ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

Discussion Paper No. 1080

INFERENCE IN WEAK FACTOR MODELS

Yoshimasa Uematsu
Takashi Yamagata

March 2020

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Inference in Weak Factor Models

YOSHIMASA UEMATSU* and TAKASHI YAMAGATA†

*Department of Economics and Management, Tohoku University
†Department of Economics and Related Studies, University of York
†Institute of Social Economic Research, Osaka University

March 12, 2020

Abstract

In this paper, we consider statistical inference for high-dimensional approximate factor models. We posit a weak factor structure, in which the factor loading matrix can be sparse and the signal eigenvalues may diverge more slowly than the cross-sectional dimension, N. We propose a novel inferential procedure to decide whether each component of the factor loadings is zero or not, and prove that this controls the false discovery rate (FDR) below a pre-assigned level, while the power tends to unity. This "factor selection" procedure is primarily based on a de-sparsified (or debiased) version of the WF-SOFAR estimator of Uematsu and Yamagata (2020), but is also applicable to the principal component (PC) estimator. After the factor selection, the *re-sparsified* WF-SOFAR and *sparsified* PC estimators are proposed and their consistency is established. Finite sample evidence supports the theoretical results. We apply our procedure to the FRED-MD macroeconomic and financial data, consisting of 128 series from June 1999 to May 2019. The results strongly suggest the existence of sparse factor loadings and exhibit a clear association of each of the extracted factors with a group of macroeconomic variables. In particular, we find a price factor, housing factor, output and income factor, and a money, credit and stock market factor.

Keywords. Approximate factor models, Debiased SOFAR estimator, Multiple testing, FDR and Power, Re-sparsification.

1 Introduction

The factor models have become an increasingly important tool for the analysis of psychology, finance, economics, and biology, among many others. This paper discusses statistical inference for high-dimensional *approximate factor models*. These were first introduced by Chamberlain and Rothschild (1983), then developed in subsequent articles by Connor and Korajczyk (1986, 1993), Bai and Ng (2002), Bai (2003), Fan et al. (2008), and Fan et al. (2011, 2013), among many others.

*Yoshimasa Uematsu is Associate Professor, Department of Economics and Management, Tohoku University, 27-1 Kawauchi, Aobaku, Sendai 980-8576, Japan (E-mail: yoshimasa.uematsu.e7@tohoku.ac.jp). He gratefully acknowledges the partial support of JSPS KAKENHI JP19K13665.

†Takashi Yamagata is Professor, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK (E-mail: takashi.yamagata@york.ac.uk).He gratefully acknowledges the partial support of JSPS KAKENHI JP15H05728 and JP18K01545. The authors appreciate Kun Chen giving helpful suggestions and modification of the R package, rrpack.

1.1 Factor models

Suppose that a vector of zero-mean stationary time series $\mathbf{x}_t \in \mathbb{R}^N$, $t = 1, \ldots, T$, is generated from the factor model $\mathbf{x}_t = \mathbf{B}^* \mathbf{f}_t^* + \mathbf{e}_t$, where $\mathbf{B}^* = (b_{ik}^*) \in \mathbb{R}^{N \times r}$ is a matrix of deterministic factor loadings, $\mathbf{f}_t^* \in \mathbb{R}^r$ is a vector of zero-mean latent factors, and $\mathbf{e}_t \in \mathbb{R}^N$ is an idiosyncratic error vector. To separately identify factors and factor loadings, we choose a specific (but frequently employed) rotation which imposes r^2 restrictions, and hereafter we consider this model without loss of generality:

