DSSR

Discussion Paper No. 130

Estimation of Large Covariance Matrices with Mixed Factor Structures

Runyu Dai, Yoshimasa Uematsu and Yasumasa Matsuda

August 24, 2022

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Estimation of Large Covariance Matrices with Mixed Factor Structures

Runyu Dai*, Yoshimasa Uematsu†* and Yasumasa Matsuda*

*Department of Economics and Management, Tohoku University

†Center for the Promotion of Social Data Science Education and Research, Hitotsubashi University

August 24, 2022

Abstract

We extend the Principal Orthogonal complEment Thresholding (POET) framework introduced by Fan et al. (2013) to estimate large static covariance matrices with a "mixed" structure of observable and unobservable common factors, and we call this method the extended POET (ePOET). A stable covariance estimator for large-scale data is developed by combining observable factors and sparsity-induced weak latent factors, with an adaptive threshold estimator of idiosyncratic covariance. Under some mild conditions, we derive the uniform consistency of the proposed estimator for the cases with or without observable factors. Furthermore, several simulation studies show that the ePOET achieves good finite-sample performance regardless of data with strong, weak, or mixed factors structure. Finally, we conduct empirical studies to present the practical usefulness of the ePOET.

Keywords: Sparsity-induced weak factor model, SOFAR estimator, Factor error structure, Sparse covariance matrix, Thresholding.

1 Introduction

Covariance matrix estimation plays an essential role in a wide range of fields, including finance and economics. With advances in computer technology, many high-dimensional data, where the cross-sectional dimension N is close to or larger than the sample size T, have become increasingly accessible. The simplest estimator of the covariance matrix is the naive sample covariance, but it is prone to instability or even a singularity problem in a high-dimensional setting. Alternatively, many covariance estimators have been proposed; for example, Bickel and Levina (2008a,b), Rothman et al. (2009), Cai and Liu (2011), Ledoit and Wolf (2004), Ledoit and Wolf (2012), Lam (2016), and so forth.

Since economic and financial data frequently exhibit multicollinearity, a factor-based approach can be more appealing; see Fan et al. (2008) and Fan et al. (2011). Fan et al. (2013) extend the ideas and propose the well-known *principal orthogonal complement thresholding* (POET) estimator. Roughly speaking, the POET supposes the approximate factor model for the target and performs a principal component analysis (PCA) on the sample covariance matrix to extract the top K principal components as the signal part, and then applies a thresholding technique to the remaining noise part to obtain a sparse idiosyncratic covariance estimator. Related studies include, but not limited to, Fan et al. (2018) and Wang et al. (2021).

The POET and aforementioned factor-based approaches employ the so-called *pervasiveness* (*strong factor*) assumption, which makes the first K largest eigenvalues of data covariance matrix diverge proportionally to N while the others bounded. This results in exhibiting a single large gap between the Kth and $(K+1)$th largest eigenvalues. Although the strong factor assumption is simple and widely used in the literature, including Bai and Ng (2002) and Bai (2003), it is questionable whether the assumption is consistent with real data. In fact, in the discussion of Fan et al. (2013), Alexei Onatski says that it may be misleading in many economic and financial applications. He indicates that the absence of such a large gap may have a negative effect on the performance of POET. Lam et al. (2011) also point out the possibility that the PC estimator does not work well in the absence of such a clear eigen-separation. We reconsider this issue.

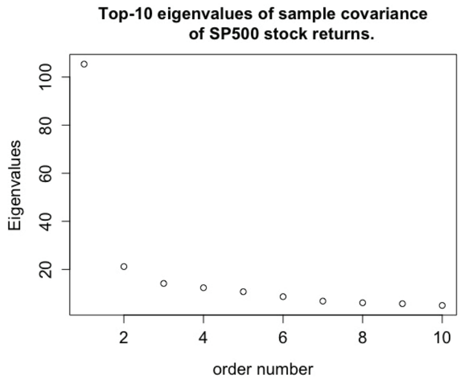


Figure 1: The sample eigenvalues of S&P monthly excess returns.

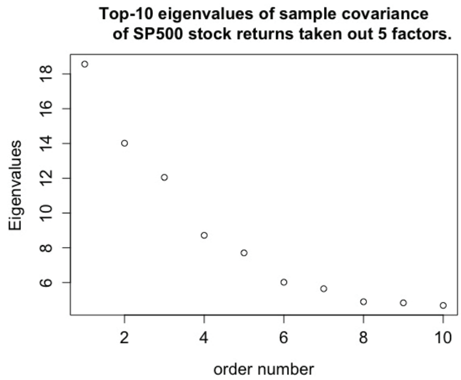


Figure 2: The sample eigenvalues of residuals of S&P monthly excess returns regressed by the five observable factors.

Figure 1 shows the scree plot of the sample covariance matrix of S&P 500 stock returns from Datastream over the period from May 1998 to April 2018, including 376 companies. In the figure, we can observe a huge eigen-gap between the first and the second eigenvalues, implying that there is only a single factor. Perhaps surprisingly, however, the edge distributions (ED) method of Onatski (2010) stably detects six factors. This suggests there are a single strong and several "weaker factors" in the data.

More interestingly, there still remain some unobserved "weak factors" even after taken out some observable factors. Figure 2 is the scree plot of the same data after taken out the five observable factors of Fama and French (2015).[1] For this sample covariance matrix,

[1]The Fama-French five factors refer to the market return, SMB, HML, RMW, CMA. Here, SMB means the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML is the difference between the returns on diversified portfolios of high and low B/M stocks,

Onatski's ED method detects five factors left behind with exhibiting no large gaps. This indicates the existence of "weak factors" that are not explained by the observable factors.

On the basis of the above observations, we propose the new framework that allows both observable factors and latent "weak factors," called the *extended POET* (*ePOET*). This modelling strategy is expected to enhance the interpretability and flexibility, compared with the original POET. Focusing on the latent "weak factor" part, we adopt the *sparsity-induced weak factor (sWF) models* by Uematsu and Yamagata (2022); namely, we suppose sparsity condition on the factor loadings that directly link to the magnitude of signal eigenvalues under a specific rotation. Then the kth largest eigenvalues of the covariance matrix diverges proportionally to N^{α_k} for some $\alpha_k \in (0, 1]$ for each $k \in \{1, \ldots, K\}$. The "weakness" is well-estimated as long as the loadings are sparsely estimated, and is indeed achieved by the *sparse orthogonal factor regression (SOFAR)* estimator of Uematsu et al. (2019) and Uematsu and Yamagata (2022).

To our ePOET, we derive the rates of convergence for the estimated covariance and precision matrices under strong (α-)mixing assumptions. Simulation studies show that the ePOET can ensure reliable estimation accuracy, whatever data are generated from observable, weak, strong, or "mixed" factor models. We conduct an empirical study about minimum variance portfolio (MVP) constructions. It is observed that, in terms of the out-of-sample risk, the proposed ePOET method with observable factors outperforms the other candidates. Finally, a heatmap comparison shows that the proposed method can well capture the potential WF structure in the residuals obtained from a financial dataset regressed by the Fama-French five factors.

RMW represents the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA stands for the difference between the returns on diversified portfolios of the stocks of low and high investment firm. They are obtained from the Kenneth R. French Data Library.

1.1 Notations and organization

Throughout the paper, we use $\|\mathbf{M}\|$, $\|\mathbf{M}\|_{\mathrm{F}}$, $\|\mathbf{M}\|_{\max}$, and $\|\mathbf{M}\|_1$ to denote l_2 norm, the Frobenius norm, the maximum norm and the elementwise l_1 norm for any matrix $\mathbf{M} = (m_{ti}) \in \mathbb{R}^{T\times N}$, respectively. Given a $N \times N$ positive definite matrix $\mathbf{\Sigma}$, the weighted quadratic norm of an $N \times N$ matrix $\mathbf{P}$ of $\mathbf{\Sigma}$ is defined as $\|\mathbf{P}\|_{\mathbf{\Sigma}} = N^{-1/2}\|\mathbf{\Sigma}^{-1/2}\mathbf{P}\mathbf{\Sigma}^{-1/2}\|_{\mathrm{F}}$. For any square matrix $\mathbf{A}$, we denote the largest, the smallest and the kth largest eigenvalues by $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$, and $\lambda_k(\mathbf{A})$, respectively. $\mathbf{I}_N$ means a $N \times N$ identity matrix. Let $\lesssim$ and $\gtrsim$ represent $\leqslant$ and $\geqslant$ up to a positive constant factor. For two positive values x and y, we use $x \wedge y$ and $x \vee y$ to denote $\min\{x, y\}$ and $\max\{x, y\}$, respectively. Finally, for two positive sequences a_n and b_n, we denote $a_n \asymp b_n$ if $a_n \gtrsim b_n$ and $a_n \lesssim b_n$.

The rest of the paper is organized as follows. Section 2 formally defines the ePOET model. Section 3 describes the estimation methodologies and the steps to determine the optimal number of factors. The theoretical results of ePOET are presented in Section 4. In Section 5, we conduct three sets of simulation studies are conducted. Two empirical applications are shown in Section 6. Section 7 concludes the paper. All the proofs of the theoretical results and additional discussions are reported in the appendix. Section B gives some remarks on the practical issues of choosing the tuning parameters.

2 Model

We consider estimation of the covariance matrix of N-dimensional vector $\mathbf{y}_t$, generated by the linear regression model

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \mathbf{u}_t, \tag{2.1}$$

where $\mathbf{x}_t$ and $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_r)$ represent the r-dimensional observable factor and its factor loadings, respectively. Furthermore, the error term $\mathbf{u}_t$ has the latent factor structure

$$\mathbf{u}_t = \mathbf{B}\mathbf{f}_t + \mathbf{e}_t, \tag{2.2}$$

where $\mathbf{f}_t$ and $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_K)$ represent the K-dimensional unobservable factor and its factor loadings, respectively, and $\mathbf{e}_t$ is the idiosyncratic error term. Stacking the observations for $t = 1, \ldots, T$, we can rewite the models as $\mathbf{Y} = \mathbf{X}\mathbf{A}' + \mathbf{U}$ and $\mathbf{U} = \mathbf{F}\mathbf{B}' + \mathbf{E}$. Similar factor structures can be found in Bai (2009), , Bai et al. (2016), Gagliardini et al. (2019) and Fan et al. (2021). Following the latter two papers, we impose the orthogonality conditions between observable factors $\mathbf{x}_t$ and latent parts $\mathbf{u}_t$. This assumption enable us to separately estimate $\mathbf{A}$ and $\mathbf{B}$.

Without loss of generality, we suppose the identification conditions,

$$E[\mathbf{f}_t\mathbf{f}_t'] = \mathbf{I}_K \quad \text{and} \quad \mathbf{B}'\mathbf{B} \text{ diagonal}, \tag{2.3}$$

throughout the paper. Denote by $\boldsymbol{\Sigma}_x$, $\boldsymbol{\Sigma}_u$, and $\boldsymbol{\Sigma}_e$ the covariance matrices of $\mathbf{x}_t$, $\mathbf{u}_t$, and $\mathbf{e}_t$, respectively. Given the condition that $\mathbf{e}_t$, $\mathbf{x}_t$, and $\mathbf{f}_t$ are mutually uncorrelated, the covariance matrix of $\mathbf{y}_t$ under (2.3) is

$$\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}' + \boldsymbol{\Sigma}_u, \quad \boldsymbol{\Sigma}_u = \mathbf{B}\mathbf{B}' + \boldsymbol{\Sigma}_e. \tag{2.4}$$

2.1 Sparsity-induced weak factor model

As discussed in Introduction, it is natural to allow a "weaker" factor structure in (2.2). Following Uematsu and Yamagata (2022), we formally introduce the sWF model. Assume that $\lambda_{\min}(\mathbf{B}'\mathbf{B})$ is bounded away from zero and $\lambda_{\max}(\boldsymbol{\Sigma}_e)$ is bounded from above. Then Weyl's inequality entails $\lambda_k(\boldsymbol{\Sigma}_u) \asymp \lambda_k(\mathbf{B}\mathbf{B}')$ for $k = 1, \ldots, K$. The latent factor model in (2.2) is called the *weak factor (WF) model* if

$$\lambda_k(\mathbf{B}\mathbf{B}') \asymp N^{\alpha_k}, \quad k = 1, \ldots, K \tag{2.5}$$

for some $\alpha_k \in (0, 1]$. The sWF model achieves (2.5) by imposing a sparsity assumption on $\mathbf{B}$ with the aid of (2.3). Specifically, suppose that $\mathbf{b}_k$ has $N_k := \lfloor N^{\alpha_k} \rfloor$ nonzero elements for some constant $\alpha_k \in (0, 1]$ for each $k = 1, \ldots, K$. Then the signal strength is controlled via the sparsity:

$$\lambda_k(\mathbf{B}\mathbf{B}') = \lambda_k(\mathbf{B}'\mathbf{B}) = \|\mathbf{b}_k\|_2^2 \asymp N_k, \quad k = 1, \ldots, K.$$

The sparsity assumption on $\mathbf{B}$ is not the only assumption that brings about the WF structure; e.g., the condition that $\mathbf{b}_k$ is dense with $b_{ik} \asymp N^{(\alpha_k - 1)/2}$ for all $i = 1, \ldots, N$ achieves the same rate, but we do not pursue this direction. See Uematsu and Yamagata (2021) for a statistical evidence of sparse loadings with macroeconomic and financial data.

3 Estimation Methodology

We propose a new estimation framework for $\boldsymbol{\Sigma}$ in (2.4). The procedure first apply the ordinary least squares (OLS) to (2.1) to obtain the residuals, and then estimate the residual covariance matrix applying the POET algorithm to (2.2) with replacing the PC estimate by the SOFAR estimate. The detailed procedure is shown as follows:

Step 1. Regress $\mathbf{Y}$ on $\mathbf{X}$ to obtain the OLS estimate $\hat{\mathbf{A}}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and make the residual matrix $\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{A}}'$.

Step 2. Obtain the SOFAR estimate of $(\mathbf{B}, \mathbf{F})$ by

$$(\hat{\mathbf{B}}, \hat{\mathbf{F}}) = \operatorname*{arg\,min}_{(\tilde{\mathbf{B}}, \tilde{\mathbf{F}}) \in \mathbb{R}^{N \times \hat{K}} \times \mathbb{R}^{T \times \hat{K}}} \frac{1}{2} \|\hat{\mathbf{U}} - \tilde{\mathbf{F}}\tilde{\mathbf{B}}'\|_{\mathrm{F}}^2 + \eta_{NT} \|\tilde{\mathbf{B}}\|_1 \tag{3.6}$$
$$\text{subject to} \quad \tilde{\mathbf{F}}'\tilde{\mathbf{F}}/T = \mathbf{I}_{\hat{K}} \quad \text{and} \quad \tilde{\mathbf{B}}'\tilde{\mathbf{B}} \text{ diagonal},$$

where $\hat{K}$ is the estimated number of factors defined in Section 3.1 and $\eta_{NT} > 0$ is a penalty coefficient, and make the residual matrix $\hat{\mathbf{E}} = \hat{\mathbf{U}} - \hat{\mathbf{F}}\hat{\mathbf{B}}'$.

Step 3. Obtain the POET estimate of $\boldsymbol{\Sigma}_e$ by

$$\hat{\boldsymbol{\Sigma}}_e^{\tau} = (\hat{\sigma}_{ij}^{\tau})_{N\times N} \quad \text{with} \quad \hat{\sigma}_{ij}^{\tau} = \begin{cases} \hat{\sigma}_{ii}^{e} & \text{for } i = j, \\ s_{ij}^{\tau}(\hat{\sigma}_{ij}^{e}) & \text{for } i \neq j. \end{cases} \tag{3.7}$$

where $\hat{\sigma}_{ij}^{e}$ is the (i,j)-th entry of the sample covariance matrix of $\hat{\mathbf{e}}_t$, $s_{ij}^{\tau}(\cdot)$ is a soft thresholding function given by

$$s_{ij}^{\tau}(\hat{\sigma}_{ij}^{e}) = \text{sign}(\hat{\sigma}_{ij}^{e}) \max(0, |\hat{\sigma}_{ij}^{e}| - \tau_{ij})$$

with

$$\tau_{ij} = C_{\tau}\omega_{NT}\sqrt{\hat{\theta}_{ij}} \quad \text{and} \quad \hat{\theta}_{ij} = \frac{1}{T}\sum_{t=1}^{T}(\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}_{ij}^{e})^2 \tag{3.8}$$

for some sequence $\omega_{NT} > 0$ and sufficiently large constant $C_{\tau} > 0$.

Step 4. The covariance matrix estimate of $\mathbf{y}_t$ is defined as

$$\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{A}}\hat{\boldsymbol{\Sigma}}_x\hat{\mathbf{A}}' + \hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\boldsymbol{\Sigma}}_e^{\tau}, \tag{3.9}$$

where $\hat{\boldsymbol{\Sigma}}_x$ is the sample covariance matrix of the observable factors, $\mathbf{x}_t$.

It is reasonable to use the OLS in Step 1 and employ sample covariance matrix in Step 4 since $\mathbf{x}_t$ is observable and low-dimensional; see Fan et al. (2008) for a similar procedure. In Step 2, the SOFAR proposed by Uematsu et al. (2019) and Uematsu and Yamagata (2022) yields a sparse estimate of $\mathbf{B}$ with simultaneously satisfying the restrictions, $\hat{\mathbf{F}}'\hat{\mathbf{F}}/T = \mathbf{I}_{\hat{K}}$ and $\hat{\mathbf{B}}'\hat{\mathbf{B}}$ diagonal, which leads to an efficient estimation of the sWF model. The SOFAR estimation with tuning the regularisation coefficinent η_{NT} is numerically implemented by the R package, `rrpack`[2]. The SOFAR is regarded as a natural extension of the PC since they are identical if $\eta_{NT} = 0$. In Step 3, we adopt the soft threshold referring to our preliminary numerical trials though there are other types of threshold functions, such as the hard threshold and smoothly clipped absolute deviation (SCAD) method. Choosing the regularization coefficient C_{τ} is explained in Section B.

[2]See `https://cran.r-project.org/web/packages/rrpack/index.html`.

3.1 Determining the number of factors

There are several methods to determine the number of factors in approximate factor models, including Bai and Ng (2002) and Ahn and Horenstein (2013), but they are designed for the strong factor models with all the K signal eigenvalues diverging proportionally to N. In this paper, we recommend to use the approach of Onatski (2010). Briefly, it is to determine the number of (weak) factors by $\hat{K} = \hat{K}(\delta)$ with

$$\hat{K}(\delta) = \{k = 1, \dots, k_{\max} - 1 : \ \lambda_k - \lambda_{k+1} \geqslant \delta\},$$

where λ_k represents the kth largest eigenvalue of $(N \vee T)^{-1}\hat{\mathbf{U}}\hat{\mathbf{U}}'$ and $\delta > 0$ is a fixed constant. Uematsu and Yamagata (2022) prove that $\hat{K} \to K$ with high probability for any fixed $\delta > 0$ for a wide class of the sWF models. In practice, δ is predetermined by the *edge distribution* (ED) method of Onatski (2010) that is based on a calibration; see the article for full details.

