sigma}}_u(z) - \boldsymbol{\Sigma}_u(z)\right\|_F, \\
\mathrm{MSL}_X &= \mathrm{Median}\{\nabla^*_{S_X}(z_i);\ i = 1, \cdots, 9\}, \quad \nabla^*_{S_X}(z) = \left\|\boldsymbol{\Sigma}_X^{-1/2}(z)\widetilde{\boldsymbol{\Sigma}}_X(z)\boldsymbol{\Sigma}_X^{-1/2}(z) - \mathbf{I}_N\right\|, \\
\mathrm{MFL}_X &= \mathrm{Median}\{\nabla^*_{F_X}(z_i);\ i = 1, \cdots, 9\}, \quad \nabla^*_{F_X}(z) = \left\|\boldsymbol{\Sigma}_X^{-1/2}(z)\widetilde{\boldsymbol{\Sigma}}_X(z)\boldsymbol{\Sigma}_X^{-1/2}(z) - \mathbf{I}_N\right\|_F,
\end{aligned}$$

where $z_i = i/10$ for $i = 1, \cdots, 9$. As discussed in Section 3.2, we examine the relative error $\boldsymbol{\Sigma}_X^{-1/2}(z)\widetilde{\boldsymbol{\Sigma}}_X(z)\boldsymbol{\Sigma}_X^{-1/2}(z) - \mathbf{I}_N$ when measuring the accuracy of $\widetilde{\boldsymbol{\Sigma}}_X(z)$ (with spiked eigenvalues). Table 2 reports the medians and standard deviations (in parentheses) of MSL_u, MFL_u, MSL_X and MFL_X over the 100 replications. The MSL_u value decreases as the sample size T goes up but remains stable as the dimension N increases, which should be expected. Regarding the relative estimation errors for $\boldsymbol{\Sigma}_X(z)$, both the MSL_X and MFL_X values increase as the dimension N increases. This justifies the normalisation rate of $\sqrt{N}$ in the definition of $\|\cdot\|_{\boldsymbol{\Sigma}_X(z)}$. In addition, all the standard deviations are relatively small, indicating stability of the proposed method in finite samples.

Appendix D in the supplemental document further examines the performance of the proposed covariance matrix estimation method in the over-fitting scenario with data generated by a conventional factor model with constant factor loadings, and compares with the performance of Fan, Liao and Mincheva (2013)'s POET method. It also examines how the signal-to-noise ratio affects the factor number selection and assesses the numerical performance of our method when the factor loadings contain noises.

Table 2: Accuracy of the covariance matrix estimation. The MSL_u, MFL_u, MSL_X and MFL_X values are reported for $N \in \{200, 300, 400\}$ and $T \in \{200, 300\}$ by using four different thresholding rules (i.e., soft, hard, scad, and alasso).

		$T = 200$				$T = 300$			
		soft	hard	scad	alasso	soft	hard	scad	alasso
$N = 200$	MSL_u	1.89	2.20	1.89	2.02	1.54	2.15	1.79	1.66
		(0.08)	(0.02)	(0.08)	(0.07)	(0.11)	(0.04)	(0.11)	(0.09)
	MFL_u	10.79	14.81	10.79	11.71	8.37	14.58	10.29	8.59
		(0.22)	(0.03)	(0.30)	(0.32)	(0.76)	(0.24)	(0.68)	(0.26)
	MSL_X	24.10	24.65	24.09	24.20	21.60	22.75	21.87	21.67
		(4.39)	(4.39)	(4.40)	(4.40)	(3.75)	(4.03)	(3.75)	(3.78)
	MFL_X	37.93	43.21	37.90	39.09	32.99	41.12	35.69	33.31
		(4.08)	(3.72)	(4.12)	(4.06)	(4.48)	(4.04)	(4.37)	(4.37)
$N = 300$	MSL_u	1.86	2.19	1.87	2.00	1.74	2.16	1.76	1.57
		(0.07)	(0.02)	(0.09)	(0.06)	(0.09)	(0.02)	(0.06)	(0.08)
	MFL_u	12.76	18.13	13.04	13.78	12.19	17.89	12.31	9.84
		(0.38)	(0.04)	(0.89)	(0.26)	(0.70)	(0.22)	(0.22)	(0.27)
	MSL_X	34.58	35.36	34.62	34.74	31.45	32.42	31.56	31.05
		(7.90)	(8.01)	(7.94)	(7.92)	(5.24)	(5.41)	(5.28)	(5.18)
	MFL_X	51.98	58.62	52.19	53.26	48.91	56.05	49.06	45.83
		(8.06)	(7.41)	(8.17)	(7.90)	(5.69)	(4.96)	(5.56)	(5.84)
$N = 400$	MSL_u	1.88	2.19	1.97	2.00	1.76	2.16	1.76	1.58
		(0.09)	(0.02)	(0.12)	(0.06)	(0.06)	(0.02)	(0.06)	(0.09)
	MFL_u	15.27	20.93	16.33	15.69	14.08	20.72	14.02	11.10
		(1.20)	(0.05)	(1.46)	(0.46)	(0.19)	(0.08)	(0.20)	(0.85)
	MSL_X	40.00	40.58	40.13	40.07	34.93	35.60	34.92	34.55
		(9.27)	(9.25)	(9.27)	(9.26)	(6.68)	(6.71)	(6.68)	(6.67)
	MFL_X	59.99	67.46	61.59	61.03	54.86	63.37	54.77	51.03
		(9.39)	(8.33)	(9.38)	(9.04)	(6.39)	(5.72)	(6.39)	(6.84)

5.2 An empirical study

We next apply the developed model and estimation methodology to construction of minimum variance portfolios using daily returns of S&P 500 stocks collected from https://www.kaggle.com. We collect the data over the time period between 2 January 2008 and 31 December 2018. After removing the companies which have missing stock returns during the whole time period, we end up with 319 stocks ($N = 319$). As in Pelger and Xiong (2020), we use the log-normalised CBOE volatility index (VIX) available at http://www.cboe.com as the index variable z_t in model (2.1).

Similarly to Chen, Li and Linton (2019) and Engle, Ledoit and Wolf (2019), we assume no transaction cost and allow short-sales in construction of the minimum variance portfolio. Define

$$\min_{w} \; w'\widetilde{\Sigma}_X(z)w \quad \text{subject to} \quad w'1_N = 1, \tag{5.1}$$

where 1_N is an N-dimensional vector of ones and $\widetilde{\Sigma}_X(z)$ is defined as in (2.11). The analytic solution to the above minimisation problem is

$$w^*(z) = \frac{\widetilde{\Sigma}_X^{-1}(z)1_N}{1_N'\widetilde{\Sigma}_X^{-1}(z)1_N}. \tag{5.2}$$

Our main interest lies in the out-of-sample numerical performance of the minimum variance portfolio using the proposed large dynamic covariance matrix estimation. For each trading day t in the out-sample, we estimate $\Sigma_X(z_t)$ with the sample information on the most recent 252×2 trading days available in the collected data set before t by fixing the number of factors as $1, 2, 3, 4$, respectively[4], and subsequently construct the minimum variance portfolio. We use the rolling-window based calculation as in Chen, Li and Linton (2019). For each rolling-window, we re-estimate the dynamic covariance matrix and select the tuning parameter involved as in Section 4.2. Alternatively, one may conduct the calculation using the expanding window as in Pelger and Xiong (2020), which would be much more time-consuming.