$$\mathbf{x}_t = \mathbf{B}^0 \mathbf{f}_t^0 + \mathbf{e}_t, \tag{1}$$

where $\mathbf{f}_t^0 = \mathbf{H}\mathbf{f}_t^*$ and $\mathbf{B}^{0\prime} = \mathbf{H}^{-1}\mathbf{B}^{*\prime}$ with $\boldsymbol{\Sigma}_f = \mathbb{E}[\mathbf{f}_t^0 \mathbf{f}_t^{0\prime}] = \mathbf{I}_r$ and $\mathbf{B}^{0\prime}\mathbf{B}^0$ being a diagonal matrix. Assuming uniform boundedness of the maximum eigenvalue of $\mathbb{E}[\mathbf{e}_t \mathbf{e}_t']$, the asymptotic property of $\mathbb{E}[\mathbf{x}_t \mathbf{x}_t']$ is dictated by the r largest eigenvalues of $\mathbf{B}^0 \mathbf{B}^{0\prime}$. Specifically, Chamberlain and Rothschild (1983) assume the condition, $\lambda_r(\mathbf{B}^0\mathbf{B}^{0\prime}) \to \infty$ as $N \to \infty$. In order to consider the estimation, we need a stronger condition. Most studies, including Connor and Korajczyk (1986, 1993), Stock and Watson (2002), Bai and Ng (2002, 2006, 2013), and Bai (2003), suppose $\lambda_k(\mathbf{B}^0\mathbf{B}^{0\prime}) \asymp N$ for all $k = 1, \ldots, r$. The model with this condition is called the *strong factor (SF) model*. In view of the real data, the SF assumption is much more restrictive than that of Chamberlain and Rothschild (1983). In this paper, following Uematsu and Yamagata (2020), we consider *weak factor (WF) models* with sparse factor loadings that lead to $\lambda_k(\mathbf{B}^0\mathbf{B}^{0\prime}) \asymp N^{\alpha_k}$ for some constants $1 \geq \alpha_1 \geq \cdots \geq \alpha_r > 0$.

Uematsu and Yamagata (2020) investigate the estimation of the WF models. In particular, extending Uematsu et al. (2019), they propose the WF-SOFAR (simply denoted as SOFAR hereafter) estimator and its adaptive version, the latter of which yields *factor selection consistency* (which is an analogous concept of variable selection consistency in the lasso literature). In this paper, we consider statistical inference on the factor selection without relying on the adaptive SOFAR.

1.2 Toward Global inferences

In line with the literature on the adaptive lasso for high-dimensional linear models, the asymptotic normality of the adaptive SOFAR estimator could also be established for the nonzero elements of the estimator. It was thought to be useful for statistical inference, but has been criticized by, e.g. Leeb and Pötscher (2008) and Pötscher and Leeb (2009), who argue that the property lacks uniformity over sequences of models that include even minor deviations from the so-called beta-min condition (see Chernozhukov et al., 2015, Ch. 6). The same criticism could apply to the adaptive SOFAR estimator.

Instead of the adaptive lasso, several methods have been proposed for inference in high-dimensional linear regressions. Especially, the method called *debiasing* (*desparsification*) by Javanmard and Montanari (2014), van de Geer et al. (2014), and Zhang and Zhang (2014) has gained popularity. This framework tries directly to remove the bias using the Karush-Kuhn-Tucker (KKT) conditions, and achieves the asymptotic normality.

Let $\mathcal{S}$ denote the support (index set of nonzero elements) of a p-dimensional unknown parameter of interest. Given $\mathcal{H} \subset \{1, \ldots, p\}$, consider testing for a pair of hypotheses

$$H_0 : j \in \mathcal{S}^c \ \text{ for all } \ j \in \mathcal{H} \quad \text{v.s.} \quad H_1 : j \in \mathcal{S} \ \text{ for some } \ j \in \mathcal{H}. \tag{2}$$

Such hypotheses are statistically testable based on the asymptotic normality. This conventional hypothesis testing is sometimes labeled as a *local* inference since it only focuses on a subset of indexes, $\mathcal{H}$. It is noteworthy that rejection of H_0 is not informative as it merely tells us that *not all* the elements in $\mathcal{H}$ are null variables. This fact is fostered especially when $|\mathcal{H}|$ is vary large. Alternatively, it is more interesting to investigate whether *each* entry in $\{1, \dots, p\}$ is significantly null or not. To this end, we attempt to consider a *multiple testing* for a sequence of pairs of hypotheses

$$H_0^{(j)} : j \in \mathcal{S}^c \quad \text{v.s.} \quad H_1^{(j)} : j \in \mathcal{S} \quad \text{for each} \quad j \in \{1, \dots, p\}. \tag{3}$$