4 Theoretical Properties

This section derives the rates of convergence of the estimated covariance and corresponding precision matrices as $N \wedge T$ tends to infinity while K and r are fixed.

Assumption 4.1 *Each k-th column vector of $\mathbf{B}$ has the sparsity $N_k = \lfloor N^{\alpha_k} \rfloor$ with $0 < \alpha_K \leqslant \cdots \leqslant \alpha_1 \leqslant 1$, and $\|\mathbf{B}\|_{\max} \leqslant c_b < \infty$ for some constant c_b. There exist constants $d_1, \dots, d_K$ such that $\mathbf{B}'\mathbf{B} = \operatorname{diag}(d_1^2 N_1, \dots, d_K^2 N_K)$ and $0 < d_K N_K^{1/2} \leqslant \cdots \leqslant d_1 N_1^{1/2}$. For k such that $\alpha_k = \alpha_{k-1}$, we have $d_{k-1}^2 - d_k^2 \leqslant c^{1/2} d_{k-1}^2$ for some constant $c > 0$.*

Assumption 4.2 *The vector process of idiosyncratic errors and latent factors $\{\mathbf{e}_t, \mathbf{f}_t\}_{t \geqslant 1}$ is strictly stationary with $E(\mathbf{e}_t) = \mathbf{0}$, $E(\mathbf{f}_t) = \mathbf{0}$, $E(e_{ti} f_{tk}) = 0$ and $\theta_{ij} := \operatorname{Var}(e_{ti} e_{tj}) > g$ for all $k = 1, \dots, K$ and $i, j = 1, \dots, N$ for some constant $g > 0$. Moreover, there exist some constants $c_1, c_2 > 0$ such that*

(a) $\max_{1\leqslant i\leqslant N} E[\exp(se_{ti}^2)] \leqslant c_1$ *for some* $0 < s < \infty$,

(b) $\max_{1\leqslant k\leqslant K} E[\exp(sf_{tk}^2)] \leqslant c_2$ *for some* $0 < s < \infty$.

For the idiosyncratic covariance, there exist some constants $\underline{c}, \bar{c} > 0$ *such that* $\underline{c} \leqslant \lambda_{\min}(\boldsymbol{\Sigma}_e) \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_e) \leqslant \bar{c}$. *Moreover,* $\boldsymbol{\Sigma}_e \in \Upsilon(m_N)$, *where*

$$\Upsilon(m_N) = \left\{ \boldsymbol{\Sigma}_e = (\sigma_{ij}^e)_{N\times N} \succ 0 : \max_{i\leqslant N} \sum_{j=1}^N |\sigma_{ij}^e|^q (\sigma_{ii}^e \sigma_{jj}^e)^{(1-q)/2} \leqslant m_N \right\}$$

for the sparsity measure $m_N = \max_{i\leqslant N} \sum_{j\leqslant N} |\sigma_{ij}^e|^q$ *with some constant* $q \in [0, 1]$.

Assumption 4.3 *There exist some constants* $r_1, b_1 > 0$ *such that the* α*-mixing coefficient*

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

satisfies $\alpha(T) \leqslant b_1 \exp(-CT^{r_1})$, *where* $\mathcal{F}_{-\infty}^0$ *and* $\mathcal{F}_T^\infty$ *represent the* σ*-fields generated by* $\{(\mathbf{f}_t, \mathbf{e}_t) : t \leqslant 0\}$ *and* $\{(\mathbf{f}_t, \mathbf{e}_t) : t \geqslant T\}$, *respectively.*

Assumption 4.4

(a) *There exists a constant* $c_e > 0$ *such that* $\|\mathbf{E}\|_2 \leqslant c_e(N \vee T)^{1/2}$ *holds with probability at least* $1 - O((N \vee T)^{-v})$ *for some* $v > 0$.

(b) *There exists a constant* c_4 *such that for all* t, k, *and* $x \geqslant 0$,

$$P\left(\left| \sum_{i=1}^N e_{ti} b_{ik} \right| \geqslant x \right) \leqslant 2\exp\left\{ -c_4 \frac{x^2}{N_1} \right\}.$$

Assumption 4.1 is the same as Assumption 2 in Uematsu and Yamagata (2022), which characterizes the sWF models discussed in Section 2.1. Assumption 4.2(a)(b) prescribe that f_{tk} and e_{ti} are subGaussian random variables. In Assumption 4.2, following Bickel and Levina (2008a) and Cai and Liu (2011), we assume the *conditional sparsity* condition, $\boldsymbol{\Sigma}_e \in \Upsilon(m_N)$. A smaller m_N indicates a sparser $\boldsymbol{\Sigma}_e$, which leads to faster convergence rates in the theorems. This assumption is widely used in, for instance, Fan et al. (2013), Cao et al. (2019) and Chen et al. (2019). Unlike Uematsu and Yamagata (2022), where they assume that idiosyncratic terms and factors follow the vector moving average (VMA) processes, we adopt the α-strong mixing condition in Assumption 4.3. Assumption 4.4 are relatively high-level but permissible conditions for deriving the consistency of factor and factor loadings, and such conditions are widely imposed in many existing literature. Bai and Ng (2006) basically assumes the same bound for $\mathbf{E}$ with weak serial and cross-sectional correlations. Moon and Weidner (2015, 2017) discuss several examples that various error processes may satisfy $\|\mathbf{E}\|_2 = O_p((N \vee T)^{1/2})$. Su and Wang (2017) also keeps the same order of the largest eigenvalue $\mathbf{E}$. In the same literature, they further assume that $\|\mathbf{EB}\|_{\max} = O_p(N^{1/2}\log^{1/2}(N \vee T))$ under a strong factor scheme, which is similar to and can be induced by our Assumption 4.4 (b). Similar high-level assumptions are also imposed in Fan et al. (2013) and Wang et al. (2021), for example. Moreover, in Uematsu and Yamagata (2022), they rigorously prove that all the inequalities in Assumption 4.4 hold when idiosyncratic errors and factors are specified as a VMA.

Set tuning parameters to be

$$\eta_{NT} \asymp T^{1/2}\log^{1/2}(N \vee T), \quad \omega_{NT} \asymp \frac{N_1^{3/2}\log^{1/2}(N \vee T)}{N_K(N_K \wedge T)}.$$

This choice of the tuning parameters will guarantee that $\|\mathbf{F}'\mathbf{E}\|_{\max} \lesssim \eta_{NT}$, $T^{-1/2}\|\hat{\mathbf{F}} - \mathbf{F}\|_{\mathrm{F}} \lesssim \omega_{NT}$, and $T^{-1/2}\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}} \lesssim N_1^{-1/2}\omega_{NT}$ occur with high probability (Lemmas A.7, A.9, and A.10 resp. in the Appendix).

Due to a technical reason, we restrict the parameter space of optimization (3.6) to

$$\left\{(\tilde{\mathbf{B}}, \tilde{\mathbf{F}}) \in \mathbb{R}^{N \times K} \times \mathbb{R}^{T \times K} : \max_i \|\tilde{\mathbf{b}}_i\| \leqslant C_b, \ \max_t \|\tilde{\mathbf{f}}_t\| \leqslant C_f\sqrt{\log T}\right\} \tag{4.10}$$

for some (sufficiently large) constants $C_b, C_f > 0$. This is large enough to include the true parameter $(\mathbf{B}, \mathbf{F})$ that satisfies Assumptions 4.1 and 4.2(b). Actually under Assumption 4.2(b) (sub-Gaussianity), we have $P(\max_t \|\mathbf{f}_t\| \leqslant C_f\sqrt{\log T}) \to 1$ for sufficiently large C_f by the sub-Gaussian property (Vershynin, 2018, Ch.2).

We write $T \asymp N^{\zeta}$ for some constant $\zeta > 0$ to represent the size of T relative to the cross-sectional dimension N. We further impose the technical condition that restricts the class of the sWF models:

$$\frac{N_1(N \vee T)^{1/2}}{N_K^{3/2} T^{1/2}} = o(1), \tag{4.11}$$

which is equivalent to $\alpha_1 + (1 \vee \zeta)/2 < 3\alpha_K/2 + \zeta/2$. The condition excludes the sWF models with a large gap between α_1 and α_K and/or too small T relative to N. This is used to derive the estimation error bounds of the SOFAR estimator. Moreover, with this condition, we can prove that $\hat{K}$ in Section 3.1 converges to the true number K with high probability; see Uematsu and Yamagata (2022). Thus we suppose K is known throughout the rest of this section.

4.1 Case 1: Observable factors do not exist

We first consider the case without observable factors in the model. This corresponds to the original POET setting, but allows the weak factors. The ePOET estimator of $\mathbf{y}_t$ is simply reduced to

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\mathbf{\Sigma}}_e^{\tau}. \tag{4.12}$$

Theorem 4.1 *Suppose that Assumptions 4.1–4.4 and condition (4.11) hold. Then $\hat{\mathbf{\Sigma}}_e^{\tau}$ in (4.12) satisfies with probability at least $1 - O((N \vee T)^{-v})$,*

$$\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\| \lesssim \omega_{NT}^{1-q} m_N.$$

Theorem 4.2 *Suppose the same assumptions as in Theorem 4.1 hold. If*

$$\omega_{NT}^{1-q} m_N = o(1), \tag{4.13}$$

then the inverse matrix of $\hat{\mathbf{\Sigma}}_e^{\tau}$ is well-defined with probability approaching one, and it satisfies with probability at least $1 - O((N \vee T)^{-v})$,

$$\|\hat{\mathbf{\Sigma}}_e^{\tau -1} - \mathbf{\Sigma}_e^{-1}\| \lesssim \omega_{NT}^{1-q} m_N.$$

Theorem 4.3 *Suppose the same assumptions as in Theorem 4.1 hold. Then* $\hat{\mathbf{\Sigma}}$ *in* (4.12) *satisfies that*

$$\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathbf{\Sigma}} \lesssim \frac{N^{1/2}\log(N \vee T)}{T} + \omega_{NT}^{1-q} m_N,$$
$$\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} \lesssim \omega_{NT}$$

hold with probability at least $1 - O((N \vee T)^{-v})$.

Following Fan et al. (2011) and Fan et al. (2013), we have used the relative error $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathbf{\Sigma}} = \|\mathbf{\Sigma}^{-1/2}\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_N\|_{\mathrm{F}}$ because the spectral norm is too large to be controlled for $\mathbf{\Sigma}$.

Theorem 4.4 *Suppose the same assumptions as in Theorem 4.1 hold. If*

$$m_N \omega_{NT}^{1-q} \frac{N_1^{1/2}(N_1 \vee T)^{1/2}}{N_K} = o(1), \tag{4.14}$$

then the inverse matrix of $\hat{\mathbf{\Sigma}}$ *is well-defined with probability approaching one, and it satisfies with probability at least* $1 - O((N \vee T)^{-v})$,

$$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\| \lesssim m_N \omega_{NT}^{1-q} \frac{N_1^{3/2}(N_1 \vee T)^{1/2}}{N_K^2}.$$

To prove Theorem 4.1-4.4, a key step is to derive the consistency of factors and factor loadings, which will be stated and proved in Lemma A.9 and A.10. In particular, the rates of estimated factor and factor loadings by SOFAR can recover the original POET ones if all the factors are strong (i.e., $N_k = N$ for all $k \leqslant K$). On the other hand, the rates deteriorate with an extra cost if PC is used under the sWF models. See Appendix C for more details. Therefore, our covariance estimate has an advantage of faster convergence speed under the sWF, and can work as good as POET estimates even under the SF.

4.2 Case 2: Observable factors exist

When the model contains the observable factors $\mathbf{x}_t$, we need to estimate $\mathbf{A}$ first by OLS as explained in Section 3. To evaluate the estimation error, additional conditions for the observable factors $\mathbf{x}_t$ are required.

Assumption 4.5

(a) The vector of the observed and latent factors and the idiosyncratic error $\{\mathbf{x}_t, \mathbf{f}_t, \mathbf{e}_t\}_{t\geqslant 1}$ *is strictly stationary with* $E(x_{tl}e_{ti}) = 0$ *and* $E(x_{tl}f_{tk}) = 0$ *for any* $l = 1, \ldots, r$, $k = 1, \ldots, K$, *and* $i = 1, \ldots, N$.

(b) There exits some positive constant c_3 *such that* $\max_{1\leqslant l\leqslant r} E[\exp(sx_{tl}^2)] \leqslant c_3$ *for some* $0 < s < \infty$,

(c) There exist some positive constants r_2 *and* b_2 *such that the* α*-mixing coefficient*

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

satisfies $\alpha(T) \leqslant b_2\exp(-CT^{r_2})$, *where* $\mathcal{F}_{-\infty}^0$ *and* $\mathcal{F}_T^\infty$ *represent the* σ*-fields generated by* $\{(\mathbf{x}_t, \mathbf{f}_t, \mathbf{e}_t) : t \leqslant 0\}$ *and* $\{(\mathbf{x}_t, \mathbf{f}_t, \mathbf{e}_t) : t \geqslant T\}$, *respectively.*

(d) $c_u \geqslant \lambda_{\max}(\mathrm{cov}(\mathbf{x}_t)) \geqslant \cdots \geqslant \lambda_{\min}(\mathrm{cov}(\mathbf{x}_t)) \geqslant c_l$ *for some constants* $c_u, c_l > 0$.

(e) The number of observable factors r *is fixed and known.*

(f) $\lambda_{\min}(N^{-1}\mathbf{A}'\mathbf{A}) \geqslant c_a$ *for some positive constant* c_a.

In Assumption 4.5, conditions (a) – (e) are standard in the literature. Condition (f) implies that the r observed factors are *strongly* pervasive.

Lemma 4.1 *Suppose that Assumptions 4.1–4.2, 4.4–4.5 hold. Then we have*

$$\|\hat{\mathbf{U}} - \mathbf{U}\|_{\max} \lesssim \frac{\log^{1/2}(N\vee T)\log^{1/2}T}{T^{1/2}}$$

occurs with probability at least $1 - O((N\vee T)^{-v})$.

Thanks to this lemma, we can find that the estimation error in Step 1 is small enough to treat $\mathbf{A}$ as if it were known. Consequently, the convergence results of covariance estimators in *Case* 2, which are similar to those in *Case* 1, are obtained. Define

$$\tilde{\omega}_{NT} = \omega_{NT} \ \vee \ \frac{\log^{1/2}(N \vee T) \log^{1/2} T}{T^{1/2}}.$$

Theorem 4.5 *Suppose that Assumptions 4.1–4.2, 4.4–4.5 and condition (4.11) hold. Then all the assertions of Theorems 4.1–4.3 in Case 1 with ω_{NT} replaced by $\tilde{\omega}_{NT}$ are true.*

Theorem 4.6 *Suppose that the assumptions as in Theorem 4.5 hold. If*

$$\frac{N \vee (N_1^{1/2} T^{1/2})}{N} m_N \tilde{\omega}_{NT}^{1-q} = o(1), \tag{4.15}$$

then the inverse matrix of $\hat{\mathbf{\Sigma}}$ is well-defined with probability approaching one, and it satisfies

$$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\| \lesssim m_N \tilde{\omega}_{NT}^{1-q}$$

hold with probability at least $1 - O((N \vee T)^{-v})$.

In *Case 1*, the convergence rates of covariance estimators are mainly determined by the estimation error of latent factors and loadings, ω_{NT}. However in *Case 2*, the estimation error comes from the first stage OLS and ω_{NT}, implying that the convergence rates of ePOET depend on the maximum value of them, $\tilde{\omega}_{NT}$. Theorem 4.6, which is parallel to Theorem 4.4, holds under a weaker condition (4.15) than (4.14) and gives the same rate as $\hat{\mathbf{\Sigma}}_e^{-1}$, which is faster than that in Theorem 4.4. It implies that the precision estimate of ePOET can work better when observable factors exist than the case when observable factors do not exist.

5 Simulation Studies

We carry out Monte Carlo simulations to investigate the finite-sample performances of the ePOET estimators. We consider a similar data generating process (DGP) in Uematsu and

Yamagata (2022) with some modifications on the idiosyncratic error term:

$$y_{ti} = \sum_{l=1}^{r} a_{il} x_{tl} + \sum_{k=1}^{K} b_{ik} f_{tk} + \sqrt{\theta} e_{ti}.$$

The observed factors x_{tl} and the factor loadings a_{il} are both generated from i.i.d.$N(0,1)$. The unobserved factors f_{tk} and the corresponding factor loadings b_{ik} satisfy the conditions $T^{-1}\sum_{t=1}^{T} f_{tk} f_{ts} = 1\{s = k\}$ and $N^{-1}\sum_{i=1}^{N} b_{ik} b_{is} = 1\{s = k\}$, which are realized by the Gram-Schimidt orthogonalization to f_{tk}^* and b_{ik}^*. Here, $b_{ik}^* \sim$ i.i.d.$N(0,1)$ for $i = 1, \ldots, N_k$ and $b_{ik}^* = 0$ for $N_{k+1}, \ldots, N$, and $f_{tk}^* = \rho_{fk} f_{t-1,k}^* + v_{tk}$ with $|\rho_{fk}| < 1$, $f_{t-1,k}^* \sim$ i.i.d.$N(0,1)$, and $v_{tk} \sim$ i.i.d.$N(0, 1-\rho_{fk}^2)$. For the idiosyncratic error term, $\mathbf{e}_t = (e_{ti})_{N\times 1}$, we simulate them independently by $\mathbf{e}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e)$, where $\boldsymbol{\Sigma}_e = (\sigma_{ij}^e)_{N\times N}$ with $\sigma_{ij}^e = \sigma^{|i-j|} 1\{|i-j| \leqslant 4\}$.

The following statistical losses are used to evaluate the estimation accuracy.

(a) Relative error: $N^{1/2}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\Sigma} = \|\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_N\|_{\mathrm{F}}$.

(b) Spectral loss of idisosyncratic covariance estimator: $\|\hat{\boldsymbol{\Sigma}}_e^{\mathcal{T}} - \boldsymbol{\Sigma}_e\|$.

(c) Kullback-Leibler Divergence: $\mathbf{KLD} := \mathrm{Trace}\left(\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\right) - \log\left(\left|\boldsymbol{\Sigma}\hat{\boldsymbol{\Sigma}}^{-1}\right|\right) - N$.

(d) Spectral loss of inverse of covariance estimator: $\|\hat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\|$.

(e) Spectral loss of inverse of idiosyncratic covariance estimator: $\|\hat{\boldsymbol{\Sigma}}_e^{\mathcal{T}^{-1}} - \boldsymbol{\Sigma}_e^{-1}\|$.

(f) Frobenius loss of factor loadings: $\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}$.