With all the out-of-sample global minimum variance portfolio returns obtained, we compute their standard deviation (STD), the information ratio (IR) defined as the ratio of average to STD, and the Sharpe ratio (SR) defined as the mean of returns minus risk-free rate normalised by STD. For the purpose of comparison, we also consider Fan, Liao and Mincheva (2013)'s POET method. The results are summarised in Table 3. As pointed out by Engle, Ledoit and Wolf (2019), in the context of constructing minimum variance portfolio, the primary measure for evaluating the performance should be the STD measurement. High SR and IR are desirable but may be of secondary importance when the main aim is to assess the performance of covariance matrix estimation. From Table 3, the out-of-sample performance of the minimum variance portfolio with

[4]Due to constraints of time and computational power, we no longer explore the cases with more factors.

covariance matrices estimated by the kernel-weighted and POET methods improves as the factor number increases (from 2 to 4). Overall, based on all the three criteria, the kernel-weighted method outperforms the POET method for each given number of factors (in particular, when the number is between 2 and 4). The POET method often needs to include at least one more factor to beat the proposed kernel-weighted method.

Table 3: Out-of-sample performance of the minimum variance portfolio with covariance matrices estimated by the kernel-weighted and POET methods. The STD, IR, and SR are computed with the factor number fixed as $1, 2, 3$ and 4.

	Kernel-weighted method			POET		
No. of Factors	STD	IR	SR	STD	IR	SR
1	0.0078	0.0733	0.0615	0.0081	0.0750	0.0637
2	0.0079	0.0636	0.0520	0.0083	0.0594	0.0484
3	0.0066	0.0880	0.0740	0.0070	0.0729	0.0598
4	0.0063	0.0905	0.0760	0.0066	0.0851	0.0712

6 Conclusions

In the present paper, we have considered the problem of estimating covariance matrices of high-dimensional inter-correlated variables generated from an approximate factor model with factor loadings varying smoothly with an index variable. We have presented assumptions compatible with these features of data, provided new uniform consistency results for our estimation method, discussed the implementation of our approach, and illustrated its performance via synthesised and real data.

For future work, the idea of the paper can be extended in several directions. First, we may consider estimating the dynamic covariance matrix of X_t defined in (2.1) with multiple index variables when the dimension of z_t is larger than one. This arises, for example, when one chooses to use Fama-French three factors as index variables to describe the dynamic covariance structure of stocks (e.g., Chen, Li and Linton, 2019). In this setting, the developed nonparametric covariance matrix estimation method needs to be substantially modified and some additional model restrictions (such as the additive model structure) should be imposed on the factor loading functions to avoid the so-called "curse of dimensionality" problem. Other possible extensions include semi-parametric estimation of the large dynamic covariance matrix with conditional sparsity (where some entries vary over time but others have constant values), and formally testing the constancy of the time-dependent factor loadings. For the latter, we conjecture that the quadratic form test statistic developed by Su and Wang (2017) may be applicable with modifications needed in the relevant theoretical justification.

Acknowledgements

The authors thank the Co-Editor, an Associate Editor and two reviewers for the valuable comments, which have substantially improved an earlier version of the paper. Wang's research is supported by the National Key R&D Program of China (No. 2018YFA0703900), the National Natural Science Foundation of China (No.11701331), Shandong Provincial Natural Science Foundation (No. ZR2017QA007) and Young Scholars Program of Shandong University. Li's research is partly supported by the BA/Leverhulme Small Research Grant funded by the Leverhulme Trust (No. SRG1920/100603). Leng's research is partially supported by a Turing fellowship. Part of this work was completed when Dr Hanchao Wang visited the University of York with financial support from the China Scholarship Council. Wang thanks the University of York for hosting the visit and providing relevant resource.

Supplement

The online supplementary document contains proofs of some technical lemmas and additional numerical results.

Appendix A: Proofs of the mathematical results

In Appendix A.1, we introduce a local version of the POET estimation technique and prove its equivalence to that defined in Section 2.2. In Appendix A.2, we state some technical lemmas whose proofs are given in Appendix B of the supplemental document. In Appendix A.3, we provide the detailed proofs of the main theoretical results.

Appendix A.1: Local POET

Let $\overline{\mathbf{\Sigma}}_X(z)$ be the kernel-weighted sample covariance matrix defined in (2.5). With eigen-analysis on $\overline{\mathbf{\Sigma}}_X(z)$, we obtain $\left(\overline{\lambda}_{k,z}, \overline{\boldsymbol{\eta}}_{k,z}\right)$, $k = 1, \cdots, N$, as pairs of eigenvalues and normalised eigenvectors, where the eigenvalues are arranged in a decreasing order. By the spectral decomposition of $\overline{\mathbf{\Sigma}}_X(z)$, we have

$$\overline{\mathbf{\Sigma}}_X(z) = \sum_{k=1}^{K} \overline{\lambda}_{k,z} \overline{\boldsymbol{\eta}}_{k,z} \overline{\boldsymbol{\eta}}_{k,z}' + \overline{\mathbf{\Sigma}}_u(z), \;\; \overline{\mathbf{\Sigma}}_u(z) = \sum_{k=K+1}^{N} \overline{\lambda}_{k,z} \overline{\boldsymbol{\eta}}_{k,z} \overline{\boldsymbol{\eta}}_{k,z}'. \tag{A.1}$$

Let $\overline{\sigma}_{u,ij}(z)$ be the (i,j)-th entry of $\overline{\mathbf{\Sigma}}_u(z)$. As in Section 2.2, we can apply the generalised shrinkage technique to $\overline{\mathbf{\Sigma}}_u(z)$, forcing very small off-diagonal entries $\overline{\sigma}_{u,ij}(z)$ to be zero. Define

$$\widehat{\mathbf{\Sigma}}_u(z) = \left(\widehat{\sigma}_{u,ij}(z)\right)_{N\times N}, \;\; \widehat{\sigma}_{u,ij}(z) = \begin{cases} \overline{\sigma}_{u,ii}(z), & i = j, \\ s_{\rho(z)}\left(\overline{\sigma}_{u,ij}(z)\right), & i \neq j, \end{cases} \tag{A.2}$$

where $s_{\rho(z)}(\cdot)$ is a shrinkage function with $\rho(z)$ being a tuning parameter changing with the index variable. Consequently, we estimate $\mathbf{\Sigma}_X(z)$ by

$$\widehat{\mathbf{\Sigma}}_X(z) = \sum_{k=1}^{K} \overline{\lambda}_{k,z}\overline{\boldsymbol{\eta}}_{k,z}\overline{\boldsymbol{\eta}}'_{k,z} + \widehat{\mathbf{\Sigma}}_u(z), \tag{A.3}$$

which is an extension of Fan, Liao and Mincheva (2013)'s POET method to the local nonparametric setting. Proposition 1 below is similar to Theorem 1 in Fan, Liao and Mincheva (2013), showing the equivalence of the above local POET estimators and those defined in (2.10) and (2.11).