In such multiple testing problems, it is important to control the number of false discoveries (type I errors) while pursuing a higher power. A classical measure of type I errors is the *family-wise error rate* (FWER) and can be controlled by the methods of Bonferroni (1935) or Holm (1979), for instance. However, these procedures will lead to a very conservative variable selection, especially in high dimensions. Instead of the FWER, in the context of the multiple testing problem with which we are concerned, it is more suitable to control another measure of type I errors: the *false discovery rate* (FDR). The FDR was first introduced by Benjamini and Hochberg (1995) and is defined as the expectation of the falsely discovered proportion (FDP):

$$\text{FDR} = \mathbb{E}\,\text{FDP} \qquad \text{with} \qquad \text{FDP} = \frac{|\mathcal{S}^c \cap \widehat{\mathcal{S}}|}{|\widehat{\mathcal{S}}| \vee 1},$$

where $\widehat{\mathcal{S}} \subset \{1, \dots, p\}$ is a set of discovered indexes by some statistical procedure. The associated power is defined as

$$\text{Power} = \mathbb{E}\left[\frac{|\mathcal{S} \cap \widehat{\mathcal{S}}|}{|\mathcal{S}| \vee 1}\right].$$

The FDR controlled multiple testing is expected to keep high power even in high-dimensional settings. This inferential framework can be called a *global* inference, in contrast with the local inference for (2).

1.3 Contributions

In light of the recent development of global inferences described above, we propose the debiased SOFAR estimator of the sparse loadings in the WF models, and establish its asymptotic normality. In addition, we show that the PC estimator is asymptotically normal even for the WF models. This is an extension of Bai (2003), which deals only with the SF models.

Building upon the asymptotic normality of the factor loading estimators, we consider statistical inference on the factor selection. More precisely, we consider multiple testing like (3) for the sequence, $H_0^{(i,k)} : b_{ik}^0 = 0$ v.s. $H_1^{(i,k)} : b_{ik}^0 \neq 0$ for $i = 1, \dots, N$ and $k = 1, \dots, r$, and propose a method to control the FDR which is inspired by Liu (2013) and Javanmard and Javadi (2019). We prove that this method asymptotically controls the FDR below a pre-assigned level while the power tends to unity. Although the theory is established for the debiased SOFAR estimator, the method works with any asymptotically normal estimators, such as the PC estimator: whereas the latter can be less efficient as it cannot effectively utilize the sparseness of the loadings. Indeed, the Monte Carlo experiments suggest that the debiased SOFAR estimator is normally approximated very well while the PC estimator is not, as the model becomes weaker (sparser). It also shows that the proposed method controls the FDR while keeping the high power satisfactory.

After the global inference, the natural loading matrix estimator is the debiased SOFAR estimator, with its insignificant elements being replaced with zeros. We coin it a *re-sparsified* SOFAR estimator. Moreover, we propose a *sparsified* PC estimator, which is obtained after the global inference based on the PC loading matrix estimator in a similar manner. We also establish its consistency. Since these estimators inherit the asymptotic normality of the debiased SOFAR and PC estimators, they can be attractive alternatives to the adaptive SOFAR under the recent situation in which the inference of the latter had reached an impasse.

We apply our factor selection procedure to the FRED-MD dataset of macroeconomic and financial variables, which consist of a balanced panel of 128 monthly series spanning the period from June 1999 to May 2019. The results give very strong evidence of sparse factor loadings under the identification restrictions, and exhibit a clear association of factors and groups of macroeconomic variables. The first factor is associated with five variable groups and can be seen as a semi-global factor. Each of the remaining four factors is associated with just one or two dominating groups. Specifically, we find a price factor, housing factor, output and income factor, and a money, credit and stock market factor.

1.4 Notational remarks and organization

For any matrix $\mathbf{M} = (m_{ti}) \in \mathbb{R}^{T\times N}$, we denote by $\|\mathbf{M}\|_{\mathrm{F}}$, $\|\mathbf{M}\|_2$, $\|\mathbf{M}\|_1$, and $\|\mathbf{M}\|_{\max}$ the Frobenius norm, ℓ_2-induced (spectral) norm, entrywise ℓ_1-norm, and entrywise ℓ_∞-norm, respectively. Specifically, they are defined by $\|\mathbf{M}\|_{\mathrm{F}} = (\sum_{t,i} m_{ti}^2)^{1/2}$, $\|\mathbf{M}\|_2 = \lambda_1^{1/2}(\mathbf{M}'\mathbf{M})$, $\|\mathbf{M}\|_1 = \sum_{t,i}|m_{ti}|$, and $\|\mathbf{M}\|_{\max} = \max_{t,i}|m_{ti}|$, where $\lambda_i(\mathbf{S})$ refers to the ith largest eigenvalue of any square matrix $\mathbf{S}$. Denote by $\mathbf{I}_N$ and $\mathbf{0}_{T\times N}$ the $N\times N$ identity matrix and $T\times N$ matrix with all the entries being zero, respectively. We use $\lesssim$ ($\gtrsim$) to represent $\leq$ ($\geq$) up to a positive constant factor. For any positive sequence a_n and b_n that converge to some points or diverge as $n\to\infty$, we write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$. Moreover, denote by $a_n \sim b_n$ if $a_n/b_n \to 1$. We also use $X\sim\mu$ to signify that random variable X has distribution μ. For any positive values a and b, $a\vee b$ and $a\wedge b$ stand for $\max(a,b)$ and $\min(a,b)$, respectively. The indicator function is denoted by $1\{\cdot\}$. For any $k\in\mathbb{N}$, write $[k]$ to represent $\{1,\dots,k\}$.