Criterion (a) is a relative error measure for covariance matrix estimation, which is proposed by Fan et al. (2011) and Fan et al. (2013). For (c), it has been used in Yuan and Lin (2007) and Rothman et al. (2008), for example, to evaluate the precision matrix estimation performance. The others are common criteria.

5.1 Case 1

For the case with no observable factors, we compare estimates of the ePOET and POET. The only difference is that the former applies the SOFAR to obtain $(\hat{\mathbf{B}}, \hat{\mathbf{F}})$ while the latter uses the PCA. Two simulation studies are presented; one assumes two strong factors and the other considers two weak factors. For each, we set $\rho_{fk} = 0.5$, $\theta = 1$ and $\sigma = 0.3$, and fix $T = 200$.

Table 1 reports the accuracy of the ePOET and POET estimates when the exponents $(\alpha_1, \alpha_2) = (1, 1)$, representing the strong factors. In this case, the penalty term for $\mathbf{B}$ is indeed unnecessary, but the ePOET performs as good as the POET in all aspects. Interestingly, the ePOET even slightly outperforms the POET in some cases. This suggests the use of ePOET is good even when the latent factors are expected to be strong. Table 2 shows the case with $(\alpha_1, \alpha_2) = (0.6, 0.6)$, which indicates the weak factors. Overall the ePOET performs better especially in terms of the KLD, and the tendency becomes strong as N increases though the POET can work as good as the ePOET only if $T > N$. This feature results from the advantage of the SOFAR estimates over the PC estimates for the sparse factor loadings; see Uematsu and Yamagata (2022) for more information.

5.2 Case 2

For the case when observable factors exist, we also set $\rho_{fk} = 0.5$, $\theta = 1$ and $\sigma = 0.3$, and fix $T = 200$. The numbers of observable factors and unobservable factors are fixed to be $r = 3$ and $K = 3$, respectively. We compare the three estimates: (i) *ePOET* proposed in Section 3; (ii) *POET* constructed by the PCA instead of the SOFAR; (iii) *Diagonalized Sample Covariance* obtained as follows. This construction is inspired by Fan et al. (2008).

Step 1. Regress $\mathbf{Y}$ on $\mathbf{X}$ to obtain the OLS estimate $\hat{\mathbf{A}}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and make the residual matrix $\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{A}}'$.

Table 1: Performance of the ePOET and POET estimates for the factor model with (α_1, α_2) = (1,1) with no observable factors. $T = 200$.

		ePOET			**POET**		
Design (α_1, α_2)		$(\alpha_1, \alpha_2) = (1,\ 1)$					
N	**Criteria**	**mean**	**median**	**s.d.**	**mean**	**median**	**s.d.**
100	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$	3.80	3.80	0.11	3.81	3.80	0.10
	$\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|$	0.74	0.73	0.04	0.74	0.74	0.04
	KLD	4.78	4.80	0.27	4.74	4.75	0.26
	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.64	0.64	0.03	0.64	0.64	0.03
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	0.65	0.64	0.04	0.65	0.64	0.03
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}$	18.85	19.64	6.03	18.27	19.59	6.63
200	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$	5.02	5.02	0.11	5.08	5.12	0.15
	$\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|$	0.76	0.75	0.03	0.76	0.76	0.03
	KLD	10.36	10.39	0.39	10.28	10.25	0.39
	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.69	0.69	0.03	0.69	0.69	0.03
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	0.70	0.69	0.03	0.70	0.69	0.03
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}$	26.61	27.95	9.11	26.07	28.30	10.06
300	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\Sigma}$	5.99	5.99	0.11	6.00	5.98	0.13
	$\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|$	0.78	0.78	0.03	0.78	0.78	0.03
	KLD	16.50	16.47	0.60	16.37	16.33	0.56
	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.72	0.72	0.02	0.72	0.72	0.02
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	0.73	0.72	0.02	0.72	0.72	0.02
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}$	32.62	34.26	10.67	33.74	34.50	9.56

Table 2: Performance of the ePOET and POET estimates for the factor model with (α_1, α_2) = (0.6,0.6) with no observable factors. $T = 200$.

		ePOET			**POET**		
	Design (α_1, α_2)	(α_1, α_2) = (0.6, 0.6)					
N	**Criteria**	**mean**	**median**	**s.d.**	**mean**	**median**	**s.d.**
	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\Sigma$	3.65	3.66	0.16	3.78	3.78	0.09
	$\|\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\|$	0.94	0.95	0.11	1.01	1.01	0.12
	KLD	4.11	4.00	0.29	4.72	4.74	0.27
100	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.66	0.66	0.03	0.66	0.65	0.04
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	1.62	1.43	0.77	1.93	1.79	0.77
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_F$	7.14	7.52	2.39	7.26	7.90	3.71
	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\Sigma$	4.93	4.90	0.18	5.10	5.15	0.14
	$\|\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\|$	0.85	0.84	0.07	0.90	0.90	0.08
	KLD	9.10	9.07	0.46	10.32	10.31	0.40
200	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.72	0.71	0.03	0.70	0.70	0.03
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	0.92	0.82	0.25	1.06	1.02	0.27
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_F$	8.89	9.12	2.74	9.36	9.80	2.96
	$N^{1/2}\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_\Sigma$	5.90	5.87	0.20	6.02	6.01	0.12
	$\|\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\|$	0.82	0.80	0.05	0.85	0.84	0.06
	KLD	14.48	14.44	0.62	16.41	16.40	0.58
300	$\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\|$	0.74	0.74	0.02	0.72	0.72	0.02
	$\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\|$	0.79	0.76	0.09	0.80	0.75	0.12
	$\|\hat{\mathbf{B}} - \mathbf{B}\|_F$	9.83	10.16	3.04	10.63	11.20	3.82

Step 2. Obtain the sample covariance matrix of $\hat{\mathbf{U}}$, and replace all the off-diagonal elements with zeros, which is denoted by $\hat{\mathbf{\Sigma}}_u^{\text{diag}}$.

Step 3. The covariance matrix estimate of $\mathbf{Y}$ is given by

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{A}}\hat{\mathbf{\Sigma}}_x\hat{\mathbf{A}}' + \hat{\mathbf{\Sigma}}_u^{\text{diag}},$$

where $\hat{\mathbf{\Sigma}}_x$ is the sample covariance matrix of the observable factors, $\mathbf{x}_t$.

In Table 3, our ePOET (i) works better than the other two methods (ii) and (iii) in most of the aspects except for the precision matrix estimates. In particular, it is clear that the superiority of the ePOET becomes larger as N increases. The results also demonstrate that after taking out observable factors, the ePOET works better than that of Case 1.

6 Empirical Applications

We carry out two empirical studies of proposed ePOET estimators.

6.1 Risk minimization of portfolios

In order to evaluate the performance of our ePOET in practice, we consider to construct minimum variance portfolios (MVP) using daily access return series of the S&P 500 dataset and Fama-French five factors data from Kenneth R. French-Data Library. [3]

6.1.1 MVP model and empirical designs.

The MVP attempts to allocate N financial assets to make the portfolio risk $\mathbf{w}'\tilde{\mathbf{\Sigma}}\mathbf{w}$, where $\mathbf{w}$ is a vector of weights and $\tilde{\mathbf{\Sigma}}$ is a covariance matrix estimate of the given assets, as small as possible. Specifically, the MVP solves the optimization problem:

$$\min_{\mathbf{w}} \mathbf{w}'\tilde{\mathbf{\Sigma}}\mathbf{w} \quad \text{subject to} \quad \mathbf{w}'\mathbf{1}_N = 1, \tag{6.16}$$

[3] See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 3: Performance of the three methods, where (i), (ii), and (iii) refere to the ePOET, POET, and Diagonalized Sample Covariance, respectively, for the factor model with three observable factors and three unobservable factors.

		Mean			**Median**			**s.d**		
	Design $(\alpha_1, \alpha_2, \alpha_3)$	$(\alpha_1, \alpha_2, \alpha_3) = (0.7, 0.7, 0.7)$								
N	**Criteria**	**(i)**	**(ii)**	**(iii)**	**(i)**	**(ii)**	**(iii)**	**(i)**	**(ii)**	**(iii)**
100	$N^{1/2}\Vert\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\Vert_\Sigma$	4.15	4.32	4.42	4.15	4.28	4.42	0.11	0.11	0.13
	$\Vert\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\Vert$	0.92	1.00	-	0.92	1.00	-	0.09	0.09	-
	KLD	6.06	6.83	30.58	6.03	6.82	30.63	0.36	0.37	2.51
	$\Vert\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\Vert$	0.64	0.69	1.53	0.64	0.68	1.53	0.04	0.07	0.05
	$\Vert\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\Vert$	1.21	1.65	-	1.13	1.56	-	0.39	0.46	-
	$\Vert\hat{\mathbf{B}} - \mathbf{B}\Vert_\mathrm{F}$	11.22	11.91	-	11.52	11.98	-	2.09	2.34	-
200	$N^{1/2}\Vert\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\Vert_\Sigma$	5.36	5.60	5.79	5.34	5.62	5.78	0.15	0.10	0.15
	$\Vert\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\Vert$	0.83	0.89	-	0.82	0.89	-	0.05	0.06	-
	KLD	13.03	14.59	60.20	13.03	14.59	60.30	0.58	0.58	3.27
	$\Vert\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\Vert$	0.71	0.68	1.57	0.70	0.67	1.57	0.03	0.04	0.03
	$\Vert\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\Vert$	0.77	0.88	-	0.74	0.86	-	0.09	0.15	-
	$\Vert\hat{\mathbf{B}} - \mathbf{B}\Vert_\mathrm{F}$	14.61	15.69	-	14.70	15.85	-	2.62	2.59	-
300	$N^{1/2}\Vert\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\Vert_\Sigma$	6.29	6.54	6.82	6.26	6.55	6.81	0.16	0.13	0.17
	$\Vert\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\Vert$	0.81	0.86	-	0.81	0.86	-	0.04	0.04	-
	KLD	20.39	22.82	86.86	20.37	22.88	86.63	0.68	0.75	3.95
	$\Vert\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\Vert$	0.74	0.71	1.59	0.74	0.70	1.59	0.03	0.03	0.03
	$\Vert\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\Vert$	0.76	0.74	-	0.76	0.72	-	0.03	0.07	-
	$\Vert\hat{\mathbf{B}} - \mathbf{B}\Vert_\mathrm{F}$	16.61	17.55	-	16.82	17.80	-	3.27	3.13	-

where $\mathbf{1}_N = (1, \ldots, 1)'$. We allow short sales and ignore any transaction cost for simplicity. It is well-known that the optimal weight $\mathbf{w}^*$ obtained by the quadratic problem (6.16) and the corresponding risk $\mathbf{R}^*$ are computed as

$$\mathbf{w}^* = \frac{\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}{\mathbf{1}_N{}'\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}_N}, \qquad \mathbf{R}^* = \mathbf{w}^{*\prime}\tilde{\boldsymbol{\Sigma}}\mathbf{w}^*. \tag{6.17}$$

We compare the out-of-sample forecasting performance of candidate methods in terms of the MVP construction. The following five methods are conducted.

(i) **ePOET-2**: ePOET proposed in Section 2 when observable factors exist.

(ii) **POET**: the method proposed in Fan et al. (2013), which does not require observable factors.

(iii) **EFM-POET**: POET constructed by the PCA instead of the SOFAR when observable factors exist.

(iv) **ePOET-1**: ePOET proposed in Section 2 when no observable factors.

(v) **EFM**: Exact factor model proposed by Fan et al. (2008), which use observable factors and the diagonal part of sample idiosyncratic covariance.

We collect the S&P 500 data that consists of 2520 daily excess return for the period from April 2, 2002 to March 30, 2012 (about 10-year trading days with 21 days per month). For the methods used observable factors, we also collect the Fama-French five factors data from the same period as the S&P 500 returns. A portfolio is created at the first trading day of each month using a candidate method to estimate the covariance matrix of returns based on the data from the past T days. To reflect the high-dimensionality, we set the time dimension $T = 126$ (six months of trading days) and the cross-sectional dimension $N = 395$, which is the maximum number of stocks available in the dataset.

Under a rolling window scheme, the vector of optimal portfolio weights ($\mathbf{w}_t^*$) is updated monthly (21 days) for constructing next month's portfolios until March 30, 2012. Once

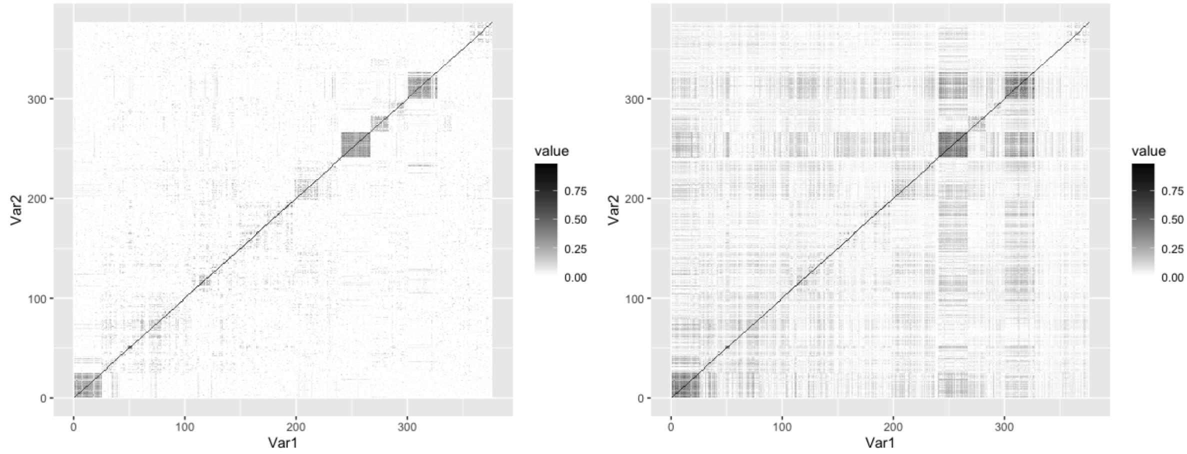
Table 4: Performance of different methods in out-of-sample minimum variance portfolio analysis.

Criteria	**ePOET-2**	**EFM-POET**	**ePOET-1**	**POET**	**EFM**
Out-of-sample standard deviation	0.45	0.50	0.48	0.47	0.51
Total excess return	39.89	33.46	35.01	38.19	38.67
Mean Sharpe ratio	1.36	1.32	1.24	1.29	1.14

obtaining all the out-of-sample portfolios, we calculate out-of-sample standard deviation, the total out-of-sample excess returns and the mean Sharpe ratio following DeMiguel et al. (2009). We determine the number of latent factors by the method of Onatski (2010) and select the optimal threshold tuning parameter by CV for each update.

6.1.2 Results

Because the aim here is risk minimization, the out-of-sample standard deviation should be our primary basis of comparison, with the total out-of-sample returns and the Sharpe ratios serving as the secondary bases. In Table 4, we find that the ePOET with observable factors achieves the best performance in out-of-sample standard deviation among the candidates. Meanwhile it can maintain certain high level excess returns. When assuming there are no observable factors, ePOET works similar to POET under the criteria of out-of-sample standard deviation. Exact factor model works the worst as observable factors are not sufficient to capture the covariance structure. Overall, ePOET with observable factors performs the best in both risk minimization and out-of-sample return.



(a) ePOET (b) POET

Figure 3: Heatmaps of estimated residual correlations (absolute values) from the Fama-French five-factor model.

6.2 Unobserved weak factors left out in the residuals of Fama-French five-factor model

As the scree plots in Introduction demonstrate, weak factors may exist in the residuals of the Fama-French five-factor model. We use the same dataset in Section 1 to explore the covariance structure of the residuals of the five-factor model based on our sWF framework. The residuals are obtained by taking out all five observable factors of the five-factor model from the collected stock returns through the OLS method. From Figure 3, it is clear that the POET estimates exhibit much noisier patterns than the ePOET estimates. The ePOET can remove some of the estimation noise through its sparseness mechanism and successfully retain the clustered non-zeros among specific industries. Such findings are consistent with the conclusion of Section 6.3 of Uematsu and Yamagata (2022) and inspire us to estimate covariance matrices by the ePOET method when some factors are already known.

7 Conclusion

This paper proposes the extended POET (ePOET), which fully extends the original POET of Fan et al. (2013) to allowing the model to have not only latent strong factors but also observable and latent weak factors. Regarding estimation of covariance matrices, ePOET combines the observable factors, sparsity-induced weak factors (sWF), and the sparse idiosyncratic noise to estimate high-dimensional covariance matrices. Compared to POET, when observable factor exist, our ePOET method can better detect the potential WF structure in the residuals. When observable factors do not exist, the ePOET is able to distinguish which factors are essential to data variations through the sparsity patterns of the estimated factor loadings, thereby enhancing the explanatory power of the proposed model. Simulation studies show that if data are generated from relatively weak latent factors only or a mixed structure of observable factors and weak latent factors, the performance of our ePOET model is uniformly better than the POET estimators. In addition, ePOET can work as good as POET even if data contain strong factors only. The MVP studies conclude that the ePOET with mixed factors brings significantly less risky portfolios than other candidates, maintaining the highest returns. Moreover, as our model can be seen as a structure of observable regressors plus factor errors, which is similar to the structure proposed in Bai (2009) and Fan et al. (2021), there may have potential research interests for panel regressions such as GLS estimation.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J., K. Li, and L. Lu (2016). Estimation and inference of favar models. *Journal of Business & Economic Statistics 34*(4), 620–641.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Determining the number of factors in approximate factor models, errata. *manuscript, Columbia University*.

Bickel, P. J. and E. Levina (2008a). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Bickel, P. J. and E. Levina (2008b). Regularized estimation of large covariance matrices. *The Annals of Statistics 36*(1), 199–227.

Cai, T. and W. Liu (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association 106*(494), 672–684.

Cao, Y., W. Lin, and H. Li (2019). Large covariance estimation for compositional data via composition-adjusted thresholding. *Journal of the American Statistical Association 114*(526), 759–772.

Chen, J., D. Li, and O. Linton (2019). A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics 212*(1), 155–176.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science 55*(5), 798–812.

Ding, Y., Y. Li, and X. Zheng (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *Journal of Econometrics 222*(1), 502–515.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics 39*(6), 3320.

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

Fan, J., H. Liu, and W. Wang (2018). Large covariance estimation through elliptical factor models. *Annals of statistics 46*(4), 1383.

Fan, J., R. Masini, and M. C. Medeiros (2021). Bridging factor and sparse models. *Available at SSRN 3789141*.

Gagliardini, P., E. Ossola, and O. Scaillet (2019). A diagnostic criterion for approximate factor structure. *Journal of Econometrics 212*(2), 503–521.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *The Annals of Statistics 44*(3), 928–953.