Proposition 1. *Suppose that the shrinkage function and the variable tuning parameter used in (2.10) are the same as those in (A.2). Then we have $\widehat{\mathbf{\Sigma}}_u(z) = \widetilde{\mathbf{\Sigma}}_u(z)$ and $\widehat{\mathbf{\Sigma}}_X(z) = \widetilde{\mathbf{\Sigma}}_X(z)$ for any z.*

Proof. The proof is similar to the proof of Theorem 1 in Fan, Liao and Mincheva (2013). As $\widetilde{\mathbf{B}}(z) = \frac{1}{T}\mathbf{X}(z)\widetilde{\mathbf{F}}(z)$ and $\frac{1}{T}\widetilde{\mathbf{F}}(z)'\widetilde{\mathbf{F}}(z) = \mathbf{I}_K$, the sample covariance matrix of $\widetilde{u}_t(z)$ can be written as

$$\begin{aligned}\check{\mathbf{\Sigma}}_u(z) &= \frac{1}{T}\sum_{t=1}^{T}\left[X_t(z) - \widetilde{\mathbf{B}}(z)\widetilde{F}_t(z)\right]\left[X_t(z) - \widetilde{\mathbf{B}}(z)\widetilde{F}_t(z)\right]' \\ &= \frac{1}{T}\sum_{t=1}^{T} X_t(z)X_t(z)' - \widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)'. \end{aligned} \tag{A.4}$$

On the other hand, combining (2.5) and (A.1), we readily have that

$$\frac{1}{T}\sum_{t=1}^{T} X_t(z)X_t(z)' = \sum_{k=1}^{K} \overline{\lambda}_{k,z}\overline{\boldsymbol{\eta}}_{k,z}\overline{\boldsymbol{\eta}}'_{k,z} + \overline{\mathbf{\Sigma}}_u(z). \tag{A.5}$$

By (A.4) and (A.5), it suffices to show

$$\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' = \sum_{k=1}^{K} \overline{\lambda}_{k,z}\overline{\boldsymbol{\eta}}_{k,z}\overline{\boldsymbol{\eta}}'_{k,z}. \tag{A.6}$$

In order to prove (A.6), we next switch the role of the (varying-coefficient) factor loadings and common factors in the kernel-weighted least squares objective function in (2.7) and consider the following identification condition:

$$\frac{1}{N}\mathbf{B}(z)'\mathbf{B}(z) = \mathbf{I}_K, \quad \frac{1}{T}\mathbf{F}(z)'\mathbf{F}(z) \text{ is diagonal.}$$

Consequently, we can obtain the following solution to the kernel-weighted local least squares objective function: $\widehat{\mathbf{B}}(z) = [\widehat{B}_1(z), \cdots, \widehat{B}_K(z)] = [\overline{\boldsymbol{\eta}}_{1,z}, \cdots, \overline{\boldsymbol{\eta}}_{K,z}]$, the K eigenvectors (multiplied by $\sqrt{N}$) of the matrix $\mathbf{X}(z)\mathbf{X}(z)'$ corresponding to the first K largest eigenvalues, and $\widehat{\mathbf{F}}(z) = \frac{1}{N}\mathbf{X}(z)'\widehat{\mathbf{B}}(z)$. Furthermore, we may show that $\widehat{\mathbf{B}}(z)\widehat{\mathbf{F}}(z)' = \widetilde{\mathbf{B}}(z)\widetilde{\mathbf{F}}(z)'$ and $\frac{1}{T}\widehat{\mathbf{F}}(z)'\widehat{\mathbf{F}}(z) = \mathrm{diag}\{\overline{\lambda}_{1,z}, \cdots, \overline{\lambda}_{K,z}\}$. Hence, we have

$$\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' = \frac{1}{T}\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{F}}(z)'\widetilde{\mathbf{F}}(z)\widetilde{\mathbf{B}}(z)' = \frac{1}{T}\widehat{\mathbf{B}}(z)\widehat{\mathbf{F}}(z)'\widehat{\mathbf{F}}(z)\widehat{\mathbf{B}}(z)' = \sum_{k=1}^{K} \overline{\lambda}_{k,z}\overline{\boldsymbol{\eta}}_{k,z}\overline{\boldsymbol{\eta}}'_{k,z},$$

completing the proof of (A.6). ■

Appendix A.2: Some technical lemmas

In order to prove the main asymptotic results, we need the following technical lemmas. Their proofs are available in Appendix B of the supplemental document.

Lemma 1. *Suppose that Assumptions 1, 3 and 4(a)(b) are satisfied. Then, we have*

$$\max_{1\leqslant t\leqslant T}\sup_{|z|\leqslant L_T}\frac{1}{N}\sum_{s=1}^{T}\left|\mathsf{E}\left[u_s(z)'u_t(z)\right]\right| = O\left(a_T^{-1}(f)\right), \tag{A.7}$$

where $u_t(z) = u_t K_{t,b}^{1/2}(z)$, L_T *and* $a_T(f)$ *are defined in Assumption 1(b).*

Lemma 2. *Suppose that Assumptions 1, 2(a)(b), 3 and 4(a)(b) are satisfied and* $\omega(T,N,b)/a_T(f) = o(1)$, *where* $\omega(T,N,b)$ *is defined in Assumption 4(c). Then, we have*

$$\max_{1\leqslant i\leqslant N}\sup_{|z|\leqslant L_T}\frac{1}{T}\sum_{t=1}^{T}[\widetilde{u}_{it}(z)-u_{it}(z)]^2 = O_P\left([\omega(T,N,b)/a_T(f)]^2\right), \tag{A.8}$$

where $\widetilde{u}_{it}(z)$ *is defined in Section 2.2 and* $u_{it}(z)$ *is the* i*-th element of* $u_t(z)$.

Lemma 3. *Suppose that Assumptions 1, 2(c), 4(a)(b) and (3.3) are satisfied and* $\nu(T,N,b)+b^2 = o\left(a_T(f)\right)$, *where* $\nu(T,N,b) = \sqrt{\frac{\log(N\vee T)}{Tb}}$. *Then, we have*

$$\max_{1\leqslant i,j\leqslant N}\sup_{|z|\leqslant L_T}\left|\frac{1}{T}\sum_{t=1}^{T}u_{it}(z)u_{jt}(z)-\sigma_{u,ij}(z)\right| = O_P\left(a_T^{-1}(f)\left(\nu(T,N,b)+b^2\right)\right). \tag{A.9}$$

Lemma 4. *Suppose that Assumptions 1–3 and 4(a)(b) are satisfied and* $\omega(T,N,b)/a_T(f) = o(1)$. *Then, we have*

$$\max_{1\leqslant i,j\leqslant N}\sup_{|z|\leqslant L_T}\left|\check{\sigma}_{u,ij}(z)-\sigma_{u,ij}(z)\right| = O_P\left(\omega(T,N,b)/a_T(f)\right), \tag{A.10}$$

where $\check{\sigma}_{u,ij}(z)$ *is defined in (2.9).*

Appendix A.3: Proofs of the main asymptotic theorems

With the technical lemmas given in Appendix A.2, we next provide the detailed proof of the main asymptotic results.

Proof of Theorem 1. By the definition of $\widetilde{\mathbf{\Sigma}}_u(z)$ and the property of the shrinkage function $s_{\rho(z)}(\cdot)$, we have

$$\sup_{|z|\leqslant L_T}\|\widetilde{\mathbf{\Sigma}}_u(z)-\mathbf{\Sigma}_u(z)\|$$

$$
\begin{aligned}
&\leqslant \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|\widetilde{\sigma}_{u,ij}(z) - \sigma_{u,ij}(z)\right| \\
&= \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|s_{\rho(z)}\left(\check{\sigma}_{u,ij}(z)\right) I(\check{\sigma}_{u,ij}(z) > \rho(z)) - \sigma_{u,ij}(z)\right| \\
&= \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|s_{\rho(z)}\left(\check{\sigma}_{u,ij}(z)\right) I(\check{\sigma}_{u,ij}(z) > \rho(z)) - \sigma_{u,ij}(z) I(\check{\sigma}_{u,ij}(z) > \rho(z)) - \right. \\
&\qquad \left. \sigma_{u,ij}(z) I(\check{\sigma}_{u,ij}(z) \leqslant \rho(z))\right| \\
&\leqslant \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|s_{\rho(z)}\left(\check{\sigma}_{u,ij}(z)\right) - \check{\sigma}_{u,ij}(z)\right| I(\check{\sigma}_{u,ij}(z) > \rho(z)) + \\
&\qquad \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|\check{\sigma}_{u,ij}(z) - \sigma_{u,ij}(z)\right| I(\check{\sigma}_{u,ij}(z) > \rho(z)) + \\
&\qquad \sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} \left|\sigma_{u,ij}(z)\right| I(\check{\sigma}_{u,ij}(z) \leqslant \rho(z)) \\
&=: \Pi_1 + \Pi_2 + \Pi_3.
\end{aligned}
$$