The paper is organized as follows. Section 2 formally defines the WF models. Section 3 proposes the methodology of global inference for the sparse loadings. Section 4 explores the statistical theory for the FDR control and power guarantee of our method. Section 5 confirms the finite sample validity via Monte Carlo experiments. Section 6 applies our method to a large macroeconomic dataset. Section 7 concludes. All the proofs of our theoretical results are collected in the Appendix, and supplementary analyses are in the Online Appendix.

2 Weak Factor Models

Suppose that an N-dimensional vector of zero-mean stationary time series $\{\mathbf{x}_t\}_{t=1}^T$ is generated from the factor model of (1). Under the identification restrictions imposed in the Introduction, $\mathbb{E}[\mathbf{f}_t^0\mathbf{f}_t^{0\prime}] = \mathbf{I}_r$ and $\mathbf{B}^{0\prime}\mathbf{B}^0$ being a diagonal matrix with different elements, while assuming an exogeneity condition, we have

$$\mathbf{\Sigma}_x = \mathbf{B}^0\mathbf{B}^{0\prime} + \mathbf{\Sigma}_e, \tag{4}$$

where $\mathbf{\Sigma}_x = \mathbb{E}[\mathbf{x}_t\mathbf{x}_t']$ and $\mathbf{\Sigma}_e = \mathbb{E}[\mathbf{e}_t\mathbf{e}_t']$. We investigate the case in which N and T diverge at the same time. For the sake of convenience, we assume the existence of an underlying divergent sequence n such that $N = N(n) \to \infty$ and $T = T(n) \to \infty$ as $n \to \infty$. For example, we may simply suppose $n = N \wedge T \to \infty$. In Section 4, we also write $T = N^\tau$ for the constant $\tau > 0$ to understand the size of T relative to N. The number of factors r is unknown and to be determined in advance. Stacking the vectors vertically like $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_T)'$, $\mathbf{F}^0 = (\mathbf{f}_1^0, \dots, \mathbf{f}_T^0)'$, and $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$, we equivalently rewrite model (1) as the matrix form

$$\mathbf{X} = \mathbf{F}^0\mathbf{B}^{0\prime} + \mathbf{E} = \mathbf{C}^0 + \mathbf{E}, \tag{5}$$

where $\mathbf{C}^0$ is called the matrix of common components.

As mentioned in the Introduction, Chamberlain and Rothschild (1983) consider approximate factor models (5) allowing possibly different divergence rates of $\lambda_j(\mathbf{\Sigma}_x)$ for $j = 1, \dots, r$ while $\lambda_{r+1}(\mathbf{\Sigma}_x)$ is bounded, which has recently been called the WF structure. In this paper, we consider the *sparsity-induced* WF models. Specifically, we assume *exactly sparse* factor loadings $\mathbf{B}^0$ such that the sparsity of kth column (i.e., the number of nonzero elements in $\mathbf{b}_k^0 \in \mathbb{R}^N$) is given by $N_k = N^{\alpha_k}$ for $k \in [r]$, where $N \geq N_1 \geq \cdots \geq N_r$ (i.e., $1 \geq \alpha_1 \geq \cdots \geq \alpha_r > 0$) and α_k's are unknown. Note that N_r must diverge since $\alpha_r > 0$ and $N \to \infty$. Combining the sparsity assumption with the identification restriction, we then observe that there exist some constants $d_1 \geq \cdots \geq d_r > 0$ such that

$$\mathbf{B}^{0\prime}\mathbf{B}^0 = \mathrm{diag}(d_1^2 N_1, \dots, d_r^2 N_r).$$