Lam, C., Q. Yao, and N. Bathia (2011). Estimation of latent factors for high-dimensional time series. *Biometrika 98*(4), 901–918.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis 88*(2), 365–411.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Merlevède, F., M. Peligrad, and E. Rio (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, pp. 273–292. Institute of Mathematical Statistics.

Moon, H. R. and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica 83*(4), 1543–1579.

Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory 33*(1), 158–195.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Press, S. J. (1972). *Applied multivariate analysis.* Series in quantitative methods for decision-making. New York: Holt, Rinehart and Winston.

Rothman, A. J., P. J. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics 2*, 494–515.

Rothman, A. J., E. Levina, and J. Zhu (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association 104*(485), 177–186.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

Uematsu, Y., Y. Fan, K. Chen, J. Lv, and W. Lin (2019). SOFAR: large-scale association network learning. *IEEE transactions on information theory 65*(8), 4924–4939.

Uematsu, Y. and T. Yamagata (2021). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, 1–14.

Uematsu, Y. and T. Yamagata (2022). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, 1–15.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Wang, H., B. Peng, D. Li, and C. Leng (2021). Nonparametric estimation of large covariance matrices with conditional sparsity. *Journal of Econometrics 223*(1), 53–72.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the gaussian graphical model. *Biometrika 94*(1), 19–35.

Supplementary Material for

Estimation of Large Covariance Matrices with Mixed Factor Structure

Dai Runyu*, Yoshimasa Uematsu†* and Yasumasa Matsuda*

*Department of Economics and Management, Tohoku University

†Center for the Promotion of Social Data Science Education and Research, Hitotsubashi University

A Proofs of the Main Results and Lemmas

In this section, we present the lemmas and proofs for the main results for two cases in Section 4.1 and 4.2 separately.

A.1 Case 1: Observable factors do not exist

Lemma A.1 *If all the assumptions in Theorem 4.1 are satisfied, then*

$$T^{-1}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2 \lesssim \frac{N_1^3\log(N\vee T)}{N_K^2(N_K\wedge T)^2}=\omega_{NT}^2$$

holds with probability at least $1-O((N\vee T)^{-v})$.

Proof: Note that

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2=\frac{1}{T}\|\hat{\mathbf{F}}-\mathbf{F}\|_{\mathrm{F}}^2\lesssim\omega_{NT}^2,$$

where $\lesssim$ holds because of Lemma A.9. □

Lemma A.2 *Under the assumptions in Theorem 4.1, we have*

$$\max_{i\leqslant N}\|\hat{\mathbf{b}}_i-\mathbf{b}_i\|\lesssim\frac{\log^{1/2}(N\vee T)}{T^{1/2}}\leqslant\frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}$$

holds with probability at least $1-O((N\vee T)^{-v})$.

Proof: According to Uematsu and Yamagata (2021), we obtain $\hat{\mathbf{B}}$ under the Karush-Kuhn-Tucker (KKT) conitions by

$$\hat{\mathbf{B}}-\mathbf{B}=T^{-1}(\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}}-\mathbf{F})+\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F}))+T^{-1}\mathbf{E}'\mathbf{F}-T^{-1}\eta_n\mathbf{V}(\hat{\mathbf{B}}),$$

where the $(i,k)th$ element of $\mathbf{V}(\tilde{\mathbf{B}})$ for given $\tilde{\mathbf{B}}=\tilde{b}_{ik}\in\mathbb{R}^{N\times K}$ is defined as

$$v_{ik}(\tilde{\mathbf{B}})\begin{cases}=\operatorname{sgn}(\tilde{b}_{ik}) & \text{for } \tilde{b}_{ik}\neq 0\\ \in[-1,1] & \text{for } \tilde{b}_{ik}=0.\end{cases}$$

Then, using the triangle inequality, we have

$$\begin{aligned}&\max_{i\leqslant N}\|\hat{\mathbf{b}}_i-\mathbf{b}_i\|\\ \leqslant\;& T^{-1}\eta_n+T^{-1}(\|\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max}+\|\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max})+T^{-1}\|\mathbf{E}'\mathbf{F}\|_{\max}.\end{aligned}\tag{A.1}$$

Note that we set $\eta_n\asymp T^{1/2}\log^{1/2}(N\vee T)$ in Section 4. The proof of Theorem 1 in Uematsu and Yamagata (2021) implies the middle term of (A.1) satisfies

$$\begin{aligned}&T^{-1}(\|\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max}+\|\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max})\\ \lesssim\;& T^{-1/2}\frac{N_1^{3/2}\log(N\vee T)}{N_K(N_K\wedge T)}.\end{aligned}$$

Also note that Lemma A.7 gives $T^{-1}\|\mathbf{E}'\mathbf{F}\|_{\max}\lesssim T^{-1/2}\log^{1/2}(N\vee T)$. Combining the terms in (A.1), we have

$$\max_{i\leqslant N}\|\hat{\mathbf{b}}_i-\mathbf{b}_i\|\lesssim\frac{\log^{1/2}(N\vee T)}{T^{1/2}}\leqslant\frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}=\omega_{NT},\tag{A.2}$$

where $\leqslant$ holds because condition (4.11) ensures that

$$\frac{N_1^{3/2}T^{1/2}}{N_K(N_K\wedge T)}\geqslant\frac{N_1^{1/2}T^{1/2}}{(N_K\wedge T)}\geqslant 1.$$

□

Lemma A.3 *If all the assumptions in Theorem 4.1 are satisfied, the following inequalities hold with probability at least* $1 - O((N \vee T)^{-v})$:

(a) $\max_{i \leqslant N} \frac{1}{T} \sum_{t=1}^{T} |\hat{e}_{ti} - e_{ti}|^2 \lesssim \omega_{NT}^2$,

(b) $\max_{i \leqslant N, t \leqslant T} |\hat{e}_{ti} - e_{ti}|^2 \lesssim \log(N \vee T)$.

Proof: Let $\Delta_i^{\mathbf{b}} = \hat{\mathbf{b}}_i - \mathbf{b}_i$ and $\Delta_t^{\mathbf{f}} = \hat{\mathbf{f}}_t - \mathbf{f}_t$. We first rewrite $|\hat{e}_{ti} - e_{ti}|$ as

$$|\hat{e}_{ti} - e_{ti}| = |\Delta_t^{\mathbf{f}} \mathbf{b}_i' + \Delta_t^{\mathbf{f}} \Delta_i^{\mathbf{b}\prime} + \mathbf{f}_t \Delta_i^{\mathbf{b}\prime}|.$$

For (a), by the inequality $(A+B+C)^2 \leqslant 3(A^2+B^2+C^2)$, Lemma A.1 and Lemma A.2, we get

$$\begin{aligned} & \max_{i \leqslant N} \frac{1}{T} \sum_{t=1}^{T} |\hat{e}_{ti} - e_{ti}|^2 \\ \leqslant \quad & 3 \max_i \|\mathbf{b}_i\|^2 \frac{1}{T} \sum_{t=1}^{T} \|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 + 3 \max_i \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \frac{1}{T} \sum_{t=1}^{T} \|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 \\ + \quad & 3 \max_i \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \frac{1}{T} \sum_{t=1}^{T} \|\mathbf{f}_t\|^2 \end{aligned}$$

Note that with the proof of Lemma 3.1(ii) in Fan et al. (2011), $T^{-1} \sum_{t=1}^{T} \|\mathbf{f}_t\|^2 \lesssim K$. Therefore,

$$\max_{i \leqslant N} \frac{1}{T} \sum_{t=1}^{T} |\hat{e}_{ti} - e_{ti}|^2 \lesssim \omega_{NT}^2$$

holds with probability at least $1 - O((N \vee T)^{-v})$. Similarly for (b), we upper bound $\max_{i \leqslant N, t \leqslant T} |\hat{e}_{ti} - e_{ti}|^2$ as

$$\begin{aligned} & \max_{i \leqslant N, t \leqslant T} |\hat{e}_{ti} - e_{ti}|^2 \\ \leqslant \quad & 3 \max_i \|\mathbf{b}_i\|^2 \max_t \|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 + 3 \max_i \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \max_t \|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 \\ \leqslant \quad & 3 \max_i \|\mathbf{b}_i\|^2 (\max_t \|\hat{\mathbf{f}}_t\|^2 + \max_t \|\mathbf{f}_t\|^2) + 3 \max_i \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 (\max_t \|\hat{\mathbf{f}}_t\|^2 + \max_t \|\mathbf{f}_t\|^2) \\ + \quad & 3 \max_i \|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \max_t \|\mathbf{f}_t\|^2 \\ \lesssim \quad & \log(N \vee T), \end{aligned}$$

where we have used the fact that $\hat{\mathbf{f}}_t$ and $\mathbf{f}_t$ lie in the parameter space in (4.10) with high probability under Assumption 4.2(b). $\square$

Lemma A.4 *Under the assumptions of Theorem 4.1, there exists a large enough constant $C > 0$ such that*

(a)

$$P\left(\max_{i,j\leqslant N}\left|\frac{1}{T}\sum_{t=1}^{T}e_{ti}e_{tj}-\sigma_{ij}^{e}\right|>\frac{C\log^{1/2}(N\vee T)}{T^{1/2}}\right)<O\left((N\vee T)^{-v}\right),$$

(b)

$$P\left(\max_{i,j\leqslant N}\left|\frac{1}{T}\sum_{t=1}^{T}(e_{ti}e_{tj}-\sigma_{ij}^{e})^2-\theta_{ij}\right|>\frac{C\log^{1/2}(N\vee T)}{T^{1/2}}\right)<O\left((N\vee T)^{-v}\right),$$

(c)

$$P\left(\max_{i,j\leqslant N}\frac{1}{T}\sum_{t=1}^{T}(\hat{e}_{ti}\hat{e}_{tj}-e_{ti}e_{tj})^2>C\omega_{NT}\log^{1/2}(N\vee T)\right)<O\left((N\vee T)^{-v}\right).$$

Proof: We follow the similar proof procedures to Lemma A.3 of Fan et al. (2011) and borrowing some strategies in Ding et al. (2021). The proofs of (a) and (b) are generally based on Theorem 1 in Merlevède et al. (2009). First for (a), by Lemma A.2 in Fan et al. (2011) and Assumption 4.2, we have $e_{ti}e_{tj}$ satisfies the sub-exponential tail condition with parameter 2/3. Let $\gamma = (3/2 + r_1^{-1})^{-1}$ so that $0 < \gamma < 1$, then Theorem 1 of Merlevède et al. (2009) gives that there exist some constants $C_i > 0,\ i = 1, \ldots, 5$ that do not depend on N and T such that for any $i, j \leqslant N$ and any positive s,

$$\begin{aligned}P\left(\left|\frac{1}{T}\sum_{t=1}^{T}e_{ti}e_{tj}-\sigma_{ij}^{e}\right|>s\right) &\leqslant T\exp\left(-\frac{(Ts)^{\gamma}}{C_1}\right)+\exp\left(-\frac{(Ts)^2}{C_2(1+TC_3)}\right)\\ &+\exp\left(-\frac{(Ts)^2}{C_4T}\exp\left(\frac{(Ts)^{\gamma(1-\gamma)}}{C_5(\log Ts)^{\gamma}}\right)\right).\end{aligned}$$

Notice that by Bonferroni's inequality,

$$P\left(\max_{i,j}\left|\frac{1}{T}\sum_{t=1}^{T}e_{ti}e_{tj}-\sigma_{ij}^{e}\right|>s\right)\leqslant N^2\max_{i,j}P\left(\left|\frac{1}{T}\sum_{t=1}^{T}e_{ti}e_{tj}-\sigma_{ij}^{e}\right|>s\right).$$

Let $s = C\sqrt{\log(N \vee T)/T}$ for some large enough constant C. We have

$$\begin{aligned} & N^2 \exp\left(-\frac{(Ts)^2}{C_2(1+TC_3)}\right) = N^2 \exp\left(-\frac{C^2 T \log(N \vee T)}{C_2(1+TC_3)}\right) \\ \leqslant \quad & N^2 \exp\left(-\frac{C^2 T \log(N \vee T)}{C_2(1+TC_3)}\right) = O((N \vee T)^{-v}). \end{aligned} \tag{A.3}$$

For all $N, T \geqslant 2$, we have $Ts = C(\log(N \vee T)T)^{1/2} \geqslant T^{1/2} > 1$. Then it is easy to see that

$$\frac{(Ts)^{\gamma(1-\gamma)}}{\log^{\gamma}(N \vee T)} \geqslant \exp(\gamma)(1-\gamma)^{\gamma}$$

for $0 < \gamma < 1$ so that

$$\exp\left(\frac{(Ts)^{\gamma(1-\gamma)}}{C_5(\log(N \vee T))^{\gamma}}\right) \geqslant \exp\left(\exp(\gamma)(1-\gamma)^{\gamma} C_5^{-1}\right) = C_6.$$

Hence, we obtain

$$\begin{aligned} & N^2 \exp\left(-\frac{(Ts)^2}{C_4 T} \exp\left(\frac{(Ts)^{\gamma(1-\gamma)}}{C_5(\log Ts)^{\gamma}}\right)\right) \\ < \quad & N^2 \exp\left(-\frac{C_6 (Ts)^2}{C_4 T}\right) = O((N \vee T)^{-v}). \end{aligned} \tag{A.4}$$

Next, following the steps in the proof of Lemma A.1 in Ding et al. (2021), we have when $(\log(N \vee T))^{2/\gamma_1 - 1} = O(T)$ for some $\gamma_1 < \gamma$,

$$N^2 T \exp\left(-\frac{(Ts)^{\gamma}}{C_1}\right) \leqslant O((N \vee T)^{-v}). \tag{A.5}$$

Combining (A.3), (A.4) and (A.5), we complete the proof of (a). Part (b) can be proved in a same manner as (a). For (c), by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \max_{i,j} \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - e_{ti}e_{tj})^2 \\ = \quad & \max_{i,j} \frac{1}{T} \sum_{t=1}^{T} \left(e_{ti}(\hat{e}_{tj} - e_{tj}) + (\hat{e}_{ti} - e_{ti})e_{tj} + (\hat{e}_{ti} - e_{ti})(\hat{e}_{tj} - e_{tj})\right)^2 \\ \leqslant \quad & \frac{3}{T} \max_{i,j} \sum_{t=1}^{T} ((\hat{e}_{ti} - e_{ti})(\hat{e}_{tj} - e_{tj}))^2 + \frac{6}{T} \max_{i,j} \sum_{t=1}^{T} (e_{ti}(\hat{e}_{tj} - e_{tj}))^2 \\ \leqslant \quad & \frac{3}{T} \max_{i} \sum_{t=1}^{T} (\hat{e}_{ti} - e_{ti})^2 \max_{i,t} |\hat{e}_{ti} - e_{ti}|^2 \\ + \quad & \frac{6}{T} \left(\max_{i} \sum_{t=1}^{T} e_{ti}^4\right)^{1/2} \left(\left(\max_{i} \sum_{t=1}^{T} (\hat{e}_{ti} - e_{ti})^2\right) \left(\max_{i,t} |\hat{e}_{ti} - e_{ti}|^2\right)\right)^{1/2} \\ =: \quad & I + II. \end{aligned}$$

By Lemma A.3, $I \lesssim \omega_{NT}^2 \log(N \vee T)$. With simple algebras, the first parentheses in II can be upper bounded as for some positive constant C_e,

$$
\begin{aligned}
& \frac{1}{T^{1/2}}\left(\max_i \sum_{t=1}^T e_{ti}^4\right)^{1/2} \\
= & \left(\max_i \frac{1}{T}\sum_{t=1}^T (e_{ti}^2 - \sigma_{ii}^e + \sigma_{ii}^e)^2\right)^{1/2} \\
\leqslant & \left(2\max_i \frac{1}{T}\sum_{t=1}^T (e_{ti}^2 - \sigma_{ii}^e)^2 + 2\max_i (\sigma_{ii}^e)^2\right)^{1/2} \\
\leqslant & \left(2\max_i \left|\frac{1}{T}\sum_{t=1}^T (e_{ti}^2 - \sigma_{ii}^e)^2 - \theta_{ii}\right| + 2\max_i \theta_{ii} + 2\max_i (\sigma_{ii}^e)^2\right)^{1/2} \\
\leqslant & C_e,
\end{aligned}
$$

where the last inequality holds by part (b) and Assumption 4.2. Next, using Lemma A.3 for the second parentheses in II, we have

$$
\begin{aligned}
& \frac{1}{T^{1/2}}\left(\left(\max_i \sum_{t=1}^T (\hat{e}_{ti} - e_{ti})^2\right)\left(\max_{i,t} |\hat{e}_{ti} - e_{ti}|^2\right)\right)^{1/2} \\
= & \left(\max_i \frac{1}{T}\sum_{t=1}^T (\hat{e}_{ti} - e_{ti})^2\right)^{1/2}\left(\max_{i,t} |\hat{e}_{ti} - e_{ti}|^2\right)^{1/2} \\
\lesssim & \ \omega_{NT} \log^{1/2}(N \vee T).
\end{aligned}
$$

Thus,

$$II \lesssim \omega_{NT} \log^{1/2}(N \vee T).$$

Combining I and II, we have

$$\max_{i,j} \frac{1}{T}\sum_{t=1}^T (\hat{e}_{ti}\hat{e}_{tj} - e_{ti}e_{tj})^2 \lesssim \omega_{NT} \log^{1/2}(N \vee T)$$

holds with probability at least $1 - O((N \vee T)^{-v})$. □

Lemma A.5 *Under the assumptions of Theorem 4.1,*

$$\max_{i,j \leqslant N} |\hat{\sigma}_{ij}^e - \sigma_{ij}^e| \lesssim \omega_{NT}$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: Using the triangle inequality, we have

$$\begin{aligned}|\hat{\sigma}^e_{ij}-\sigma^e_{ij}| &\leqslant \frac{1}{T}\left|\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})(\hat{e}_{tj}-e_{tj})\right| + \frac{1}{T}\left|\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})e_{tj}\right| \\ &+ \frac{1}{T}\left|\sum_{t=1}^T(\hat{e}_{tj}-e_{tj})e_{ti}\right| + \frac{1}{T}\left|\sum_{t=1}^T(e_{ti}e_{tj}-\sigma^e_{ij})\right| \\ &=: I+II+III+IV.\end{aligned}$$

Applying the Cauchy-Schwarz inequality to I yields

$$\begin{aligned}&\max_{i,j\leqslant N}\frac{1}{T}\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})(\hat{e}_{tj}-e_{tj}) \\ \leqslant\ &\max_{i,j\leqslant N}\frac{1}{T}\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})^2 \lesssim \omega^2_{NT},\end{aligned}$$

where $\lesssim$ holds by Lemma A.3(a). We employ the Cauchy-Schwarz inequality again for II and III, and achieve

$$\begin{aligned}&\max_{i,j}\left[\frac{1}{T}\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})e_{tj},\ \frac{1}{T}\sum_{t=1}^T(\hat{e}_{tj}-e_{tj})e_{ti}\right] \\ \leqslant\ &\left\{\max_{i,j}\frac{1}{T}\sum_{t=1}^T(\hat{e}_{ti}-e_{ti})^2\right\}^{1/2}\left\{\max_{i,j}\frac{1}{T}\sum_{t=1}^T e^2_{ti}\right\}^{1/2} \\ \lesssim\ &\omega_{NT}.\end{aligned}$$