Define the event

$$
\mathcal{B}(C) = \left\{ \max_{1\leqslant i,j\leqslant N} \sup_{|z|\leqslant L_T} \left|\check{\sigma}_{u,ij}(z) - \sigma_{u,ij}(z)\right| \leqslant C\omega_L(T,N,b) \right\}
$$

where C is a positive constant and $\omega_L(T,N,b) = \omega(T,N,b)/a_T(f)$. For any small $\varepsilon > 0$, by (A.10) in Lemma 4, we may find a sufficiently large constant C_ε such that

$$
\mathsf{P}\left(\mathcal{B}(C_\varepsilon)\right) \geqslant 1 - \varepsilon. \tag{A.11}
$$

Note that

$$
\Pi_1 \leqslant \sup_{|z|\leqslant L_T} \rho(z) \left[\max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} I\left(\check{\sigma}_{u,ij}(z) > \rho(z)\right) \right]
$$

and

$$
\Pi_2 \leqslant C_\varepsilon \omega_L(T,N,b) \left[\sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} I\left(\check{\sigma}_{u,ij}(z) > \rho(z)\right) \right]
$$

conditional on the event $\mathcal{B}(C_\varepsilon)$. By the reverse triangle inequality,

$$
\left|\check{\sigma}_{u,ij}(z)\right| \leqslant \left|\sigma_{u,ij}(z)\right| + C_\varepsilon \omega_L(T,N,b)
$$

on $\mathcal{B}(C_\varepsilon)$. Letting $\underline{M} = 2C_\varepsilon$ in Assumption 4(c), as $\mathbf{\Sigma}_u(\cdot) \in \mathcal{U}(q, m_N, L)$, we readily have that

$$
\Pi_1 + \Pi_2 \leqslant \omega_L(T,N,b)(\overline{M} + C_\varepsilon) \left[\sup_{|z|\leqslant L_T} \max_{1\leqslant i\leqslant N} \sum_{j=1}^{N} I(\check{\sigma}_{u,ij}(z) > \underline{M}\omega_L(T,N,b)) \right]
$$

$$
\begin{aligned}
&\leqslant \omega_L(T,N,b)(\overline{M}+C_\varepsilon)\left[\sup_{|z|\leqslant L_T}\max_{1\leqslant i\leqslant N}\sum_{j=1}^{N} I(\sigma_{u,ij}(z) > C_\varepsilon \omega_L(T,N,b))\right] \\
&= O_P\left(\omega_L(T,N,b)\right)\left[\sup_{|z|\leqslant L_T}\max_{1\leqslant i\leqslant N}\sum_{j=1}^{N}\frac{|\sigma_{u,ij}(z)|^q}{[C_\varepsilon\omega_L(T,N,b)]^q}\right] \\
&= O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right) \\
&= O_P\left(m_N\left[\omega(T,N,b)/a_T(f)\right]^{1-q}\right). \qquad (A.12)
\end{aligned}
$$

on the event $\mathcal{B}(C_\varepsilon)$, where $\overline{M}$ is defined in Assumption 4(c).

By the triangle inequality, the events $\left\{|\check{\sigma}_{u,ij}(z)| \leqslant \rho(z)\right\}$ and $\mathcal{B}(C_\varepsilon)$ imply

$$
\left\{|\sigma_{u,ij}(z)| \leqslant \left(\overline{M}+C_\varepsilon\right)\omega_L(T,N,b)\right\}.
$$

Then, for Π_3, we may show that

$$
\begin{aligned}
\Pi_3 &\leqslant \sup_{|z|\leqslant L_T}\max_{1\leqslant i\leqslant N}\sum_{j=1}^{N}|\sigma_{u,ij}(z)| I\left(|\sigma_{u,ij}(z)| \leqslant \left(\overline{M}+C_\varepsilon\right)\omega_L(T,N,b)\right) \\
&\leqslant \left(\overline{M}+C_\varepsilon\right)^{1-q}\left[\omega_L(T,N,b)\right]^{1-q}\sup_{|z|\leqslant L_T}\max_{1\leqslant i\leqslant N}\sum_{j=1}^{N}|\sigma_{u,ij}(z)|^q \\
&= O\left(m_N\left[\omega(T,N,b)/a_T(f)\right]^{1-q}\right). \qquad (A.13)
\end{aligned}
$$

By (A.12) and (A.13), and letting $\varepsilon \to 0$ in (A.11), we prove (3.7) in Theorem 1. ■

Proof of Theorem 2. Throughout the proof, we let M be a generic positive constant whose value may change from line to line, and define

$$
\mathbf{\Sigma}_{X,T}(z) = \mathbf{B}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{B}(z)' + \mathbf{\Sigma}_u(z) \text{ with } \mathbf{\Sigma}_{F,T}(z) = \frac{1}{T}\sum_{t=1}^{T}F_t(z)F_t(z)',
$$

where $F_t(z)$ is defined in Assumption 2(b). Note that

$$
\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_X(z)-\mathbf{\Sigma}_X(z)\right\|^2_{\mathbf{\Sigma}_X(z)} \leqslant M\sup_{|z|\leqslant L_T}\left[\left\|\widetilde{\mathbf{\Sigma}}_X(z)-\mathbf{\Sigma}_{X,T}(z)\right\|^2_{\mathbf{\Sigma}_X(z)} + \|\mathbf{\Sigma}_{X,T}(z)-\mathbf{\Sigma}_X(z)\|^2_{\mathbf{\Sigma}_X(z)}\right]. \qquad (A.14)
$$

In addition, since all the eigenvalues of $\mathbf{\Sigma}_X(z)$ are bounded away from 0, for any $N \times N$ matrix $\mathbf{A}$,

$$
\|\mathbf{A}\|^2_{\mathbf{\Sigma}_X(z)} = \frac{1}{N}\|\mathbf{\Sigma}_X^{-1/2}(z)\mathbf{A}\mathbf{\Sigma}_X^{-1/2}(z)\|^2_F \leqslant \frac{M}{N}\|\mathbf{A}\|^2_F. \qquad (A.15)
$$

We start with the first term on the right hand side of (A.14). By the definitions of $\widetilde{\mathbf{\Sigma}}_X(z)$ and $\mathbf{\Sigma}_{X,T}(z)$, we

readily have that

$$\sup_{|z|\leqslant L_T}\left\|\widetilde{\boldsymbol{\Sigma}}_X(z)-\boldsymbol{\Sigma}_{X,T}(z)\right\|^2_{\boldsymbol{\Sigma}_X(z)}\leqslant M\sup_{|z|\leqslant L_T}\left[\left\|\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)'-\mathbf{B}(z)\boldsymbol{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)}+\left\|\widetilde{\boldsymbol{\Sigma}}_u(z)-\boldsymbol{\Sigma}_u(z)\right\|^2_{\boldsymbol{\Sigma}_X(z)}\right].$$

By (A.15) and Theorem 1, we can prove that

$$\begin{aligned}\sup_{|z|\leqslant L_T}\left\|\widetilde{\boldsymbol{\Sigma}}_u(z)-\boldsymbol{\Sigma}_u(z)\right\|^2_{\boldsymbol{\Sigma}_X(z)} &= O_P\left(\frac{1}{N}\sup_{|z|\leqslant L_T}\left\|\widetilde{\boldsymbol{\Sigma}}_u(z)-\boldsymbol{\Sigma}_u(z)\right\|^2_F\right)\\ &= O_P\left(\sup_{|z|\leqslant L_T}\left\|\widetilde{\boldsymbol{\Sigma}}_u(z)-\boldsymbol{\Sigma}_u(z)\right\|^2\right)\\ &= O_P\left(m_N^2\left[\omega_L(T,N,b)\right]^{2-2q}\right),\end{aligned}\tag{A.16}$$

where $\omega_L(T,N,b)=\omega(T,N,b)/a_T(f)$ as in the proof of Theorem 1.