Therefore, under the assumption of uniform boundedness of $\lambda_j(\mathbf{\Sigma}_e)$, it is not hard to see that

$$\lambda_j(\mathbf{\Sigma}_x) \begin{cases} \asymp \lambda_j(\mathbf{B}^0\mathbf{B}^{0\prime}) = d_j^2 N_j & \text{for} \quad j \in [r], \\ \text{is uniformly bounded} & \text{for} \quad j \in [N]\backslash[r], \end{cases}$$

where the equality in the first line holds because $\lambda_j(\mathbf{B}^0\mathbf{B}^{0\prime}) = \lambda_j(\mathbf{B}^{0\prime}\mathbf{B}^0)$ for $j \in [r]$. This specification appears to fulfil the requirement of the WF structure. Define $\mathcal{S} := \mathrm{supp}(\mathbf{B}^0) \subset [N] \times [r]$ and $s := |\mathcal{S}| = \sum_{k=1}^r N_k$. Thus $|\mathcal{S}^c| = Nr - s$.

3 Inferential Methodology

We introduce a new inferential framework for the WF models. First we propose a new estimator that can converge weakly to a normal distribution by *debiasing* the SOFAR estimator. Using the estimator, we next consider *global* inference on the sparsity pattern of $\mathbf{B}^0$ based on a multiple testing with the FDR control. The formal theory of these results is developed in the next section.

For the WF models introduced in Section 2, Uematsu et al. (2019) and Uematsu and Yamagata (2020) proposed the SOFAR estimator,

$$(\widehat{\mathbf{F}}, \widehat{\mathbf{B}}) = \underset{(\mathbf{F},\mathbf{B}) \in \mathbb{R}^{T\times\hat{r}} \times \mathbb{R}^{N\times\hat{r}}}{\arg\min} \left\{ \frac{1}{2} \left\| \mathbf{X} - \mathbf{F}\mathbf{B}' \right\|_{\mathrm{F}}^2 + \eta_n \|\mathbf{B}\|_1 \right\} \tag{6}$$
$$\text{subject to } \mathbf{F}'\mathbf{F}/T = \mathbf{I}_{\hat{r}} \text{ and } \mathbf{B}'\mathbf{B} \text{ diagonal},$$

where $\eta_n > 0$ is a regularization coefficient. Setting $\eta_n = 0$ eventuates in the PC estimator.

The SOFAR estimator can be more efficient than the PC estimator for WF models because it provides sparse estimates, while the PC does not. A key ingredient for inference is asymptotic normality, but it is impossible for the SOFAR estimator to have this property due to the bias caused by the regularization, as with the lasso estimator.

3.1 Debiasing the SOFAR estimator

For inference in high-dimensional linear models, Javanmard and Montanari (2014), van de Geer et al. (2014), and Zhang and Zhang (2014) proposed the debiased (desparsified) lasso estimator that can converge weakly to a normal distribution. In the same spirit, we introduce the *debiased SOFAR estimator* to recover its asymptotic normality. Regarding optimization (6), consider the KKT condition:

$$\widehat{\mathbf{B}}\widehat{\mathbf{F}}'\widehat{\mathbf{F}} - \mathbf{X}'\widehat{\mathbf{F}} + \eta_n \mathbf{V}(\widehat{\mathbf{B}}) = \mathbf{0}_{N\times r}, \tag{7}$$

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

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

Recall that $\mathbf{C}^0 = \mathbf{F}^0\mathbf{B}^{0\prime}$ and $\widehat{\mathbf{C}} = \widehat{\mathbf{F}}\widehat{\mathbf{B}}'$. From (7) with the restriction $\widehat{\mathbf{F}}'\widehat{\mathbf{F}} = T\mathbf{I}$, we have

$$\begin{aligned} T^{-1}\eta_n \mathbf{V}(\widehat{\mathbf{B}}) &= T^{-1}(\mathbf{X} - \widehat{\mathbf{C}})'\widehat{\mathbf{F}} \\ &= -(\widehat{\mathbf{B}} - \mathbf{B}^0) - T^{-1}\mathbf{B}^0\mathbf{F}^{0\prime}(\widehat{\mathbf{F}} - \mathbf{F}^0) + T^{-1}\mathbf{E}'(\widehat{\mathbf{F}} - \mathbf{F}^0) + T^{-1}\mathbf{E}'\mathbf{F}^0 \\ &=: -(\widehat{\mathbf{B}} - \mathbf{B}^0) + T^{-1/2}\mathbf{R} + T^{-1/2}\mathbf{Z}, \end{aligned} \tag{8}$$