For IV, $\max_{i,j\leqslant N}T^{-1}|\sum_{t=1}^T e_{ti}e_{tj}-\sigma^e_{ij}| \lesssim T^{-1/2}\log^{1/2}(N\vee T)$ is directly from Lemma A.4(a). Combining $I-IV$ completes the proof. □

Lemma A.6 *Under the assumptions in Theorem 4.1, for $\hat{\theta}_{ij}$ defined in (3.8), there are some positive constants θ_1 and θ_2 such that the event $\Theta=\{\theta_1>\hat{\theta}_{ij}>\theta_2,\ \forall i,j>N\}$ occurs with probability at least $1-O((N\vee T)^{-v})$.*

Proof: It is sufficient to show that for θ_{ij} defined in Assumption 4.2,

$$\max_{i,j\leqslant N}|\hat{\theta}_{ij}-\theta_{ij}| \lesssim \omega_{NT}\log^{1/2}(N\vee T),$$

holds with probability at least $1 - O((N \vee T)^{-v})$. We follow the similar procedures in the proof of Lemma A.9 in Ding et al. (2021). Using the triangle inequality, we have

$$\begin{aligned}
& \max_{i,j \leqslant N} |\hat{\theta}_{ij} - \theta_{ij}| \\
\leqslant \quad & \max_{ij} \left| \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}^e_{ij})^2 - \frac{1}{T} \sum_{t=1}^{T} (e_{ti}e_{tj} - \sigma^e_{ij})^2 \right| + \max_{ij} \left| \frac{1}{T} \sum_{t=1}^{T} (e_{ti}e_{tj} - \sigma^e_{ij})^2 - \theta_{ij} \right| \\
\leqslant \quad & \max_{ij} \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}^e_{ij} - e_{ti}e_{tj} + \sigma^e_{ij})^2 \\
+ \quad & 2 \max_{ij} \left| \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}^e_{ij} - e_{ti}e_{tj} + \sigma^e_{ij})(e_{ti}e_{tj} - \sigma^e_{ij}) \right| \\
+ \quad & \max_{ij} \left| \frac{1}{T} \sum_{t=1}^{T} (e_{ti}e_{tj} - \sigma^e_{ij})^2 - \theta_{ij} \right| \\
= \quad & I + II + III
\end{aligned}$$

For I, by Lemma A.4 and Lemma A.5,

$$\begin{aligned}
I \quad & \leqslant \quad 2 \max_{ij} \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - e_{ti}e_{tj})^2 + 2 \max_{ij} (\hat{\sigma}^e_{ij} - \sigma^e_{ij})^2 \\
& \leqslant \quad \omega_{NT} \log^{1/2}(N \vee T).
\end{aligned}$$

For II, by the Cauchy-Schwarz Inequality, term I, Lemma A.4(b), and the boundedness of θ_{ij} given in Assumption 4.2, we have

$$\begin{aligned}
II \quad & \leqslant \quad 2 \max_{ij} \left(\frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{ti}\hat{e}_{tj} - \hat{\sigma}^e_{ij} - e_{ti}e_{tj} + \sigma^e_{ij})^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} (e_{ti}e_{tj} - \sigma^e_{ij})^2 \right)^{1/2} \\
& \lesssim \quad \omega_{NT}.
\end{aligned}$$

Next, by Lemma A.4 (b),

$$III \lesssim \frac{\log^{1/2}(N \vee T)}{T^{1/2}} \leqslant \omega_{NT}.$$

Combining I, II and III, we have

$$\max_{i,j \leqslant N} |\hat{\theta}_{ij} - \theta_{ij}| \lesssim \omega_{NT} \log^{1/2}(N \vee T).$$

□

Lemma A.7 *Suppose Assumption 4.2-4.3 hold, we have*

$$\|\mathbf{E}'\mathbf{F}\|_{\max} \lesssim T^{1/2}\log^{1/2}(N \vee T)$$

occurs with probability at least $1 - O((N \vee T)^{-v})$.

Proof: Let $x = C^* T^{1/2} \log^{1/2}(N \vee T)$ for some sufficient large constant C^*. we have

$$\begin{aligned} & P\left(\|\mathbf{E}'\mathbf{F}\|_{\max} > x\right) \\ \leqslant \; & P\left(\max_{i \leqslant N,\; k \leqslant K}\left|\sum\nolimits_{t=1}^{T} e_{ti} f_{tk}\right| > x\right) \\ \leqslant \; & KN \max_{i \leqslant N,\; k \leqslant K} P\left(\left|\sum\nolimits_{t=1}^{T} e_{ti} f_{tk}\right| > x\right), \end{aligned}$$

where the second inequality comes from the Boole's inequality. By Theorem 1 of Merlevède et al. (2009), it is easy to get

$$P\left(\max_{i \leqslant N,\; k \leqslant K}\left|\sum\nolimits_{t=1}^{T} e_{ti} f_{kt}\right| > C^* T^{1/2} \log^{1/2}(N \vee T)\right) \leqslant O((N \vee T)^{-v}).$$

Therefore, we have $\|\mathbf{E}'\mathbf{F}\|_{\max} \lesssim T^{1/2}\log^{1/2}(N \vee T)$ occurs with probability at least $1 - O((N \vee T)^{-v})$.

□

Lemma A.8 *Suppose Assumption 4.1 and 4.4 hold, we have*

$$\|\mathbf{E}\mathbf{B}\|_{\max} \lesssim N_1^{1/2}\log^{1/2}(N \vee T)$$

occurs with probability at least $1 - O((N \vee T)^{-v})$.

Proof: Let $x = C^* N_1^{1/2} \log^{1/2}(N \vee T)$ for some sufficient large constant C^*. we have

$$\begin{aligned} & P\left(\|\mathbf{E}\mathbf{B}\|_{\max} > x\right) \\ \leqslant \; & P\left(\max_{t \leqslant T,\; k \leqslant K}\left|\sum\nolimits_{i=1}^{N} e_{ti} b_{ik}\right| > x\right) \\ \leqslant \; & KT \max_{t \leqslant T,\; k \leqslant K} P\left(\left|\sum\nolimits_{i=1}^{N} e_{ti} b_{ik}\right| > x\right), \end{aligned}$$

where the second inequality comes from the Boole's inequality. Assumption 4.4 (b) yields for some sufficient large constant C^*,

$$
\begin{aligned}
& P\left(\max_{t\leqslant T,\ k\leqslant K}\left|\sum\nolimits_{i=1}^{N} e_{ti}b_{ik}\right| > C^* N_1^{1/2}\log^{1/2}(N\vee T)\right) \\
\leqslant\ & 2KT\exp\left\{-C^\star \log(N\vee T)\right\} \\
\leqslant\ & O((N\vee T)^{-v}),
\end{aligned}
$$

which can complete the proof.

□

Lemma A.9 *If all the assumptions in Theorem 4.1 hold, then*

$$
T^{-1/2}\|\hat{\mathbf{F}}-\mathbf{F}\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}
$$

holds with probability at least $1-O((N\vee T)^{-v})$.

Lemma A.10 *If all the assumptions in Theorem 4.1 hold, then*

$$
N_1^{-1/2}\|\hat{\mathbf{B}}-\mathbf{B}\|_{\mathrm{F}} \lesssim \frac{T^{1/2}N_1\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}
$$

holds with probability at least $1-O((N\vee T)^{-v})$.

Proof: [Lemma A.9 and Lemma A.10] The proof can be done in the same way of Theorem 2 in Uematsu and Yamagata (2022). Note that the proof will heavily depend on Assumption 4.4, Lemma A.7 and Lemma A.8. □

Proof: [Theorem 4.1]

We follow the proof strategy of Theorem 5 in Fan et al. (2013). Lemma A.6 and Lemma A.5 imply that for any $\epsilon>0$, there are some positive constants M, θ_1 and θ_2 such that the events

$$
\begin{aligned}
\boldsymbol{\Omega} &= \{\max\nolimits_{i,j\leqslant N}|\hat{\sigma}^e_{ij}-\sigma^e_{ij}|\leqslant M\omega_{NT}\},\ \text{and} \\
\boldsymbol{\Theta} &= \{\theta_2 \geqslant \hat{\theta}_{ij}^{1/2}\geqslant \theta_1,\ \forall i,j\leqslant N\}
\end{aligned}
$$

occur with probability at least $1 - O((N \vee T)^{-v})$. Then, under the event $\mathbf{\Omega} \cap \mathbf{\Theta}$, we can apply the inequalities of Theorem 5 in Fan et al. (2013) to get with probability at least $1 - O((N \vee T)^{-v})$,

$$
\begin{aligned}
\|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\| &\leqslant \max_{i \leqslant N} \sum_{j=1}^{N} |s_{ij}^{\tau}(\hat{\sigma}_{ij}^{e}) - \sigma_{ij}^{e}| \\
&\leqslant \left\{ \frac{C\theta_1 + M}{M^q} + (C\theta_1 + M)^{1-q} \right\} \omega_{NT}^{1-q} \max_{i \leqslant N} \sum_{j=1}^{N} |\sigma_{ij}^{e}|^q \\
&\lesssim \omega_{NT}^{1-q} m_N.
\end{aligned}
$$

□

Proof: [Theorem 4.2] Given $m_N \omega_{NT}^{1-q} = o(1)$ and the assumption that $\lambda_{\min}(\mathbf{\Sigma}_e) \geqslant \underline{c} > 0$, we can achieve that all the eigenvalues of $\hat{\mathbf{\Sigma}}_e^{\tau}$ are bounded from 0 with probability approaching 1 and

$$
\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\| \lesssim \omega_{NT}^{1-q} m_N
$$

holds with probability at least $1 - O((N \vee T)^{-v})$ by the similar arguments in the proof of Theorem 2.1 in Fan et al. (2011). □

Proof: [Theorem 4.3]

(a) By the triangle inequality, we have for

$$
\begin{aligned}
& \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathbf{\Sigma}}^2 \\
\lesssim \;& \|\hat{\mathbf{B}}\hat{\mathbf{B}}' - \mathbf{B}\mathbf{B}\|_{\mathbf{\Sigma}}^2 + \|\hat{\mathbf{\Sigma}}_e - \mathbf{\Sigma}_e^{\tau}\|_{\mathbf{\Sigma}}^2 \\
\lesssim \;& \|(\hat{\mathbf{B}}' - \mathbf{B}')(\hat{\mathbf{B}} - \mathbf{B})\|_{\mathbf{\Sigma}}^2 + \|\mathbf{B}(\hat{\mathbf{B}}' - \mathbf{B}')\|_{\mathbf{\Sigma}}^2 + \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|_{\mathbf{\Sigma}}^2 \\
\lesssim \;& N^{-1}\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}^4 + N^{-1}\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\|\|\mathbf{\Sigma}^{-1}\|\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}^2 + \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|_{\mathbf{\Sigma}}^2 \\
= \;& I + II + III.
\end{aligned}
$$

First,

$$
\begin{aligned}
I &= N^{-1}\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}^4 \\
&\leqslant N^{-1}\left(N \max_{i\leqslant N}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2\right)^2 \\
&\lesssim \frac{N\log^2(N\vee T)}{T^2} \qquad \text{(A.6)}
\end{aligned}
$$

where the last inequality holds due to Lemma A.2. Using the same procedures in the proof of Theorem 2 in Fan et al. (2008), we have $\|\mathbf{B}'\mathbf{\Sigma}^{-1}\mathbf{B}\| = O(1)$ so that

$$
II \leqslant N_K^{-1}\max_{i\leqslant N}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|^2 \lesssim \frac{\log(N\vee T)}{N_K T}.
$$

Next, by Theorem 4.1,

$$
\begin{aligned}
III &= N^{-1}\|\mathbf{\Sigma}^{-1/2}(\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e)\mathbf{\Sigma}^{-1/2}\|_{\mathrm{F}}^2 \\
&\leqslant \|\mathbf{\Sigma}^{-1/2}(\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e)\mathbf{\Sigma}^{-1/2}\|^2 \\
&\leqslant \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|^2\lambda_{\max}(\mathbf{\Sigma}^{-1}). \\
&\lesssim \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|^2
\end{aligned}
$$

Combining I, II, and III together, we have with probability at least $1-O((N\vee T)^{-v})$

$$
\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathbf{\Sigma}}^2 \lesssim \frac{N\log^2(N\vee T)}{T^2} + \omega_{NT}^{2-2q}m_N^2.
$$

(b) Using the triangle inequality, we derive

$$
\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} \leqslant \|\hat{\mathbf{B}}\hat{\mathbf{B}}' - \mathbf{B}\mathbf{B}'\|_{\max} + \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|_{\max} =: I + II.
$$

For I,

$$
\begin{aligned}
\|\hat{\mathbf{B}}\hat{\mathbf{B}}' - \mathbf{B}\mathbf{B}'\|_{\max} &= \max_{i,j}|\hat{\mathbf{b}}_i'\hat{\mathbf{b}}_j - \mathbf{b}_i'\mathbf{b}_j| \\
&\leqslant (\max_i\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|)^2 + 2\max_{i,j}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\|\|\mathbf{b}_j\| \\
&\lesssim \omega_{NT},
\end{aligned}
$$

where the last equality holds because of Lemma A.2. For II, by the triangle inequality and Lemma A.5,

$$\begin{aligned}\|\hat{\mathbf{\Sigma}}_e^\tau - \mathbf{\Sigma}_e\|_{\max} &= \max_{i,j}|s_{ij}^\tau(\hat{\sigma}_{ij}^e) - \sigma_{ij}^e| \\ &\leqslant \max_{i,j}|s_{ij}^\tau(\hat{\sigma}_{ij}^e) - \hat{\sigma}_{ij}^e| + \max_{i,j}|\hat{\sigma}_{ij}^e - \sigma_{ij}^e| \\ &\lesssim \tau_{ij} + \max_{i,j}|\hat{\sigma}_{ij}^e - \sigma_{ij}^e| \\ &\lesssim \omega_{NT} \end{aligned} \tag{A.7}$$

Combining I and II, we complete the proof.

□

Lemma A.11 *Under the assumptions of Theorem 4.1, we have*

$$\|\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\mathbf{\Sigma}_e^{-1}\mathbf{B}\| \lesssim m_N\omega_{NT}^{1-q}N_1$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: We follow the similar argument in the proof of Lemma B.5(i) in Fan et al. (2011). Using the triangle inequality, we have

$$\begin{aligned}& \|\hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\mathbf{\Sigma}_e^{-1}\mathbf{B}\| \\ \leqslant\ & \|(\hat{\mathbf{B}} - \mathbf{B})'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}(\hat{\mathbf{B}} - \mathbf{B})\| + 2\|(\hat{\mathbf{B}} - \mathbf{B})'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\mathbf{B}\| + \|\mathbf{B}'(\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e{}^{-1})\mathbf{B}\| \\ \leqslant\ & \|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}^2\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\| + 2\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\|\|\mathbf{B}\| + \|\mathbf{B}'\mathbf{B}\|\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e{}^{-1}\| \\ \lesssim\ & T\omega_{NT}^2 + T^{1/2}N_1^{1/2}\omega_{NT} + m_N\omega_{NT}^{1-q}N_1 \\ \leqslant\ & m_N\omega_{NT}^{1-q}N_1^{1/2}(N_1 \vee T)^{1/2},\end{aligned}$$

where $\lesssim$ holds because of Lemma A.10, Assumption 4.1 and Theorem 4.2. □

Proof: [Theorem 4.4] We use the similar proof strategy in Fan et al. (2013) with some modifications. By the Sherman-Morrison-Woodbury formula we have $\|\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}\| \leqslant$

$\sum_{i=1}^{6} A_i$, where

$$
\begin{aligned}
A_1 &= \|\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}} - \boldsymbol{\Sigma}_e^{-1}\|, \\
A_2 &= \|(\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}} - \boldsymbol{\Sigma}_e^{-1})\hat{\mathbf{B}}(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\|, \\
A_3 &= \|(\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}} - \boldsymbol{\Sigma}_e^{-1})\hat{\mathbf{B}}(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}'\boldsymbol{\Sigma}_e^{-1}\|, \\
A_4 &= \|\boldsymbol{\Sigma}_e^{-1}(\hat{\mathbf{B}} - \mathbf{B})(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}'\boldsymbol{\Sigma}_e^{-1}\|, \\
A_5 &= \|\boldsymbol{\Sigma}_e^{-1}(\hat{\mathbf{B}} - \mathbf{B})(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1}\mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\|, \text{ and} \\
A_6 &= \|\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\{(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1} - (\mathbf{I}_K + \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B})^{-1}\}\mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\|.
\end{aligned}
$$

Let $M := (\mathbf{I}_K + \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B})^{-1}$ and $L := (\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1}$. For M^{-1}, by Assumption 4.1 and Assumption 4.2,

$$
\lambda_{\min}(\mathbf{I}_K + \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}) \geqslant \lambda_{\min}(\mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}) \geqslant \lambda_{\min}(\boldsymbol{\Sigma}_e^{-1})\lambda_{\min}(\mathbf{B}'\mathbf{B}) \geqslant cN_K, \tag{A.8}
$$

for some constant $c > 0$. For L^{-1}, we first notice that condition (4.14) and Lemma A.11 imply for some constant $c' > 0$,

$$
\begin{aligned}
& P\left(\left\|\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\right\| \leqslant c' m_N \omega_{NT}^{1-q} N_1^{1/2}(N_1 \vee T)^{1/2}\right) \\
\leqslant\ & P\left(\left\|\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\right\| \leqslant c' N_K\right).
\end{aligned}
$$

Then, using line (A.8), Lemma A.1 in Fan et al. (2011), Lemma A.11 and the above result, we have

$$
\begin{aligned}
& P\left(\lambda_{\min}(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}}) \geqslant cN_K\right) \\
\geqslant\ & P\left(\left\|\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\right\| \leqslant c' N_K\right) \\
\geqslant\ & P\left(\left\|\hat{\mathbf{B}}'\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}} - \mathbf{B}'\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\right\| \leqslant c' m_N \omega_{NT}^{1-q} N_1^{1/2}(N_1 \vee T)^{1/2}\right) \\
\geqslant\ & 1 - O((N \vee T)^{-v}).
\end{aligned} \tag{A.9}
$$

For each term $A_1 - A_6$, we first observe from Theorem 4.2 that $A_1 = O_p(m_N \omega_{NT}^{1-q})$. Employing triangle inequality to A_2, we achieve

$$
A_2 \leqslant \|\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}} - \boldsymbol{\Sigma}_e^{-1}\|\|\hat{\mathbf{B}}L\hat{\mathbf{B}}'\|\|\hat{\boldsymbol{\Sigma}}_e^{\tau^{-1}}\|.
$$