Letting $\boldsymbol{\Delta}_B(z)=\widetilde{\mathbf{B}}(z)-\mathbf{B}_H(z)$ with $\mathbf{B}_H(z)=\mathbf{B}(z)\mathbf{H}^{-1}(z)$ and $\mathbf{H}(z)$ defined in the proof of Lemma 2 in Appendix B, we have

$$\begin{aligned}&\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)'-\mathbf{B}(z)\boldsymbol{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)}\\ \leqslant\ & M\sup_{|z|\leqslant L_T}\left[\left\|\boldsymbol{\Delta}_B(z)\boldsymbol{\Delta}_B(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)}+\left\|\mathbf{B}_H(z)\boldsymbol{\Delta}_B(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)}+\left\|\mathbf{B}_H(z)\mathbf{B}_H(z)'-\mathbf{B}(z)\boldsymbol{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)}\right].\end{aligned}$$

By (A.15) and (S.21) in the proof of Lemma 2, we may show that

$$\begin{aligned}\sup_{|z|\leqslant L_T}\left\|\boldsymbol{\Delta}_B(z)\boldsymbol{\Delta}_B(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)} &\leqslant M\sup_{|z|\leqslant L_T}\frac{1}{N}\|\boldsymbol{\Delta}_B(z)\|_F^4\\ &\leqslant MN\sup_{|z|\leqslant L_T}\max_{1\leqslant i\leqslant N}\left\|\widetilde{B}_i(z)-\left[\mathbf{H}^{-1}(z)\right]'B_i(z)\right\|^4\\ &= O_P\left(N\left[\omega_L(T,N,b)\right]^4\right).\end{aligned}\tag{A.17}$$

By Assumptions 2(a)(c) and using the argument in the proof of Theorem 2 in Fan, Fan and Lv (2008), we have

$$\sup_{|z|\leqslant L_T}\left\|\mathbf{B}(z)'\boldsymbol{\Sigma}_X^{-1}(z)\mathbf{B}(z)\right\|=O(1).\tag{A.18}$$

By (A.18), (S.21) and following the proof of Lemma 13 in Fan, Liao and Mincheva (2013), we have

$$\begin{aligned}\sup_{|z|\leqslant L_T}\left\|\mathbf{B}_H(z)\boldsymbol{\Delta}_B(z)'\right\|^2_{\boldsymbol{\Sigma}_X(z)} &= \frac{1}{N}\sup_{|z|\leqslant L_T}\text{trace}\left\{\mathbf{H}^{-1}(z)\boldsymbol{\Delta}_B(z)'\boldsymbol{\Sigma}_X^{-1}(z)\boldsymbol{\Delta}_B(z)\left[\mathbf{H}^{-1}(z)\right]'\mathbf{B}(z)'\boldsymbol{\Sigma}_X^{-1}(z)\mathbf{B}(z)\right\}\\ &\leqslant \frac{M}{N}\sup_{|z|\leqslant L_T}\left\|\mathbf{H}^{-1}(z)\right\|^2\|\boldsymbol{\Delta}_B(z)\|_F^2=O_P\left(\left[\omega_L(T,N,b)\right]^2\right).\end{aligned}\tag{A.19}$$

Similarly to the proof of (A.19) above, using (S.15) in the proof of Lemma 2, we can also prove that

$$\begin{aligned}
& \sup_{|z|\leqslant L_T} \left\|\mathbf{B}_H(z)\mathbf{B}_H(z)' - \mathbf{B}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|^2_{\mathbf{\Sigma}_X(z)} \\
= & \sup_{|z|\leqslant L_T} \left\|\mathbf{B}_H(z)\left[\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t(z)\widetilde{F}_t(z)' - \frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(z)F_t(z)F_t(z)'\mathbf{H}(z)'\right]\mathbf{B}_H(z)'\right\|^2_{\mathbf{\Sigma}_X(z)} \\
\leqslant & \frac{M}{N}\sup_{|z|\leqslant L_T}\left\|\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t(z)\widetilde{F}_t(z)' - \frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(z)F_t(z)F_t(z)'\mathbf{H}(z)'\right\|^2_F \cdot \left\|\mathbf{H}^{-1}(z)\right\|^4 \\
= & O_P\left(\frac{1}{N}\left[\omega_L(T,N,b)\right]^2\right). \qquad \text{(A.20)}
\end{aligned}$$

With (A.17), (A.19) and (A.20), we have

$$\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' - \mathbf{B}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|^2_{\mathbf{\Sigma}_X(z)} = O_P\left(N\left[\omega_L(T,N,b)\right]^4 + \left[\omega_L(T,N,b)\right]^2\right), \qquad \text{(A.21)}$$

which, together with (A.16), indicates that

$$\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_{X,T}(z)\right\|^2_{\mathbf{\Sigma}_X(z)} = O_P\left(N\left[\omega(T,N,b)/a_T(f)\right]^4 + m_N^2\left[\omega(T,N,b)/a_T(f)\right]^{2-2q}\right). \qquad \text{(A.22)}$$

On the other hand, by (3.2) in Assumption 2(b), following the proof of (A.20), we have

$$\begin{aligned}
\sup_{|z|\leqslant L_T}\left\|\mathbf{\Sigma}_{X,T}(z) - \mathbf{\Sigma}_X(z)\right\|^2_{\mathbf{\Sigma}_X(z)} &= O_P\left(\frac{1}{Na_T^2(f)}\left(\sqrt{\frac{\log T}{Tb}} + b^2\right)^2\right) \\
&= o_P\left(N\left[\omega(T,N,b)/a_T(f)\right]^4 + m_N^2\left[\omega(T,N,b)/a_T(f)\right]^{2-2q}\right),
\end{aligned}$$

which, together with (A.4) and (A.22), leads to (3.8).

We next turn to the proof of (3.9). Note that

$$\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_X(z)\right\|_{\max} \leqslant \sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_{X,T}(z)\right\|_{\max} + \sup_{|z|\leqslant L_T}\left\|\mathbf{\Sigma}_{X,T}(z) - \mathbf{\Sigma}_X(z)\right\|_{\max},$$

where

$$\sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_X(z) - \mathbf{\Sigma}_{X,T}(z)\right\|_{\max} \leqslant \sup_{|z|\leqslant L_T}\left[\left\|\widetilde{\mathbf{\Sigma}}_u(z) - \mathbf{\Sigma}_u(z)\right\|_{\max} + \left\|\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' - \mathbf{B}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|_{\max}\right].$$