where $\mathbf{Z} := T^{-1/2}\mathbf{E}'\mathbf{F}^0$ and $\mathbf{R} := \mathbf{R}^{(1)} + \mathbf{R}^{(2)}$ with $\mathbf{R}^{(1)} := T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}(\widehat{\mathbf{F}} - \mathbf{F}^0)$ and $\mathbf{R}^{(2)} := T^{-1/2}\mathbf{E}'(\widehat{\mathbf{F}} - \mathbf{F}^0)$. We may expect that each row of $\mathbf{Z}$ converges weakly to a multivariate normal distribution while the bias term $\mathbf{R}$ is asymptotically negligible. From this observation, we define the debiased SOFAR estimator:

$$\widehat{\mathbf{B}}^{\mathsf{d}} := \widehat{\mathbf{B}} + T^{-1}(\mathbf{X} - \widehat{\mathbf{C}})'\widehat{\mathbf{F}} = \mathbf{B}^0 + T^{-1/2}\mathbf{R} + T^{-1/2}\mathbf{Z}. \tag{9}$$

Remark 1. Unlike the debiased lasso for high-dimensional linear models, the debiased SOFAR for the WF models does not require approximation of the inverse covariance matrix. This is because the "covariate" $\widehat{\mathbf{f}}_t$ is low-dimensional and satisfies $\widehat{\mathbf{F}}'\widehat{\mathbf{F}} = T\mathbf{I}$. As a result, the behavior of the estimator is stable.

Remark 2. It is well-known that Bai (2003) established the asymptotic normality of the PC estimator for the SF models (i.e., $\alpha_r = 1$), but the inferential theory has not been fully investigated for the WF models with $\alpha_1 < 1$. In the next section, we will derive the asymptotic normality and consider the theoretical properties through comparison with the debiased SOFAR.

3.2 Asymptotic t-test

Each row of the debiased SOFAR estimator (9) can admit asymptotic normality under regularity conditions:

$$T^{1/2}\left(\widehat{\mathbf{b}}_i^{\mathsf{d}} - \mathbf{b}_i^0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Gamma}_i\right), \tag{10}$$

where $\mathbf{\Gamma}_i = \lim_{T\to\infty} T^{-1}\sum_{s,t=1}^{T} \mathbb{E}[\mathbf{f}_s^0 \mathbf{f}_t^{0\prime} e_{si} e_{ti}]$. In order to consider inference based on the asymptotic normality (10), a consistent estimator of the covariance matrix $\mathbf{\Gamma}_i$ is needed. As suggested for the PC estimator in the SF model of Bai (2003), the HAC estimator of Newey and West (1987) is provided:

$$\widehat{\mathbf{\Gamma}}_i = \widehat{\mathbf{\Gamma}}_{0i} + \sum_{h=1}^{H}\left(1 - \frac{h}{H+1}\right)(\widehat{\mathbf{\Gamma}}_{hi} + \widehat{\mathbf{\Gamma}}_{hi}'), \tag{11}$$

where $\widehat{\mathbf{\Gamma}}_{hi} = T^{-1}\sum_{t=h+1}^{T} \widehat{\mathbf{f}}_t \hat{e}_{ti} \hat{e}_{t-h,i} \widehat{\mathbf{f}}_{t-h}'$ with H diverging at the rate $H = o(T^{1/4})$. Once the consistent estimator is obtained, the conventional asymptotic t-test can be implemented.

3.3 Global inference for the loadings

From the discussion so far, the debiased SOFAR estimator can be used for significance tests thanks to the expected asymptotic normality. As mentioned in the Introduction, we consider a *multiple testing* of a sequence of a pair of hypotheses like (3):

$$H_0^{(i,k)} : b_{ik}^0 = 0 \quad \text{v.s.} \quad H_1^{(i,k)} : b_{ik}^0 \neq 0 \quad \text{for each} \quad (i,k) \in [N] \times [r]. \tag{12}$$