It then follows from the property of matrix norms, Lemma A.10 and the proofs of Lemma B.5(iii) and Lemma A.1 in Fan et al. (2011) that for some positive constant c'' ,

$$\begin{aligned} P(\|\hat{\mathbf{B}}\| \leqslant c''N_1^{1/2}) &\geqslant P(\|\hat{\mathbf{B}}\|_{\mathrm{F}} \leqslant c''N_1^{1/2}) \\ &\geqslant 1 - O((N \vee T)^{-v}). \end{aligned} \tag{A.10}$$

Then with line (A.9) and the proof of Theorem 4.2,

$$A_2 \lesssim \frac{N_1}{N_K} m_N \omega_{NT}^{1-q}.$$

Similarly, we obtain

$$A_3 \lesssim \frac{N_1}{N_K} m_N \omega_{NT}^{1-q}.$$

For A_4, using triangle inequality again, we can derive

$$\begin{aligned} A_4 &\leqslant \|\mathbf{\Sigma}_e^{-1}(\hat{\mathbf{B}} - \mathbf{B})\|\|L\|\|\hat{\mathbf{B}}'\mathbf{\Sigma}_e{}^{-1}\| \\ &\leqslant \|\mathbf{\Sigma}_e^{-1}\|^2\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}}\|L\|\|\hat{\mathbf{B}}\|. \end{aligned} \tag{A.11}$$

Employing the upper bounds of norms in line (A.11),

$$A_4 \lesssim \frac{N_1^{1/2}T^{1/2}}{N_K}\omega_{NT}.$$

In a similar spirit, we have

$$A_5 \lesssim \frac{N_1^{1/2}T^{1/2}}{N_K}\omega_{NT}.$$

For A_6, the middle term of it can be derived as

$$\begin{aligned} & \|(\mathbf{I}_K + \hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}})^{-1} - (\mathbf{I}_K + \mathbf{B}'\mathbf{\Sigma}_e^{-1}\mathbf{B})^{-1}\| \\ = & \|L - M\| \\ = & \|L(L^{-1} - M^{-1})M\| \leqslant \|L\|\|M\|\|L^{-1} - M^{-1}\| \\ \lesssim & N_K^{-2}\|\mathbf{B}'\mathbf{\Sigma}_e^{-1}\mathbf{B} - \hat{\mathbf{B}}'\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\hat{\mathbf{B}}\| \\ \lesssim & m_N\omega_{NT}^{1-q}\frac{N_1^{1/2}(N_1 \vee T)^{1/2}}{N_K^2}, \end{aligned}$$

where the last inequality holds by (A.8) and (A.9), and the last equality concludes from Lemma A.11. Then employing Assumption 4.1 and 4.2, we have

$$\begin{aligned} A_6 &\leqslant \|\boldsymbol{\Sigma}_e^{-1}\mathbf{B}\|^2\|L-M\| \\ &\lesssim m_N\omega_{NT}^{1-q}\frac{N_1^{3/2}(N_1\vee T)^{1/2}}{N_K^2}. \end{aligned}$$

Combining $A_1 - A_6$ together, we have

$$\|\hat{\boldsymbol{\Sigma}}^{-1}-\boldsymbol{\Sigma}^{-1}\|\lesssim m_N\omega_{NT}^{1-q}\frac{N_1^{3/2}(N_1\vee T)^{1/2}}{N_K^2} \tag{A.12}$$

occurs with probability at least $1-O((N\vee T)^{-v})$. □

A.2 Case 2: Observable factors exist

Lemma A.12 *Suppose Assumption 4.2 and 4.5 hold. Then*

$$\left\|\frac{1}{T}\mathbf{X}'\mathbf{F}\right\|_{\max} \lesssim \frac{\log^{1/2}T}{T^{1/2}}\leqslant\omega_{NT}$$

occurs with probability at least $1-O((N\vee T)^{-v})$.

Proof: Let $s=C^*T^{-1/2}\log^{1/2}T$ for some sufficient large constant C^*, we have

$$\begin{aligned} & P\left(\left\|\frac{1}{T}\mathbf{X}'\mathbf{F}\right\|_{\max}>s\right) \\ \leqslant\ & P\left(\max_{l\leqslant r,\ k\leqslant K}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T}x_{tl}f_{tk}\right|>s\right) \\ \leqslant\ & Kr\max_{l\leqslant r,\ k\leqslant K}P\left(\left|\frac{1}{T}\sum\nolimits_{t=1}^{T}x_{tl}f_{tk}\right|>s\right), \end{aligned}$$

where the second inequality comes from the Boole's inequality. By Theorem 1 of Merlevède et al. (2009), it follows that

$$P\left(\max_{l\leqslant r,\ k\leqslant K}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T}x_{tl}f_{tk}\right|>C^*\frac{\log^{1/2}T}{T^{1/2}}\right)\leqslant O((N\vee T)^{-v}).$$

Thus, we have $\|T^{-1}\mathbf{X}'\mathbf{F}\|_{\max}\lesssim T^{-1/2}\log^{1/2}T$ occurs with probability at least $1-O((N\vee T)^{-v})$. Moreover, (A.2) gives that $T^{-1/2}\log^{1/2}T\leqslant\omega_{NT}$. □

Lemma A.13 *Suppose Assumption 4.2, 4.4 and 4.5 hold. Then*

$$\left\|\frac{1}{T}\mathbf{X}'\mathbf{E}\right\|_{\max} \lesssim \frac{\log^{1/2}(N \vee T)}{T^{1/2}} \leqslant \omega_{NT}$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: Let $s = C^{\star}T^{-1/2}\log^{1/2}(N \vee T)$ for some sufficient large constant $C^{\star}$, we have

$$\begin{aligned}
& P\left(\left\|\frac{1}{T}\mathbf{X}'\mathbf{E}\right\|_{\max} > s\right) \\
\leqslant \quad & P\left(\max_{l \leqslant r,\ i \leqslant N}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} x_{tl}e_{ti}\right| > s\right) \\
\leqslant \quad & Nr \max_{l \leqslant r,\ i \leqslant N} P\left(\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} x_{tl}e_{ti}\right| > s\right),
\end{aligned}$$

where the second inequality comes from the Boole's inequality. By Theorem 1 of Merlevède et al. (2009), it is easy to obtain

$$P\left(\max_{l \leqslant r,\ i \leqslant N}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} x_{tl}e_{ti}\right| > C^{\star}\frac{\log^{1/2}(N \vee T)}{T^{1/2}}\right) \leqslant O((N \vee T)^{-v}),$$

which completes the proof. □

Lemma A.14 *Suppose Assumption 4.5 holds. The following inequality holds with probability at least* $1 - O((N \vee T)^{-v})$:

$$\max_{i,j \leqslant r}\left|\frac{1}{T}\sum_{t=1}^{T} x_{ti}x_{tj} - E(x_{ti}x_{tj})\right| \lesssim \frac{\log^{1/2} T}{T^{1/2}}.$$

Proof: It can be proved in a similar way to Lemma B.1 in Fan et al. (2011). □

Proof: [Lemma 4.1] Note that $\|\hat{\mathbf{U}}-\mathbf{U}\|_{\max}=\max_{1\leqslant i\leqslant N,1\leqslant t\leqslant T}|\hat{U}_{ti}-U_{ti}|$. Using Hölder's inequality and properties of the norms, we have

$$\begin{aligned}|\hat{U}_{ti}-U_{ti}| &= |(\hat{\mathbf{a}}_i-\mathbf{a}_i)\mathbf{x}_t'| \\ &\leqslant \|\hat{\mathbf{a}}_i-\mathbf{a}_i\|_1\|\mathbf{x}_t\|_\infty \\ &= \|(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i-\mathbf{a}_i\|_1\|\mathbf{x}_t\|_\infty \\ &\leqslant \left\|\left(\frac{1}{T}\mathbf{X}'\mathbf{X}\right)^{-1}\right\|_{\max}\left\|\frac{1}{T}\mathbf{X}'\mathbf{F}\mathbf{b}_i+\frac{1}{T}\mathbf{X}'\mathbf{e}_i\right\|_{\max}\|\mathbf{x}_t\|_\infty\end{aligned}$$

Because $r\log^{1/2}T=o(T^{1/2})$, Lemma A.14 and $\lambda_{\min}(E(\mathbf{x}_t\mathbf{x}_t'))>0$, which comes from Assumption 4.5(d), following the proof of Lemma 3.1 in Fan et al. (2011), we have

$$\left\|(\mathbf{X}'\mathbf{X})^{-1}\right\|_{\max}\leqslant\left\|(\mathbf{X}'\mathbf{X})^{-1}\right\|\lesssim T^{-1} \tag{A.13}$$

occurs with probability at least $1-O((N\vee T)^{-v})$. By Assumption 4.5(b), $\|\mathbf{x}_t\|_\infty\lesssim\log^{1/2}T$ with probability at least $1-O((N\vee T)^{-v})$. Further with Lemma A.12 and Lemma A.13, we achieve

$$\max_{t,i}|\hat{U}_{ti}-U_{ti}|\lesssim\left(\left\|\frac{1}{T}\mathbf{X}'\mathbf{F}\right\|_{\max}+\left\|\frac{1}{T}\mathbf{X}'\mathbf{E}\right\|_{\max}\right)\log^{1/2}T\lesssim\frac{\log^{1/2}(N\vee T)\log^{1/2}T}{T^{1/2}}$$

occurs with probability at least $1-O((N\vee T)^{-v})$, which complete the proof. □

Lemma A.15 *Suppose Assumption 4.2 and 4.5 hold, we have*

$$\|\mathbf{E}'\mathbf{F}\|_{\max}\lesssim T^{1/2}\log^{1/2}(N\vee T)$$

occurs with probability at least $1-O((N\vee T)^{-v})$.

Proof: The proof is exactly the same as the proof of Lemma A.7. □

Lemma A.16 *Suppose Assumption 4.1 and 4.4 hold, we have*

$$\|\mathbf{E}\mathbf{B}\|_{\max}\lesssim N_1^{1/2}\log^{1/2}(N\vee T)$$

occurs with probability at least $1-O((N\vee T)^{-v})$.

Proof: The proof is exactly the same as the proof of Lemma A.8. □

Lemma A.17 *(Consistency of estimated unobserved factors and loadings) If all the assumptions in Theorem A.1 hold, the following error bounds hold with probability at least* $1 - O((N \vee T)^{-v})$,

(a)

$$T^{-1/2}\|\hat{\mathbf{F}} - \mathbf{F}\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2}\log^{1/2}(N \vee T)}{N_K(N_K \wedge T)},$$

(b)

$$N_1^{-1/2}\|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathrm{F}} \lesssim \frac{T^{1/2}N_1\log^{1/2}(N \vee T)}{N_K(N_K \wedge T)}.$$

Proof: We follow the proof strategy of Theorem 2 in Uematsu and Yamagata (2022). The optimisation of the SOFAR estimators implies that

$$\begin{aligned} & \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\mathbf{A}' + \mathbf{X}\mathbf{A}' - \mathbf{X}\hat{\mathbf{A}}' - \hat{\mathbf{F}}\hat{\mathbf{B}}'\|_{\mathrm{F}}^2 + \eta_n\|\hat{\mathbf{B}}\|_1 \\ \leqslant\ & \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\mathbf{A}' + \mathbf{X}\mathbf{A}' - \mathbf{X}\hat{\mathbf{A}}' - \mathbf{F}\mathbf{B}'\|_{\mathrm{F}}^2 + \eta_n\|\mathbf{B}'\|_1, \end{aligned}$$

which is equivalent to

$$\begin{aligned} & \frac{1}{2}\|\mathbf{E} + \hat{\mathbf{U}} - \mathbf{U} - \hat{\mathbf{F}}\hat{\mathbf{B}} + \mathbf{F}\mathbf{B}'\|_{\mathrm{F}}^2 + \eta_n\|\hat{\mathbf{B}}\|_1 \\ \leqslant\ & \frac{1}{2}\|\mathbf{E} + \hat{\mathbf{U}} - \mathbf{U}\|_{\mathrm{F}}^2 + \eta_n\|\mathbf{B}\|_1. \end{aligned} \tag{A.14}$$

Let $\mathbf{\Delta} = \hat{\mathbf{F}}\hat{\mathbf{B}}' - \mathbf{F}\mathbf{B}'$. We can rewrite (A.14) as

$$\begin{aligned} & \frac{1}{2}\|\mathbf{E} + \hat{\mathbf{U}} - \mathbf{U} - \mathbf{\Delta}\|_{\mathrm{F}}^2 + \eta_n\|\hat{\mathbf{B}}\|_1 \\ \leqslant\ & \frac{1}{2}\|\mathbf{E} + \hat{\mathbf{U}} - \mathbf{U}\|_{\mathrm{F}}^2 + \eta_n\|\mathbf{B}\|_1. \end{aligned} \tag{A.15}$$

Define $\Delta^f = \hat{\mathbf{F}} - \mathbf{F}$, $\Delta^b = \hat{\mathbf{B}} - \mathbf{B}$. We have

$$\mathbf{\Delta} = \Delta^f\mathbf{B}' + \Delta^f\Delta^{b'} + \mathbf{F}\Delta^{b'}. \tag{A.16}$$

Plugging (A.16) into (A.15) with some decomposition to (A.15) leads to

$$\begin{aligned}\frac{1}{2}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^{2} &\leqslant \operatorname{tr}(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\boldsymbol{\Delta}'+\eta_n\left(\|\mathbf{B}\|_1-\|\hat{\mathbf{B}}\|_1\right)\\ &\leqslant \left|\operatorname{tr}(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\mathbf{B}\Delta^{f'}\right|+\left|\operatorname{tr}(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\Delta^{b}\Delta^{f'}\right|\\ &+ \left|\operatorname{tr}\Delta^{b}\mathbf{F}'(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\right|+\eta_n\left(\|\mathbf{B}\|_1-\|\hat{\mathbf{B}}\|_1\right). \end{aligned}\tag{A.17}$$

Then, we use Hölder's inequality and properties of matrix norms on the traces in (A.17). The first trace is bounded as

$$\begin{aligned}&\left|\operatorname{tr}(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\mathbf{B}\Delta^{f'}\right|\\ \leqslant\ &\|(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\mathbf{B}\|_{\max}\|\Delta^{f}\|_1\\ \leqslant\ &(rT)^{1/2}\|(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\mathbf{B}\|_{\max}\|\Delta^{f}\|_{\mathrm{F}}\\ \leqslant\ &(rT)^{1/2}\|\mathbf{E}\mathbf{B}\|_{\max}\|\Delta^{f}\|_{\mathrm{F}}+rT\|\hat{\mathbf{U}}-\mathbf{U}\|_{\max}\|\mathbf{B}\|_{\max}\|\Delta^{f}\|_{\mathrm{F}}.\end{aligned}\tag{A.18}$$

In a similar spirit of bounding the first trace, the second trace is bounded as

$$\begin{aligned}&\left|\operatorname{tr}(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\Delta^{b}\Delta^{f'}\right|\\ =\ &\left|\operatorname{tr}\mathbf{E}\Delta^{b}\Delta^{f'}+\operatorname{tr}(\hat{\mathbf{U}}-\mathbf{U})\Delta^{b}\Delta^{f'}\right|\\ \leqslant\ &\|\mathbf{E}\Delta^{b}\|_2\|\Delta^{f}\|_*+\|\hat{\mathbf{U}}-\mathbf{U}\|_{\max}\|\Delta^{b}\Delta^{f'}\|_1\\ \leqslant\ &\|\mathbf{E}\|_2\|\Delta^{b}\|_{\mathrm{F}}\|\Delta^{f}\|_{\mathrm{F}}+(NT)^{1/2}\|\hat{\mathbf{U}}-\mathbf{U}\|_{\max}\|\Delta^{b}\|_{\mathrm{F}}\|\Delta^{f'}\|_{\mathrm{F}},\end{aligned}\tag{A.19}$$

where the last $\lesssim$ holds because of properties of matrix norms and Lemma 4.1. Similarly, the third trace in (A.17) is bounded as

$$\left|\operatorname{tr}\Delta^{b}\mathbf{F}'(\mathbf{E}+\hat{\mathbf{U}}-\mathbf{U})\right|\leqslant\|\Delta^{b}\|_1\|\mathbf{F}'\mathbf{E}\|_{\max}+\|\Delta^{b}\|_1\|\mathbf{F}'(\hat{\mathbf{U}}-\mathbf{U})\|_{\max}.\tag{A.20}$$

By properties of matrix norms, the second term of the right-hand side of (A.20) is bounded as

$$
\begin{aligned}
& \|\Delta^b\|_1 \|\mathbf{F}'(\hat{\mathbf{U}} - \mathbf{U})\|_{\max} \\
\leqslant \; & (rK)^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|\hat{\mathbf{A}} - \mathbf{A}\|_{\max} \qquad \text{(A.21)} \\
= \; & (rK)^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{F}\mathbf{B} + \mathbf{X}'\mathbf{E})\|_{\max} \\
\leqslant \; & rK^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|(\mathbf{X}'\mathbf{X})^{-1}\|_{\max} \|\mathbf{X}'\mathbf{F}\mathbf{B} + \mathbf{X}'\mathbf{E}\|_{\max} \\
\leqslant \; & rK^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|(\mathbf{X}'\mathbf{X})^{-1}\|_{\max} \left((rK)^{1/2} \|\mathbf{X}'\mathbf{F}\|_{\max} \|\mathbf{B}\|_{\max} + \|\mathbf{X}'\mathbf{E}\|_{\max} \right) \\
\leqslant \; & rK^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|(\mathbf{X}'\mathbf{X})^{-1}\|_{\max} \left((rK)^{1/2} c_b \|\mathbf{X}'\mathbf{F}\|_{\max} + \|\mathbf{X}'\mathbf{E}\|_{\max} \right), \qquad \text{(A.22)}
\end{aligned}
$$

where the last inequality holds because of Assumption 4.1. Hence,

$$
\begin{aligned}
& \left| \mathrm{tr} \Delta^b \mathbf{F}'(\mathbf{E} + \hat{\mathbf{U}} - \mathbf{U}) \right| \\
\leqslant \; & \|\Delta^b\|_1 \|\mathbf{F}\mathbf{E}\|_{\max} \\
+ \; & rK^{1/2} \|\Delta^b\|_1 \|\mathbf{F}'\mathbf{X}\|_{\max} \|(\mathbf{X}'\mathbf{X})^{-1}\|_{\max} \left((rK)^{1/2} c_b \|\mathbf{X}'\mathbf{F}\|_{\max} + \|\mathbf{X}'\mathbf{E}\|_{\max} \right). \qquad \text{(A.23)}
\end{aligned}
$$