Using Lemma 4, the property of $\left|s_{\rho(z)}(w) - w\right| \leqslant \rho(z)$ and Assumption 4(c), we have

$$\begin{aligned}
& \sup_{|z|\leqslant L_T}\left\|\widetilde{\mathbf{\Sigma}}_u(z) - \mathbf{\Sigma}_u(z)\right\|_{\max} \\
= & \sup_{|z|\leqslant L_T}\max_{1\leqslant i,j\leqslant N}\left|\widetilde{\sigma}_{u,ij}(z) - \sigma_{u,ij}(z)\right|
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sup_{|z|\leqslant L_T} \max_{1\leqslant i,j\leqslant N} \left[\left|s_{\rho(z)}(\check{\sigma}_{u,ij}(z)) - \check{\sigma}_{u,ij}(z)\right| + \left|\check{\sigma}_{u,ij}(z) - \sigma_{u,ij}(z)\right|\right] \\
&= O_P\left(\omega_L(T,N,b)\right). \qquad \text{(A.23)}
\end{aligned}
$$

On the other hand, by (S.21) in the proof of Lemma 2, similarly to the proof of (A.21), we have

$$
\begin{aligned}
&\sup_{|z|\leqslant L_T} \left\|\widetilde{\mathbf{B}}(z)\widetilde{\mathbf{B}}(z)' - \mathbf{B}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{B}(z)'\right\|_{max} \\
= &\sup_{|z|\leqslant L_T} \max_{1\leqslant i,j\leqslant N} \left|\widetilde{B}_i(z)'\widetilde{B}_j(z) - B_i(z)'\mathbf{\Sigma}_{F,T}(z)B_j(z)\right| \\
\leqslant &\sup_{|z|\leqslant L_T} \max_{1\leqslant i,j\leqslant N} \Big\{ \left|\Delta_{B,i}(z)'\Delta_{B,j}(z)\right| + \left|\Delta_{B,i}(z)'[\mathbf{H}^{-1}(z)]'B_j(z)\right| + \left|B_i(z)'\mathbf{H}^{-1}(z)\Delta_{B,j}(z)\right| + \\
&\left|B_i(z)'\mathbf{H}^{-1}(z)[\mathbf{H}^{-1}(z)]'B_j(z) - B_i(z)'\mathbf{\Sigma}_{F,T}(z)B_j(z)\right| \Big\} \\
= &O_P\left(\omega_L(T,N,b)\right), \qquad \text{(A.24)}
\end{aligned}
$$

where $\Delta_{B,i}(z) = \widetilde{B}_i(z) - \left[\mathbf{H}^{-1}(z)\right]' B_i(z)$. By (A.23), (A.24) and noting that

$$
\sup_{|z|\leqslant L_T} \|\mathbf{\Sigma}_{X,T}(z) - \mathbf{\Sigma}_X(z)\|_{max} = O_P\left(a_T^{-1}(f)\left(\sqrt{\frac{\log T}{Tb}} + b^2\right)\right)
$$

from Assumptions 2(a)(b), we prove (3.9). The proof of Theorem 2 has been completed. ■

Proof of Theorem 3. By the triangle inequality, we readily have that

$$
\left\|\widetilde{\mathbf{\Sigma}}_X^{-1}(z) - \mathbf{\Sigma}_X^{-1}(z)\right\| \leqslant \left\|\widetilde{\mathbf{\Sigma}}_X^{-1}(z) - \mathbf{\Sigma}_{X,T}^{-1}(z)\right\| + \left\|\mathbf{\Sigma}_{X,T}^{-1}(z) - \mathbf{\Sigma}_X^{-1}(z)\right\|, \qquad \text{(A.25)}
$$

where $\mathbf{\Sigma}_{X,T}(z)$ is defined as in the proof of Theorem 2.

By (2.2) and the Sherman-Morrison-Woodbury formula, we have

$$
\begin{aligned}
\mathbf{\Sigma}_X^{-1}(z) &= \mathbf{\Sigma}_u^{-1}(z) - \mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\mathbf{\Sigma}_F^{1/2}\left[\mathbf{I}_K + \mathbf{\Sigma}_F^{1/2}\mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\mathbf{\Sigma}_F^{1/2}\right]^{-1}\mathbf{\Sigma}_F^{1/2}\mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z) \\
&= \mathbf{\Sigma}_u^{-1}(z) - \mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\left[\mathbf{\Sigma}_F^{-1} + \mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\right]^{-1}\mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z),
\end{aligned}
$$

and similarly

$$
\mathbf{\Sigma}_{X,T}^{-1}(z) = \mathbf{\Sigma}_u^{-1}(z) - \mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\left[\mathbf{\Sigma}_{F,T}^{-1}(z) + \mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\right]^{-1}\mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z),
$$

where $\mathbf{\Sigma}_{F,T}(z) = \frac{1}{T}\sum_{t=1}^{T} F_t(z)F_t(z)'$. Then, it is easy to show that

$$
\mathbf{\Sigma}_{X,T}^{-1}(z) - \mathbf{\Sigma}_X^{-1}(z) = -\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\mathbf{\Omega}_T(z)\mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z) \qquad \text{(A.26)}
$$

with

$$
\mathbf{\Omega}_T(z) = \left[\mathbf{\Sigma}_{F,T}^{-1}(z) + \mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\right]^{-1} - \left[\mathbf{\Sigma}_F^{-1} + \mathbf{B}(z)'\mathbf{\Sigma}_u^{-1}(z)\mathbf{B}(z)\right]^{-1} = \mathbf{\Omega}_{T,*}^{-1}(z) - \mathbf{\Omega}_*^{-1}(z).
$$

Note that

$$\begin{aligned}\left\|\mathbf{\Omega}_{T,*}^{-1}(z)-\mathbf{\Omega}_{*}^{-1}(z)\right\| &= \left\|\mathbf{\Omega}_{T,*}^{-1}(z)\left[\mathbf{\Omega}_{*}(z)-\mathbf{\Omega}_{T,*}(z)\right]\mathbf{\Omega}_{*}^{-1}(z)\right\| \\ &\leqslant \left[\left\|\mathbf{\Omega}_{T,*}^{-1}(z)-\mathbf{\Omega}_{*}^{-1}(z)\right\|+\left\|\mathbf{\Omega}_{*}^{-1}(z)\right\|\right]\|\mathbf{\Omega}_{*}(z)-\mathbf{\Omega}_{T,*}(z)\|\left\|\mathbf{\Omega}_{*}^{-1}(z)\right\|,\end{aligned}$$

leading to

$$\|\mathbf{\Omega}_{T}(z)\|=\left\|\mathbf{\Omega}_{T,*}^{-1}(z)-\mathbf{\Omega}_{*}^{-1}(z)\right\|\leqslant\frac{\|\mathbf{\Omega}_{T,*}(z)-\mathbf{\Omega}_{*}(z)\|\left\|\mathbf{\Omega}_{*}^{-1}(z)\right\|^{2}}{1-\|\mathbf{\Omega}_{T,*}(z)-\mathbf{\Omega}_{*}(z)\|\left\|\mathbf{\Omega}_{*}^{-1}(z)\right\|}. \tag{A.27}$$

By (3.2) in Assumption 2(b), we have

$$\sup_{|z|\leqslant L_T}\|\mathbf{\Omega}_{T,*}(z)-\mathbf{\Omega}(z)\|=O_P\left(a_T^{-1}(f)\left(\sqrt{\frac{\log T}{Tb}}+b^2\right)\right).$$

Meanwhile, by Assumption 2, we can prove that $\left\|\mathbf{\Omega}_{*}^{-1}(z)\right\|=O(1/N)$ and $\left\|\mathbf{\Sigma}_{u}^{-1}(z)\mathbf{B}(z)\right\|=O(N^{1/2})$ uniformly over $|z|\leqslant L_T$. Then, by (A.26) and (A.27), we have

$$\sup_{|z|\leqslant L_T}\left\|\mathbf{\Sigma}_{X,T}^{-1}(z)-\mathbf{\Sigma}_{X}^{-1}(z)\right\|=O_P\left(\frac{1}{Na_T(f)}\left(\sqrt{\frac{\log T}{Tb}}+b^2\right)\right)=o_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right), \tag{A.28}$$

where $\omega_L(T,N,b)=\omega(T,N,b)/a_T(f)$.