For each (i, k), we define the t-statistic as

$$\mathtt{T}_{ik} := \frac{\sqrt{T}\hat{b}_{ik}^{\mathsf{d}}}{\hat{\sigma}_{ik}}, \tag{13}$$

where $\hat{\sigma}_{ik}^2$ is the kth diagonal element of $\widehat{\mathbf{\Gamma}}_i$ introduced in (11). Repeating the t-test with the "conventional" critical value, 1.96, for each hypothesis will apparently fail in controlling the type I error. Instead, we construct a new critical value $\mathtt{t} \geq 0$ that leads to the FDR control of discoveries $\widehat{\mathcal{S}}$, defined as the rejected indexes, $\{(i,k) : |\mathtt{T}_{ik}| \geq \mathtt{t}\}$. More precisely, the following procedure yields a relevant critical value and corresponding active set that asymptotically controls the FDR to be less than or equal to a predetermined level.

Procedure 1. Denote by $R(\mathtt{t}) = \sum_{(i,k)\in[N]\times[r]} 1\{|\mathtt{T}_{ik}| \geq \mathtt{t}\}$ the total number of rejections in the multiple testing for (12).

1. For any target FDR level $q \in [0, 1]$, define $\bar{\mathtt{t}} = \sqrt{2\log(Nr) - a\log\log(Nr)}$ with arbitrary fixed $a > 2$ and

$$\mathtt{t}_0 = \inf\left\{\mathtt{t} \in [0, \bar{\mathtt{t}}] : \frac{NrG(\mathtt{t})}{R(\mathtt{t}) \vee 1} \leq q\right\}, \tag{14}$$

where $G(\mathtt{t}) = 2(1 - \Phi(\mathtt{t}))$ with Φ the standard normal distribution function. If (14) does not exist, set $\mathtt{t}_0 = \sqrt{2\log(Nr)}$.

2. For each $(i, k) \in [N] \times [r]$, reject $H_0^{(i,k)}$ if $|\mathtt{T}_{ik}| \geq \mathtt{t}_0$. Finally $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(q)$ is formed by the whole rejected indexes, $\widehat{\mathcal{S}} = \{(i, k) \in [N] \times [r] : |\mathtt{T}_{ik}| \geq \mathtt{t}_0\}$.

Note that $R(\mathtt{t}_0) = |\widehat{\mathcal{S}}|$ by the definition. In the next section, we will see that the FDR of $\widehat{\mathcal{S}}$ is asymptotically controlled to be less than or equal to q. A similar procedure is found in Liu (2013) and Javanmard and Javadi (2019); they consider FDR control in a Gaussian graphical model and linear regression, respectively. The result for approximate factor models is new to the literature.

Finally we propose a new estimator based on "re-sparsification" of the debiased SOFAR estimator, using $\widehat{\mathcal{S}}$. That is, the *re-sparsified SOFAR estimator* is defined as

$$\widehat{\mathbf{B}}^{\mathsf{r}} = (\hat{b}_{ik}^{\mathsf{r}}) \quad \text{with} \quad \hat{b}_{ik}^{\mathsf{r}} = \hat{b}_{ik}^{\mathsf{d}} 1\{(i, k) \in \widehat{\mathcal{S}}\}. \tag{15}$$

The estimator is attractive in that the sparsity pattern controls the FDR over $(i, k) \in [N] \times [r]$ and that given $\widehat{\mathcal{S}}$ each nonzero component admits the asymptotic normality inherited from the debiased estimator. The consistency of this estimator is shown in the next section.

Remark 3. Procedure 1 works in principle with any other estimator that is asymptotically normal, such as the PC estimator, instead of the debiased SOFAR estimator $\hat{b}_{ik}^{\mathsf{d}}$ in (13). The associated re-sparsified estimator will be consistent as well.

4 Theory

We investigate the theoretical properties of the inferential framework proposed in Section 3. First we formally prove that the debiased SOFAR estimator and the PC estimator have asymptotic linear representations, implying asymptotic normality. Next we prove that $\widehat{\mathcal{S}}$ obtained by Procedure 1 controls the FDR and exhibits high power. Throughout this section, set $\eta_n \asymp T^{1/2} \log^{1/2}(N \vee T)$ in optimization (6).