From Assumption 4.4 on $\mathbf{E}$, (A.13), Lemma 4.1, Lemma A.16 about $\mathbf{EB}$ Lemma A.15 about $\mathbf{E}'\mathbf{F}$, Lemma A.12 about $\mathbf{X}'\mathbf{F}$ and Lemma A.13 about $\mathbf{X}'\mathbf{E}$, there exist some positive constants c_1–c_6 such that the event $\mathcal{E}$ with probability at least $1 - O((N \vee T)^{-v})$ for large fixed constant $v > 0$:

$$
\begin{aligned}
\mathcal{E} \; = \; & \{\|\mathbf{E}\|_2 \leqslant c_1 (N \vee T)^{1/2}\} \cap \left\{ \|\mathbf{E}\mathbf{B}\|_{\max} \leqslant c_2 N_1^{1/2} \log^{1/2}(N \vee T) \right\} \\
\cap \; & \left\{ \|\mathbf{F}'\mathbf{E}\|_{\max} \leqslant c_3 T^{1/2} \log^{1/2}(N \vee T) \right\} \cap \left\{ \|\mathbf{X}'\mathbf{E}\|_{\max} \leqslant c_4 T^{1/2} \log^{1/2}(N \vee T) \right\} \\
\cap \; & \left\{ \|\mathbf{X}'\mathbf{F}\|_{\max} \leqslant c_5 T^{1/2} \log^{1/2} T \right\} \cap \left\{ \| (\mathbf{X}'\mathbf{X})^{-1} \|_{\max} \leqslant c_6 T^{-1} \right\}. \qquad \text{(A.24)}
\end{aligned}
$$

On this event, setting the penalty term $\eta_n = 2c_3 T^{1/2} \log^{1/2}(N \vee T)$ and employing the trace bounds (A.18), (A.19) and (A.23) on (A.17) give

$$\begin{aligned}\|\mathbf{\Delta}\|_{\mathrm{F}}^2 &\lesssim (N_1T)^{1/2}\log^{1/2}(N\vee T)\|\Delta^f\|_{\mathrm{F}} + (N\vee T)^{1/2}\log^{1/2}T\|\Delta^f\|_{\mathrm{F}}\|\Delta^b\|_{\mathrm{F}} \\ &+ \eta_n\left(\|\Delta^b\|_1 + 2\eta_n\|\mathbf{B}\|_1 - 2\|\hat{\mathbf{B}}\|_1\right). \qquad (A.25)\end{aligned}$$

Next, directly following the steps starting from (A.6) in the supplementary material of Uematsu and Yamagata (2022), we can complete the proof. Note that although compared with (A.6) in the aforementioned material, the second term of the right-hand side of (A.25) inludes an extra term $\log^{1/2}T$ due to the first stage OLS estimation, this extra term will be divided by a much faster polynomial rate in the proof so that there will be no effect on the final result. □

Lemma A.18 *If all the assumptions in Theorem A.1 are satisfied, then*

$$T^{-1}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 \lesssim \frac{N_1^3\log(N\vee T)}{N_K^2(N_K\wedge T)^2} = \omega_{NT}^2$$

holds with probability at least $1 - O((N\vee T)^{-v})$.

Proof: Note that

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t - \mathbf{f}_t\|^2 = \frac{1}{T}\|\hat{\mathbf{F}} - \mathbf{F}\|_{\mathrm{F}}^2 \lesssim \omega_{NT}^2,$$

where $\lesssim$ holds because of Lemma A.17. □

Lemma A.19 *Under the assumptions in Theorem A.1, we have*

$$\max_{i\leqslant N}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \lesssim \frac{\log^{1/2}(N\vee T)}{T^{1/2}} \leqslant \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}$$

holds with probability at least $1 - O((N\vee T)^{-v})$.

Proof: Similar to the proof of Lemma (A.2), the SOFAR estimator $\hat{\mathbf{B}}$ under KKT conitions is given by

$$\begin{aligned}\hat{\mathbf{B}} - \mathbf{B} &= T^{-1}(\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}} - \mathbf{F}) + \mathbf{E}'(\hat{\mathbf{F}} - \mathbf{F})) + T^{-1}\mathbf{E}'\mathbf{F} - T^{-1}\eta_n\mathbf{V}(\hat{\mathbf{B}}) \\ &- T^{-1}(\hat{\mathbf{A}}\mathbf{X}' - \mathbf{A}\mathbf{X}')(\hat{\mathbf{F}} - \mathbf{F}) - T^{-1}(\hat{\mathbf{A}}\mathbf{X}' - \mathbf{A}\mathbf{X}')\mathbf{F}, \qquad (A.26)\end{aligned}$$

where the $(i,k)th$ element of $\mathbf{V}(\tilde{\mathbf{B}})$ for given $\tilde{\mathbf{B}} = \tilde{b}_{ik} \in \mathbb{R}^{N\times K}$ is defined as

$$v_{ik}(\tilde{\mathbf{B}}) \begin{cases} = \operatorname{sgn}(\tilde{b}_{ik}) & \text{for } \tilde{b}_{ik} \neq 0 \\ \in [-1,1] & \text{for } \tilde{b}_{ik} = 0. \end{cases}$$

Using the triangle inequality,

$$\begin{aligned} & \max_{i\leqslant N} \|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \\ \leqslant \; & T^{-1}\eta_n + T^{-1}\left(\|\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} + \|\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max}\right) + T^{-1}\|\mathbf{E}'\mathbf{F}\|_{\max} \\ + \; & T^{-1}\left(\|(\hat{\mathbf{A}}\mathbf{X}'-\mathbf{A}\mathbf{X}')(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} + \|(\hat{\mathbf{A}}\mathbf{X}'-\mathbf{A}\mathbf{X}')\mathbf{F}\|_{\max}\right). \end{aligned} \tag{A.27}$$

Note that we set $\eta_n \asymp T^{1/2}\log^{1/2}(N\vee T)$ in Section 4. By the proof of Theorem 1 in Uematsu and Yamagata (2021), we can directly get that in (A.27),

$$T^{-1}\left(\|\mathbf{B}\mathbf{F}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} + \|\mathbf{E}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max}\right) \lesssim T^{-1/2}\frac{N_1^{3/2}\log(N\vee T)}{N_K(N_K\wedge T)}$$

holds with probability at least $1-O((N\vee T)^{-v})$.

Define $\mathcal{F} = \{\hat{\mathbf{F}}-\mathbf{F} \in \mathbb{R}^{T\times K} : \|\hat{\mathbf{F}}-\mathbf{F}\|_{\mathrm{F}} \leqslant CR_n\}$, where C is some constant >0 and

$$R_n = \frac{N_1^{3/2}T^{1/2}\log(N\vee T)}{N_K(N_K\wedge T)}.$$

Then, Lemma A.17 yields that $\hat{\mathbf{F}}-\mathbf{F} \in \mathcal{F}$ occurs with probability at least $1-O((N\vee T)^{-v})$. Conditional on $\hat{\mathbf{F}}-\mathbf{F} \in \mathcal{F}$, we also have $\hat{\mathbf{F}}-\mathbf{F} = R_n\mathbf{M}$ for some matrix $\mathbf{M}$ such that $\|\mathbf{M}\|_{\mathrm{F}} \leqslant C$.

Next, using the similar proof strategy of Theorem 1 in Uematsu and Yamagata (2021), we have, for any $x>0$,

$$\begin{aligned} & P\left(T^{-1/2}\|(\hat{\mathbf{A}}\mathbf{X}'-\mathbf{A}\mathbf{X}')(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} > x\right) \\ \leqslant \; & P\left(\|T^{-1/2}(\hat{\mathbf{A}}\mathbf{X}'-\mathbf{A}\mathbf{X}')(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} > x \mid \hat{\mathbf{F}}-\mathbf{F}\in\mathcal{F}\right) + P\left(\hat{\mathbf{F}}-\mathbf{F}\notin\mathcal{F}\right) \\ \leqslant \; & P\left((rK)^{1/2}\|\hat{\mathbf{A}}-\mathbf{A}\|_{\max}\|T^{-1/2}\mathbf{X}'(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} > x \mid \hat{\mathbf{F}}-\mathbf{F}\in\mathcal{F}\right) + O((N\vee T)^{-v}) \\ \leqslant \; & P\left(R_nT^{-1/2}\log^{1/2}(N\vee T)\|T^{-1/2}\mathbf{X}'\mathbf{M}\|_{\max} \gtrsim x \mid \hat{\mathbf{F}}-\mathbf{F}\in\mathcal{F}\right) + O((N\vee T)^{-v}), \end{aligned}$$

where the third $\leqslant$ uses (A.21) and (A.22) to get $\|\hat{\mathbf{A}}-\mathbf{A}\|_{\max} \lesssim T^{-1/2}\log^{1/2}(N\vee T)$. Setting $x = T^{-1/2}R_n\log^{1/2}(N\vee T)$ can yield that the upper bound is $O((N\vee T)^{-v})$. In a similar spirit, we obtain that for any $x > 0$,

$$P\left(T^{-1/2}\|(\hat{\mathbf{A}}\mathbf{X}' - \mathbf{A}\mathbf{X}')\mathbf{F}\|_{\max} > x\right)$$

is also smaller than $O((N\vee T)^{-v})$. Thus, we have

$$T^{-1}\left(\|(\hat{\mathbf{A}}\mathbf{X}' - \mathbf{A}\mathbf{X}')(\hat{\mathbf{F}}-\mathbf{F})\|_{\max} + \|(\hat{\mathbf{A}}\mathbf{X}' - \mathbf{A}\mathbf{X}')\mathbf{F}\|_{\max}\right) \lesssim T^{-1/2}\frac{N_1^{3/2}\log(N\vee T)}{N_K(N_K\wedge T)}$$

holds with probability at least $1 - O((N\vee T)^{-v})$. Lemma A.15 gives $T^{-1}\|\mathbf{E}'\mathbf{F}\|_{\max} \lesssim T^{-1/2}\log^{1/2}(N\vee T)$. Consequently, we have

$$\max_{i\leqslant N}\|\hat{\mathbf{b}}_i - \mathbf{b}_i\| \lesssim \frac{\log^{1/2}(N\vee T)}{T^{1/2}} \leqslant \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)} = \omega_{NT},$$

where $\leqslant$ holds because condition (4.11) ensures that

$$\frac{N_1^{3/2}T^{1/2}}{N_K(N_K\wedge T)} \geqslant \frac{N_1^{1/2}T^{1/2}}{(N_K\wedge T)} \geqslant 1.$$

□

Lemma A.20 *If all the conditions in Theorem A.1 are satisfied, then the following inequalities hold with probability at least* $1 - O((N\vee T)^{-v})$:

(a) $\max_{i\leqslant N}\frac{1}{T}\sum_{t=1}^{T}|\hat{e}_{ti} - e_{ti}|^2 \lesssim \tilde{\omega}_{NT}^2$,

(b) $\max_{i\leqslant N, t\leqslant T}|\hat{e}_{ti} - e_{ti}|^2 \lesssim \log(N\vee T)$.

Proof: Let $\Delta_i^{\mathbf{b}} = \hat{\mathbf{b}}_i - \mathbf{b}_i$ and $\Delta_t^{\mathbf{f}} = \hat{\mathbf{f}}_t - \mathbf{f}_t$. We first rewrite $|\hat{e}_{ti} - e_{ti}|$ as

$$|\hat{e}_{ti} - e_{ti}| = |\hat{U}_{ti} - U_{ti} + \Delta_t^{\mathbf{f}}\Delta_i^{\mathbf{b}\prime} + \Delta_t^{\mathbf{f}}\mathbf{b}_i{}' + \mathbf{f}_t\Delta_i^{\mathbf{b}\prime}|.$$

Then, by a usual inequality $(A+B+C+D)^2 \leqslant 4(A^2+B^2+C^2+D^2)$, Theorem 4.1, Lemma A.12 and Lemma A.13, we get

$$\begin{aligned}
& \max_{i\leqslant N} \frac{1}{T}\sum_{t=1}^{T}|\hat{e}_{ti}-e_{ti}|^2 \\
\leqslant\ & \frac{4}{T}\sum_{t=1}^{T}(\hat{U}_{ti}-U_{ti})^2 + 4\max_i \|\mathbf{b}_i\|^2 \frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2 \\
+\ & 4\max_i \|\hat{\mathbf{b}}_i-\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t\|^2 + 4\max_i \|\hat{\mathbf{b}}_i-\mathbf{b}_i\|^2\frac{1}{T}\sum_{t=1}^{T}\|\mathbf{f}_t\|^2 \\
\lesssim\ & \max\left\{\frac{\log(N\vee T)\log T}{T},\ \omega_{NT}^2\right\} = \tilde{\omega}_{NT}^2.
\end{aligned}$$

For (b), by Lemma 4.1 and the proof of Lemma A.3(b), we upper bound $|\hat{e}_{ti}-e_{ti}|^2$ as

$$\begin{aligned}
& \max_{i\leqslant N, t\leqslant T}|\hat{e}_{ti}-e_{ti}|^2 \\
=\ & \max_{i\leqslant N, t\leqslant T}|(\hat{U}_{ti}-U_{ti}) + (\hat{\mathbf{b}}_i'\hat{\mathbf{f}}_t-\mathbf{b}_i'\mathbf{f}_t)|^2 \\
\leqslant\ & \max_{i\leqslant N, t\leqslant T}|\hat{U}_{ti}-U_{ti}|^2 + 2\max_{i\leqslant N, t\leqslant T}|(\hat{U}_{ti}-U_{ti})(\hat{\mathbf{b}}_i'\hat{\mathbf{f}}_t-\mathbf{b}_i'\mathbf{f}_t)| + \max_{i\leqslant N, t\leqslant T}|\hat{\mathbf{b}}_i'\hat{\mathbf{f}}_t-\mathbf{b}_i'\mathbf{f}_t|^2 \\
\lesssim\ & \log(N\vee T).
\end{aligned}$$

□

Lemma A.21 *Under the assumptions of Theorem A.1, there exists a large enough constant $C>0$ such that for large T and N,*

(a)

$$P\left(\max_{i,j\leqslant N}\left|\frac{1}{T}\sum_{t=1}^{T}e_{ti}e_{tj}-\sigma_{ij}^e\right| \geqslant \frac{C\log^{1/2}(N\vee T)}{T^{1/2}}\right) > O\left((N\vee T)^{-v}\right),$$

(b)

$$P\left(\max_{i,j\leqslant N}\left|\frac{1}{T}\sum_{t=1}^{T}(e_{ti}e_{tj}-\sigma_{ij}^e)^2-\theta_{ij}\right| > \frac{C\log^{1/2}(N\vee T)}{T^{1/2}}\right) \leqslant O\left((N\vee T)^{-v}\right),$$

(c)

$$P\left(\max_{i,j\leqslant N}\frac{1}{T}\sum_{t=1}^{T}(\hat{e}_{ti}\hat{e}_{tj}-e_{ti}e_{tj})^2 > C\omega_{NT}\log^{1/2}(N\vee T)\right) \leqslant O\left((N\vee T)^{-v}\right).$$

Proof: The proofs can be done in the same way of Lemma A.4, so we skip the details here. Note that Lemma A.20 will be used frequently. □

Lemma A.22 *Under the assumptions in Theorem A.1, there exist some positive constants* θ_3 *and* θ_4 *such that the event* $\Theta = \{\theta_3 > \hat{\theta}_{ij} > \theta_4, \ \forall i, j > N\}$ *occurs with probability at least* $1 - O((N \vee T)^{-v})$.

Proof: The proof can be done in the same way of Lemma A.6. Note that the proof will heavily depend on Lemma A.21. □

Lemma A.23 *Under the assumptions of Theorem A.1, for some large enough constant* C,

$$P\left(\max_{i,j\leqslant N} |\hat{\sigma}^e_{ij} - \sigma^e_{ij}| \leqslant C\tilde{\omega}_{NT}\right) \geqslant 1 - O((N \vee T)^{-v}).$$

Proof: The proof is the same as the Lemma A.5 with applications of Lemma A.20 and Lemma A.21. □

Next, we present the assertions in Theorem 4.5 as Theorem A.1-A.2 as well as their proofs separately.

Theorem A.1 *Suppose Assumption 4.1-4.2, 4.4-4.5 and condition (4.11) hold. Then, for* $\hat{\boldsymbol{\Sigma}}^{\tau}_e$ *defined in (3.9), we have*

(a) $\|\hat{\boldsymbol{\Sigma}}^{\tau^{-1}}_e - \boldsymbol{\Sigma}_e\| \lesssim \tilde{\omega}^{1-q}_{NT} m_N$ *holds with probability at least* $1 - O((N \vee T)^{-v})$;

(b) If $m_N \tilde{\omega}^{1-q}_{NT} = o(1)$ *also holds, then all the eigenvalues of* $\hat{\boldsymbol{\Sigma}}^{\tau}_e$ *are bounded from 0 with probability approaching 1 and*

$$\|\hat{\boldsymbol{\Sigma}}^{\tau^{-1}}_e - \boldsymbol{\Sigma}^{-1}_e\| \lesssim \tilde{\omega}^{1-q}_{NT} m_N$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: [Theorem A.1] The proofs follow the same strategies of Theorem 4.1 and 4.2. Note that the proofs depend on the new versions of Lemma A.6 and Lemma A.5, which are Lemma A.22 and Lemma A.23 respectively. □

Theorem A.2 *Under the assumptions in Theorem A.1, for $\hat{\mathbf{\Sigma}}$ defined in (3.9), the following results hold with probability at least $1 - O((N \vee T)^{-v})$:*

(a) $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|^2_{\mathbf{\Sigma}} \lesssim T^{-2} N \log^2(N \vee T) + \tilde{\omega}_{NT}^{2-2q} m_N^2$,

(b) $\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} \lesssim \tilde{\omega}_{NT}$.