We next derive the asymptotic order of $\left\|\widetilde{\mathbf{\Sigma}}_{X}^{-1}(z)-\mathbf{\Sigma}_{X,T}^{-1}(z)\right\|$. By the Sherman-Morrison-Woodbury formula again, we have

$$\begin{aligned}\widetilde{\mathbf{\Sigma}}_{X}^{-1}(z) &= \widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)-\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\left[\mathbf{I}_K+\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z), \\ \mathbf{\Sigma}_{X,T}^{-1}(z) &= \mathbf{\Sigma}_{u}^{-1}(z)-\mathbf{\Sigma}_{u}^{-1}(z)\mathbf{B}_H(z)\left[\mathbf{\Sigma}_{F,H}^{-1}(z)+\mathbf{B}_H(z)'\mathbf{\Sigma}_{u}^{-1}(z)\mathbf{B}_H(z)\right]^{-1}\mathbf{B}_H(z)'\mathbf{\Sigma}_{u}^{-1}(z),\end{aligned}$$

where $\mathbf{B}_H(z)=\mathbf{B}(z)\mathbf{H}^{-1}(z)$ as in the proof of Theorem 2 and $\mathbf{\Sigma}_{F,H}(z)=\mathbf{H}(z)\mathbf{\Sigma}_{F,T}(z)\mathbf{H}(z)'$. By some standard arguments, we may show that

$$\widetilde{\mathbf{\Sigma}}_{X}^{-1}(z)-\mathbf{\Sigma}_{X,T}^{-1}(z)=-\sum_{r=1}^{6}\mathbf{\Lambda}_r(z),$$

where

$$\begin{aligned}\mathbf{\Lambda}_1(z) &= \mathbf{\Sigma}_{u}^{-1}(z)-\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z), \\ \mathbf{\Lambda}_2(z) &= \left[\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)-\mathbf{\Sigma}_{u}^{-1}(z)\right]\widetilde{\mathbf{B}}(z)\left[\mathbf{I}_K+\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z), \\ \mathbf{\Lambda}_3(z) &= \mathbf{\Sigma}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\left[\mathbf{I}_K+\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\widetilde{\mathbf{B}}(z)'\left[\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)-\mathbf{\Sigma}_{u}^{-1}(z)\right], \\ \mathbf{\Lambda}_4(z) &= \mathbf{\Sigma}_{u}^{-1}(z)\left[\widetilde{\mathbf{B}}(z)-\mathbf{B}_H(z)\right]\left[\mathbf{I}_K+\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\widetilde{\mathbf{B}}(z)'\mathbf{\Sigma}_{u}^{-1}(z), \\ \mathbf{\Lambda}_5(z) &= \mathbf{\Sigma}_{u}^{-1}(z)\mathbf{B}_H(z)\left[\mathbf{I}_K+\widetilde{\mathbf{B}}(z)'\widetilde{\mathbf{\Sigma}}_{u}^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\left[\widetilde{\mathbf{B}}(z)-\mathbf{B}_H(z)\right]'\mathbf{\Sigma}_{u}^{-1}(z),\end{aligned}$$

$$\boldsymbol{\Lambda}_6(z) \quad = \quad \boldsymbol{\Sigma}_u^{-1}(z)\mathbf{B}_H(z)\boldsymbol{\Omega}_H(z)\mathbf{B}_H(z)'\boldsymbol{\Sigma}_u^{-1}(z)$$

with $\boldsymbol{\Omega}_H(z) = \left[\mathbf{I}_K + \widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1} - \left[\boldsymbol{\Sigma}_{F,H}^{-1}(z) + \mathbf{B}_H(z)'\boldsymbol{\Sigma}_u^{-1}(z)\mathbf{B}_H(z)\right]^{-1}$. By Assumption 2(c), (A.27) and Theorem 1, we have

$$\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_1(z)\| = \sup_{|z|\leqslant L_T} \left\|\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z) - \boldsymbol{\Sigma}_u^{-1}(z)\right\| = O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right).$$

Following the proof of Lemma 15 in Fan, Liao and Mincheva (2013), we have

$$\sup_{|z|\leqslant L_T} \left\|\left[\mathbf{I}_K + \widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\right\| = O_P(1/N).$$

Then, we can prove that

$$\begin{aligned}\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_2(z)\| \quad &\leqslant \quad \sup_{|z|\leqslant L_T} \left\{\left\|\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z) - \boldsymbol{\Sigma}_u^{-1}(z)\right\| \left\|\widetilde{\mathbf{B}}(z)\left[\mathbf{I}_K + \widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1}\widetilde{\mathbf{B}}(z)'\right\|^{-1}\left\|\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\right\|\right\} \\ &= \quad O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right),\end{aligned}$$

and similarly

$$\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_3(z)\| = O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right).$$

By (S.21) in the proof of Lemma 2, we can show that

$$\begin{aligned}\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_4(z)\| \quad &\leqslant \quad \sup_{|z|\leqslant L_T} \left\{\left\|\boldsymbol{\Sigma}_u^{-1}(z)\left[\widetilde{\mathbf{B}}(z) - \mathbf{B}_H(z)\right]\right\| \left\|\mathbf{I}_K + \widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\widetilde{\mathbf{B}}(z)\right\|^{-1}\left\|\widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\right\|\right\} \\ &= \quad O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right),\end{aligned}$$

and similarly

$$\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_5(z)\| = O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right).$$

Note that

$$\boldsymbol{\Omega}_H(z) = \left[\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t(z)\widetilde{F}_t(z)' + \widetilde{\mathbf{B}}(z)'\widetilde{\boldsymbol{\Sigma}}_u^{-1}(z)\widetilde{\mathbf{B}}(z)\right]^{-1} - \left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{H}(z)F_t(z)F_t(z)'\mathbf{H}(z)' + \mathbf{B}_H(z)'\boldsymbol{\Sigma}_u^{-1}(z)\mathbf{B}_H(z)\right]^{-1}.$$

Using (A.27), (S.15), (S.21) and Theorem 1 and following the argument in the proof of (A.28), we have

$$\sup_{|z|\leqslant L_T} \|\boldsymbol{\Omega}_H(z)\| = O_P\left(\frac{m_N}{N}\left[\omega_L(T,N,b)\right]^{1-q}\right),$$

and consequently

$$\sup_{|z|\leqslant L_T} \|\boldsymbol{\Lambda}_6(z)\| = O_P\left(m_N\left[\omega_L(T,N,b)\right]^{1-q}\right).$$

Combining the above results, we may prove that

$$\sup_{|z|\leqslant L_T} \left\| \widetilde{\mathbf{\Sigma}}_X^{-1}(z) - \mathbf{\Sigma}_{X,T}^{-1}(z) \right\| = O_P\left(m_N \left[\omega_L(T,N,b)\right]^{1-q} \right). \tag{A.29}$$

In view of (A.25), (A.28) and (A.29), we prove (3.14). ■

Proof of Theorem 4. By the definition of $\overline{\mathbf{\Sigma}}_X(z)$ in (2.5) and the factor model structure (2.1), we have

$$\begin{aligned}
\overline{\mathbf{\Sigma}}_X(z) &= \frac{1}{T}\sum_{t=1}^{T} X_t(z)X_t(z)' \\
&= \frac{1}{T}\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)F_t(z)'\mathbf{B}(z_t)' + \frac{1}{T}\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)u_t(z)' + \\
& \quad \frac{1}{T}\sum_{t=1}^{T} u_t(z)F_t(z)'\mathbf{B}(z_t)' + \frac{1}{T}\sum_{t=1}^{T} u_t(z)u_t(z)'.
\end{aligned}$$