Proof: [Theorem A.2]

(a) By the triangle inequality, we have

$$\begin{aligned}
& \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|^2_{\Sigma} \\
\lesssim\ & \Big(\|\mathbf{A}(\hat{\text{cov}}(\mathbf{x}_t) - \text{cov}(\mathbf{x}_t))\mathbf{A}'\|^2_{\Sigma} + \|\mathbf{A}\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|^2_{\Sigma} \\
+\ & \|(\hat{\mathbf{A}} - \mathbf{A})\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|^2_{\Sigma} \Big) \\
+\ & \Big(\|\hat{\mathbf{B}}\hat{\mathbf{B}} - \mathbf{B}\mathbf{B}\|^2_{\Sigma} + \|\hat{\mathbf{\Sigma}}^{\tau}_e - \mathbf{\Sigma}_e\|^2_{\Sigma} \Big) \\
=:\ & I + II.
\end{aligned} \tag{A.28}$$

Following the proofs of Lemma B.3 in Fan et al. (2011), we have

(i) $\|\mathbf{A}(\hat{\text{cov}}(\mathbf{x}_t) - \text{cov}(\mathbf{x}_t))\mathbf{A}'\|^2_{\Sigma} + \|\mathbf{A}\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|^2_{\Sigma} \lesssim \log(N \vee T) T^{-1}$,

(ii) $\|(\hat{\mathbf{A}} - \mathbf{A})\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|^2_{\Sigma}) \lesssim \log^2(N \vee T) T^{-2}$

hold with probability at least $1 - O((N \vee T)^{-v})$. Then, by the proof of Theorem 4.3, we obtain with probability at least $1 - O((N \vee T)^{-v})$,

$$I + II \lesssim \frac{N \log^2(N \vee T)}{T^2} + \tilde{\omega}_{NT}^{2-2q} m_N^2.$$

(b) Note that

$$
\begin{aligned}
& \|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\max} \\
\leqslant \; & \Big(\|2\mathbf{A}\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|_{\max} + \|\mathbf{A}(\hat{\text{cov}}(\mathbf{x}_t) - \text{cov}(\mathbf{x}_t))\mathbf{A}'\|_{\max} \\
+ \; & \|(\hat{\mathbf{A}} - \mathbf{A})\hat{\text{cov}}(\mathbf{x}_t)(\hat{\mathbf{A}} - \mathbf{A})'\|_{\max} \\
+ \; & \|2\mathbf{A}(\hat{\text{cov}}(\mathbf{x}_t) - \text{cov}(\mathbf{x}_t))(\hat{\mathbf{A}} - \mathbf{A})'\|_{\max} \\
+ \; & \|(\hat{\mathbf{A}} - \mathbf{A})(\hat{\text{cov}}(\mathbf{x}_t) - \text{cov}(\mathbf{x}_t))(\hat{\mathbf{A}} - \mathbf{A})'\|_{\max} \Big) \\
+ \; & \Big(\|\hat{\mathbf{B}}\hat{\mathbf{B}} - \mathbf{B}\mathbf{B}\|_{\max} + \|\hat{\mathbf{\Sigma}}_e^{\tau} - \mathbf{\Sigma}_e\|_{\max} \Big) \\
=: \; & I + II.
\end{aligned} \tag{A.29}
$$

It then follows from Theorem 4.3 and the proof of Theorem 3.2 (b) in Fan et al. (2011) that $I + II \lesssim \tilde{\omega}_{NT}$ holds with probability at least $1 - O((N \vee T)^{-v})$.

□

Next, we state and prove the lemmas for proving Theorem 4.6. For notational ease, let us denote $\mathbf{B}\mathbf{B}' + \mathbf{A}\text{cov}(\mathbf{x}_t)\mathbf{A}' = \mathbf{\Sigma}_{lr}$, $\hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\mathbf{A}}\hat{\text{cov}}(\mathbf{x}_t)\hat{\mathbf{A}}' = \hat{\mathbf{\Sigma}}_{lr}$, $\mathbf{\Sigma} = \mathbf{\Sigma}_{lr} + \mathbf{\Sigma}_e$, and $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}_{lr} + \hat{\mathbf{\Sigma}}_e^{\tau}$.

Lemma A.24 *Under the assumptions in Theorem 4.6. We have*

$$
\|\hat{\mathbf{\Sigma}}_{lr} - \mathbf{\Sigma}_{lr}\| \lesssim \Big(N \vee (N_1^{1/2} T^{1/2}) \Big) m_N \tilde{\omega}_{NT}
$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: We use the triangle inequality repeatedly. It follows that for some sufficient large constant $C > 0$,

$$\begin{aligned}
& \|\hat{\mathbf{\Sigma}}_{lr} - \mathbf{\Sigma}_{lr}\| \\
\leqslant \quad & \|\hat{\mathbf{B}} - \mathbf{B}\|^2 + \|\mathbf{B}\|\|\hat{\mathbf{B}} - \mathbf{B}\| + \|(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathrm{cov}}(\mathbf{x}_t) - \mathrm{cov}(\mathbf{x}_t))\mathbf{A}\| \\
+ \quad & \|\hat{\mathbf{A}} - \mathbf{A}\|\|\mathrm{cov}(\mathbf{x}_t)\|\|\mathbf{A}\| + \|\mathbf{A}(\hat{\mathrm{cov}}(\mathbf{x}_t) - \mathrm{cov}(\mathbf{x}_t))\mathbf{A}'\| \\
\leqslant \quad & C\Big(T\tilde{\omega}_{NT}^2 + N_1^{1/2}T^{1/2}\tilde{\omega}_{NT} + N\frac{\log^{1/2} N \log^{1/2} T}{T} + N^{1/2}\frac{\log^{1/2} N}{T^{1/2}} + N\frac{\log^{1/2} T}{T^{1/2}}\Big) \\
\lesssim \quad & \Big(N \vee (N_1^{1/2}T^{1/2})\Big)\tilde{\omega}_{NT} \qquad \text{(A.30)}
\end{aligned}$$

□

Lemma A.25 *Under the assumptions in Theorem 4.6. We have*

$$\|\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_{lr}\mathbf{\Sigma}_e^{-1}\| \lesssim \Big(N \vee (N_1^{1/2}T^{1/2})\Big)m_N\tilde{\omega}_{NT}^{1-q}$$

holds with probability at least $1 - O((N \vee T)^{-v})$.

Proof: We use the triangle inequality repeatedly. It follows that for some sufficient large constant C,

$$\begin{aligned}
& \|\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_{lr}\mathbf{\Sigma}_e^{-1}\| \\
\leqslant \quad & \|\hat{\mathbf{\Sigma}}_{lr} - \mathbf{\Sigma}_{lr}\|\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\| + \|\hat{\mathbf{\Sigma}}_{lr} - \mathbf{\Sigma}_{lr}\|\|\mathbf{\Sigma}_e^{-1}\| + \|\mathbf{\Sigma}_{lr}\|\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}} - \mathbf{\Sigma}_e^{-1}\| \\
\leqslant \quad & C\Big(N \vee (N_1^{1/2}T^{1/2})\Big)\tilde{\omega}_{NT}m_N\tilde{\omega}_{NT}^{1-q} + C\Big(N \vee (N_1^{1/2}T^{1/2})\Big)\tilde{\omega}_{NT} \\
+ \quad & CNm_N\tilde{\omega}_{NT}^{1-q},
\end{aligned}$$

where the last inequality holds due to Lemma A.24, Theorem A.1, $\|\mathbf{\Sigma}_{lr}\| = O(N)$ and $\|\mathbf{\Sigma}_e\| = O(1)$. Note that with simple algebras, it can be verified that

$$\Big(N \vee (N_1^{1/2}T^{1/2})\Big)\tilde{\omega}_{NT} \leqslant N$$

by condition (4.11). Then, the result follows. □

Lemma A.26 *Under the assumptions in Theorem 4.6. We have for some sufficient large constant c,*

$$P\Big(\lambda_{\min}\left(\mathbf{I}+\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\right)\geqslant cN\Big)\quad\geqslant\quad 1-O((N\vee T)^{-v}). \tag{A.31}$$

Proof: Because we have $\lambda_{\min}(\mathbf{\Sigma}_e^{-1})\geqslant \underline{c}$, $\lambda_{\min}(\mathbf{AA}')\geqslant c_a N$, $\lambda_{\min}(\mathbf{BB}')\geqslant d_K^2 N_K$ and $\lambda_{\max}(\mathrm{cov}(\mathbf{x}_t))\geqslant c_l$ by assumptions, we achieve for some positive constant $c'>0$,

$$\begin{aligned}\lambda_{\min}\left(\mathbf{I}+\mathbf{\Sigma}_{lr}\mathbf{\Sigma}_e^{-1}\right) &\geqslant \lambda_{\min}(\mathbf{\Sigma}_{lr})\lambda_{\min}(\mathbf{\Sigma}_e^{-1})\\ &= \lambda_{\min}(\mathbf{A}\mathrm{cov}(\mathbf{x}_t)\mathbf{A}'+\mathbf{BB}')\lambda_{\min}(\mathbf{\Sigma}_e^{-1})\\ &\geqslant c'N \end{aligned}\tag{A.32}$$

Then, Lemma A.1 in Fan et al. (2011) with (A.32) gives

$$\begin{aligned}&P\Big(\lambda_{\min}\left(\mathbf{I}+\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\right)\geqslant cN\Big)\\ \geqslant\quad & P\Big(\|(\mathbf{I}+\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}})-(\mathbf{I}+\mathbf{\Sigma}_{lr}\mathbf{\Sigma}_e^{-1})\|\leqslant cN\Big)\\ \geqslant\quad & P\Big(\|\hat{\mathbf{\Sigma}}_{lr}\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}-\mathbf{\Sigma}_{lr}\mathbf{\Sigma}_e^{-1}\|\leqslant c\Big(N\vee(N_1^{1/2}T^{1/2})\Big)m_N\tilde{\omega}_{NT}^{1-q}\Big),\end{aligned}$$

where the last inequality holds because of condition (4.15). Finally by Lemma A.25, the proof completes. □

Proof: [Theorem 4.6] By the triangle inequality and Binomial Inverse Theorem[4], we have

$$
\begin{aligned}
& \|\hat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\| \\
= \quad & \|(\hat{\boldsymbol{\Sigma}}_{lr} + \hat{\boldsymbol{\Sigma}}_e^{\tau})^{-1} - (\boldsymbol{\Sigma}_{lr} + \boldsymbol{\Sigma}_e)^{-1}\| \\
\leqslant \quad & \|\hat{\boldsymbol{\Sigma}}_e^{\tau-1} - \boldsymbol{\Sigma}_e^{-1}\| \\
+ \quad & \left\| \left(\hat{\boldsymbol{\Sigma}}_e^{\tau-1} - \boldsymbol{\Sigma}_e^{-1}\right) \left(\mathbf{I} + \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}\right)^{-1} \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1} \right\| \\
+ \quad & \left\| \boldsymbol{\Sigma}_e^{-1} \left(\mathbf{I} + \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}\right)^{-1} \left(\hat{\boldsymbol{\Sigma}}_{lr} - \boldsymbol{\Sigma}_{lr}\right) \hat{\boldsymbol{\Sigma}}_e^{-1} \right\| \\
+ \quad & \left\| \boldsymbol{\Sigma}_e^{-1} \left(\mathbf{I} + \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}\right)^{-1} \boldsymbol{\Sigma}_{lr} \left(\hat{\boldsymbol{\Sigma}}_e^{-1} - \boldsymbol{\Sigma}_e^{-1}\right) \right\| \\
+ \quad & \left\| \boldsymbol{\Sigma}_e^{-1} \left(\left[\mathbf{I} + \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}\right]^{-1} - \left[\mathbf{I} + \boldsymbol{\Sigma}_{lr}\boldsymbol{\Sigma}_e^{-1}\right]^{-1} \right) \boldsymbol{\Sigma}_{lr}\boldsymbol{\Sigma}_e^{-1} \right\| \\
=: \quad & L_1 + L_2 + L_3 + L_4 + L_5.
\end{aligned}
$$

First, let $G := \mathbf{I} + \boldsymbol{\Sigma}_{lr}\boldsymbol{\Sigma}_e^{-1}$ and $\hat{G} := \mathbf{I} + \hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}$. We have

$$\|G\| \leqslant \lambda_{\max}(\boldsymbol{\Sigma}_{lr})\lambda_{\max}(\boldsymbol{\Sigma}_e^{-1}) = O(N) \tag{A.33}$$

by Assumption 4.1, 4.2 and 4.5. Theorem A.1 directly gives $L_1 \lesssim m_N \tilde{\omega}_{NT}^{1-q}$. For L_2, it follows from the triangle inequality that

$$L_2 \quad \leqslant \quad L_1 \|\hat{G}^{-1}\| \|\hat{\boldsymbol{\Sigma}}_{lr}\hat{\boldsymbol{\Sigma}}_e^{\tau-1}\|. \tag{A.34}$$

We have

$$\|\hat{G}^{-1}\| \lesssim N^{-1} \tag{A.35}$$

holds with probability at least $1 - O((N \vee T)^{-v})$ by Lemma A.26. Next, Lemma A.24 and the triangle inequality gives

$$\|\hat{\boldsymbol{\Sigma}}_{lr}\| \leqslant \|\hat{\boldsymbol{\Sigma}}_{lr} - \boldsymbol{\Sigma}_{lr}\| + \|\boldsymbol{\Sigma}_{lr}\| \lesssim N. \tag{A.36}$$

[4]See Press (1972).

Similarly, by Theorem A.1, we have $\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\|$ is bounded with probability at least $1-O((N\vee T)^{-v})$. Thus, $L_2 \lesssim m_N\tilde{\omega}_{NT}^{1-q}$ with probability at least $1-O((N\vee T)^{-v})$. Similarly for L_3, we have

$$L_3 \lesssim \|\hat{G}^{-1}\|\|\hat{\mathbf{\Sigma}}_{lr}-\mathbf{\Sigma}_{lr}\| \lesssim N^{-1}Nm_N\tilde{\omega}_{NT}^{1-q} = m_N\tilde{\omega}_{NT}^{1-q} \tag{A.37}$$

because of Lemma A.24, Lemma A.26, $\|\mathbf{\Sigma}_e^{-1}\| = O(1)$ and $\|\hat{\mathbf{\Sigma}}_e^{\tau^{-1}}\|$ being bounded with probability at least $1-O((N\vee T)^{-v})$. It is easy to get $L_4 \lesssim m_N\tilde{\omega}_{NT}^{1-q}$. For L_5, we have

$$\begin{aligned} L_5 &\leqslant \|\mathbf{\Sigma}_e^{-1}\|\|\hat{G}^{-1}-G^{-1}\|\|\mathbf{\Sigma}_{lr}^{-1}\|\|\mathbf{\Sigma}_e^{-1}\| \\ &\leqslant \|\mathbf{\Sigma}_e^{-1}\|\|\mathbf{\Sigma}_{lr}^{-1}\|\|\mathbf{\Sigma}_e^{-1}\|\|\hat{G}-G\|\|\hat{G}^{-1}\|\|G^{-1}\| \\ &\lesssim N^{-1}Nm_N\tilde{\omega}_{NT}^{1-q}N^{-2} \\ &= m_N\tilde{\omega}_{NT}^{1-q} \end{aligned} \tag{A.38}$$

because of $\|\mathbf{\Sigma}_e^{-1}\| = O(1)$, $\|\hat{G}^{-1}-G^{-1}\|\|\mathbf{\Sigma}_{lr}^{-1}\| = O(N)$, Lemma A.25, line (A.33) and line (A.35). Combing $L_1 - L_5$ can complete the proof. □

B Choice of the Threshold Tuning Parameters

In practice, the threshold constant C_τ in the threshold level $\tau_{ij} = C_\tau\omega_{NT}(\theta_{ij})^{1/2}$ is determined by users. Following the procedures of Bickel and Levina (2008a) and Fan et al. (2013), we use multi-fold cross-validations (CV) to choose C_τ:

Step 1. Obtain residuals $\{\hat{\mathbf{e}}_t\}_{t=1}^T$ from the only observable variables $\hat{\mathbf{U}}$ by our sparse-induced weak factor models.

Step 2. Divide $\{\hat{\mathbf{e}}_t\}_{t=1}^T$ randomly into two groups, M_1 and M_2. Let M_1 be the training group $\{\hat{\mathbf{e}}_t\}_{t\in M_1}$ with size $T(M_1)$, and M_2 be the validation group $\{\hat{\mathbf{e}}_t\}_{t\in M_2}$ with size $T(M_2)$, where $T(M_1) = \lfloor T(1-\log^{-1}T)\rfloor$ and $T(M_1)+T(M_2) = T$.

Step 3. Repeat Step 1-2 H times and denote each time as h. Then, select the optimal tuning parameters C_τ^* by the following Frobenius risk

$$C_\tau^* = \underset{C_\tau \in \left[C^{\min}+\epsilon,\ \bar{C}\right]}{\arg\min} \frac{1}{H} \sum_{h=1}^{H} \left\| \hat{\boldsymbol{\Sigma}}_e^\tau (C_\tau)^{M_1,h} - \hat{\boldsymbol{\Sigma}}_e^{M_2,h} \right\|_{\mathrm{F}}^2 .$$

Here, for each time h, $\hat{\boldsymbol{\Sigma}}_e^\tau (C_\tau)^{M_1,h}$ is the ePOET estimator using $\{\hat{\mathbf{e}}_t\}_{t\in M_1}^h$ with the threshold constant ω_{NT}, and $\hat{\boldsymbol{\Sigma}}_e^{M_2,h}$ is the sample covariance matrix using $\{\hat{\mathbf{e}}_t\}_{t\in M_2}^h$. Regarding the interval of ω_{NT}, $C^{\min}$ is the minimum value that guarantees the positive definiteness of estimated idiosyncratic covariance matrices, $\epsilon > 0$ is small, and $\bar{C}$ is some sufficient large constant determined by users.

C Discussion of the SOFAR and PC Estimators

Define

$$\gamma_n = \frac{N^{1/2}(N_r \wedge T)^{1/2}}{N_1^{1/2} T^{1/2}}.$$

Uematsu and Yamagata (2022) show that the PC estimates for the sWF model hold with probability at least $1 - O((N \vee T)^{-v})$:

(a)

$$T^{-1/2} \|\hat{\mathbf{F}}_{\mathrm{PC}} - \mathbf{F}\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_K (N_K \wedge T)} (1 + \gamma_n),$$

(b)

$$N_1^{-1/2} \|\hat{\mathbf{B}}_{\mathrm{PC}} - \mathbf{B}\|_{\mathrm{F}} \lesssim \frac{T^{1/2} N_1 \log^{1/2}(N \vee T)}{N_K (N_K \wedge T)} (1 + \gamma_n).$$

And recall that the SOFAR estimates hold with probability at least $1 - O((N \vee T)^{-v})$:

(a)

$$T^{-1/2} \|\hat{\mathbf{F}} - \mathbf{F}\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_K (N_K \wedge T)},$$

(b)

$$N_1^{-1/2}\|\hat{\mathbf{B}}-\mathbf{B}\|_{\mathrm{F}} \lesssim \frac{T^{1/2}N_1\log^{1/2}(N\vee T)}{N_K(N_K\wedge T)}.$$

It can be seen that when model contains strong factors only, namely $N_k = N$ for all $k = 1, ..., K$, the convergence rates of factor loadings in both cases, for example, reduce to $\log^{1/2}(N\vee T)T^{-1/2}$, which is identical to the original POET with given number of factors. The extra term γ_n appears in the PC estimates only. When γ_n is bounded in probability (i.e., $N_1 = N$ or $T > N$), PC is identical to SOFAR. Thus, in such case, the ePOET estimator converges at least as fast as the POET estimator. However, when $N_1 < N$ and relatively smaller than T, there are non-negligible extra costs for PC to recover the weak factor structure in the sWF model, and the SOFAR estimates can achieve a sharper upper bound than the PC estimates. Thus, covariance estimators of ePOET can converge faster than those of the original POET.