By Assumption 2(c), $\nu(T,N,b) + b^2 = o(a_T(f))$ and Lemma 3, we may show that

$$\sup_{|z|\leqslant L_T} \left\| \frac{1}{T}\sum_{t=1}^{T} u_t(z)u_t(z)' \right\| = o_P(N).$$

By Assumption 2(a) and using the argument in the proof of (S.17) in Appendix B, we have

$$\sup_{|z|\leqslant L_T} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)u_t(z)' + \frac{1}{T}\sum_{t=1}^{T} u_t(z)F_t(z)'\mathbf{B}(z_t)' \right\| = o_P(N).$$

By Assumptions 2(a)(b), we can prove that, with probability approaching one,

$$\min_{1\leqslant k\leqslant K} \inf_{|z|\leqslant L_T} \lambda_k\left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)F_t(z)'\mathbf{B}(z_t)' \right) \geqslant \underline{\zeta}N$$

and

$$\max_{1\leqslant k\leqslant K} \sup_{|z|\leqslant L_T} \lambda_k\left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)F_t(z)'\mathbf{B}(z_t)' \right) \leqslant \overline{\zeta}N,$$

where $\lambda_k(\cdot)$ denotes the k-th largest eigenvalue, $\underline{\zeta}$ and $\overline{\zeta}$ are two positive constants. Meanwhile, the k-th eigenvalue of $\sum_{t=1}^{T} \mathbf{B}(z_t)F_t(z)F_t(z)'\mathbf{B}(z_t)'$ is zero if $k \geqslant K+1$. Then, by Weyl's eigenvalue inequality (e.g., Horn and Johnson, 1991), we may show that

$$\frac{1}{2}\underline{\zeta}N \leqslant \min_{1\leqslant k\leqslant K} \inf_{|z|\leqslant L_T} \lambda_k\left(\overline{\mathbf{\Sigma}}_X(z)\right) \leqslant \max_{1\leqslant k\leqslant K} \sup_{|z|\leqslant L_T} \lambda_k\left(\overline{\mathbf{\Sigma}}_X(z)\right) \leqslant 2\overline{\zeta}N \tag{A.30}$$

with probability approaching one, and

$$\max_{K+1 \leqslant k \leqslant \overline{K}} \sup_{|z| \leqslant L_T} \lambda_k \left(\overline{\mathbf{\Sigma}}_X(z) \right) = o_P(N). \tag{A.31}$$

For $1 \leqslant k \leqslant K-1$, by (A.30), we have $\inf_{|z| \leqslant L_T} \left| \overline{\lambda}_{k+1,z} / \overline{\lambda}_{k,z} \right|$ is bounded away from zero; for $K+1 \leqslant k \leqslant \overline{K}$, by (4.2) and (A.31), we have $\overline{\lambda}_{k+1,z} / \overline{\lambda}_{k,z} = 1$ uniformly over $|z| \leqslant L_T$; and finally, for $k = K$, by (A.30) and (A.31), we have

$$\sup_{|z| \leqslant L_T} \frac{\overline{\lambda}_{k+1,z}}{\overline{\lambda}_{k,z}} \leqslant \frac{\sup_{|z| \leqslant L_T} \overline{\lambda}_{k+1,z}/N}{\inf_{|z| \leqslant L_T} \overline{\lambda}_{k,z}/N} = o_P(1).$$

Combining the above results, we have $\widehat{K}(z) = K$ uniformly over $|z| \leqslant L_T$ with probability approaching one, indicating that $P\left(\widehat{K} = K\right) \to 1$ by (4.3) and noting that $\mathcal{Z}$ is a subset of $\{z : |z| \leqslant L_T\}$. The proof of Theorem 4 is now completed. ■

References

AHN, S. AND HORENSTEIN, A. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

BAI, J. AND NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **90**, 191–221.

BARIGOZZI, M., CHO, H. AND FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics* **206**, 187–225.

BICKEL, P. AND LEVINA, E. (2008a). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

BICKEL, P. AND LEVINA, E. (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (2nd Edition). Lecture Notes in Statistics 110, Springer-Verlag, Berlin.

BREITUNG, J. AND EICKMEIER, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* **163**, 71–84.

CAI, T. T. AND LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

CAI, T. T., REN, Z. AND ZHOU, H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10**, 1–59.

CHAMBERLAIN, G. AND ROTHSCHILD, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* **51**, 1305–1324.

CHEN, J., LI, D. AND LINTON, O. (2019). A new semiparametric estimation approach of large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics* **212**, 155–176.

CHEN, L., DOLADO, J. J. AND GONZALO, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics* **180**, 30–48.

CHEN, X., XU, M. AND WU, W. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Annals of Statistics* **41**, 2994–3021.

CHEN, Z. AND LENG, C. (2016). Dynamic covariance models. *Journal of the American Statistical Association* **111**, 1196–1207.

CHENG, X., LIAO, Z. AND SCHORFHEIDE, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies* **83**, 1511–1543.

ENGLE, R.F., LEDOIT, O. AND WOLF, M. (2019). Large dynamic covariance matrices. *Journal of Business and Economic Statistics* **37**, 363–375.

FAN, J., FAN, Y. AND LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147**, 186–197.

FAN, J., LIAO, Y. AND LIU, H. (2016). An overview on the estimation of large covariance and precision matrices. *Econometrics Journal* **19**, 1–32.

FAN, J., LIAO, Y. AND MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B* **75**, 603–680.

FAN, J. AND YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics, Springer.

GUO, S., BOX, J. AND ZHANG, W. (2017). A dynamic structure for high dimensional covariance matrices and its application in portfolio allocation. *Journal of the American Statistical Association* **112**, 235–253.

HANSEN, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

HORN, R. AND JOHNSON, C. (1991). *Topics in Matrix Analysis*. Cambridge University Press.

LAM, C. AND FAN, J. (2009). Sparsity and rates of convergence in large covariance matrix estimation. *Annals of Statistics* **37**, 4254–4278.

LAM, C. AND YAO, Q. (2012). Factor modelling for high-dimensional time series: Inference for the number of factor. *Annals of Statistics* **40**, 694–726.

LI, D., LU, Z. AND LINTON, O. (2012). Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* **28**, 935–958.

LI, Q. AND RACINE, J. (2007). *Nonparametric Econometrics*. Princeton University Press, Princeton.

MA, S. AND SU, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics* **207**, 1–29.

MOON, H. R. AND WEIDNER, M. (2015), Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* **83**, 1543–1579.

MOTTA, G., HAFNER, C. AND VON SACHS, R. (2011). Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* **27**, 1279–1319.

PELGER, M. AND XIONG, R. (2020). State-varying factor models of large dimensions. Working paper available at https://arxiv.org/pdf/1807.02248.pdf.

POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Wiley.

ROBINSON, P. M. (1989). Nonparametric estimation of time-varying parameters. *Statistical Analysis and Forecasting of Economic Structural Change* (ed: P. Hackl). Springer, Berlin, pp. 164–253.

ROTHMAN, A. J., LEVINA, E. AND ZHU, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104**, 177–186.

STOCK, J. H. AND WATSON, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry* (ed: J. Castle and N. Shephard). Oxford University Press, pp. 173–205.

SU, L. AND WANG, X. (2017). On time-varying factor models: estimation and testing. *Journal of Econometrics* **198**, 84–101.