The theory is developed on the basis of a sub-Gaussian assumption on the factors and errors. Following Rigollet and Hütter (2017), we introduce a sub-Gaussian random variable: a random variable $X \in \mathbb{R}$ is said to be sub-Gaussian with variance proxy σ^2 if $\mathbb{E}[X] = 0$ and its moment generating function satisfies $\mathbb{E}[\exp(sX)] \leq \exp(\sigma^2 s^2/2)$ for all $s \in \mathbb{R}$. This is denoted by $X \sim \text{subG}(\sigma^2)$. Define $L_n = (N \vee T)^\nu - 1$ for an arbitrary large constant $\nu > 0$. Throughout the paper, including all the proofs in the Appendix, ν is fixed.

Assumption 1 (Latent factors)**.** The factor matrix $\mathbf{F}^0 = (\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0)'$ is specified as the vector moving average process of order L_n (VMA(L_n)) such that

$$\mathbf{f}_t^0 = \sum_{\ell=0}^{L_n} \mathbf{\Psi}_\ell \boldsymbol{\zeta}_{t-\ell}, \qquad \lim_{n\to\infty} \sum_{\ell=0}^{L_n} \mathbf{\Psi}_\ell \mathbf{\Psi}_\ell' = \mathbf{I}_r,$$

where $\boldsymbol{\zeta}_t = (\zeta_{t1}, \ldots, \zeta_{tr})'$ with $\{\zeta_{tk}\}_{t,k}$ i.i.d. subG(σ_ζ^2) that has $\mathbb{E}\,\zeta_{tk}^2 = 1$, and $\mathbf{\Psi}_0$ is a nonsingular, lower triangular matrix.

Assumption 2 (Factor loadings)**.** Each column $\mathbf{b}_k^0$ of $\mathbf{B}^0$ has the sparsity $N_k = N^{\alpha_k}$ with $0 < \alpha_r \leq \cdots \leq \alpha_1 \leq 1$ and $\mathbf{B}^{0\prime}\mathbf{B}^0 = \text{diag}\{d_1^2 N_1, \ldots, d_r^2 N_r\}$ with $0 < d_r \leq \cdots \leq d_1 < \infty$. If $N_k = N_{k-1}$, there exists a constant $\delta > 0$ such that $d_{k-1}^2 - d_k^2 \geq \delta^{1/2} d_{k-1}^2$.

Assumption 3 (Idiosyncratic errors)**.** The error matrix $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_T)'$ is specified as the VMA(L_n) such that

$$\mathbf{e}_t = \sum_{\ell=0}^{L_n} \mathbf{\Phi}_\ell \boldsymbol{\varepsilon}_{t-\ell}, \quad \limsup_{n\to\infty} \sum_{\ell=0}^{L_n} \|\mathbf{\Phi}_\ell\|_2 < \infty,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tN})'$ with $\{\varepsilon_{ti}\}_{t,i}$ i.i.d. subG(σ_ε^2) and $\mathbf{\Phi}_0$ is a nonsingular, lower triangular matrix.

Assumption 4 (Parameter space)**.** The parameter space of $\mathbf{B}$ in optimization (6) is given by $\mathcal{B}(\widetilde{N}) = \{\mathbf{B} \in \mathbb{R}^{N\times r} : \|\mathbf{B}\|_0 \lesssim \widetilde{N}/2\}$ for $\widetilde{N} \in [N_1, N]$. (Define $\tilde{\alpha}$ to be such that $\widetilde{N} = N^{\tilde{\alpha}}$.)

Assumptions 1 and 3 specify the stochastic processes $\{\mathbf{f}_t\}$ and $\{\mathbf{e}_t\}$, respectively, to be stationary VMA(L_n), where $L_n \sim (N \vee T)^\nu$ diverges with a sufficiently large fixed constant $\nu > 0$. This construction is regarded as the *asymptotic linear process*, which includes a wide range of cross-sectional and time-series dependent processes. By Assumption 3, we have $\lambda_1(\mathbb{E}\,\mathbf{e}_t\mathbf{e}_t') < \infty$. Assumption 2 is key to our analysis and provides the sparse structure of the factor loadings $\mathbf{B}^0$ that leads to the WF models. The sparsity makes the divergence rate of $\lambda_k(\mathbf{B}^{0\prime}\mathbf{B}^0)$ possibly slower than N for each k. This can be called the *weak pervasiveness* condition, in contrast to the so-called pervasive condition of Fan et al. (2013), which assumes the SF structure $\lambda_k(\mathbf{B}^{0\prime}\mathbf{B}^0) \asymp N$ for every k.

Regarding Assumption 4, note that $\mathbf{B}^0$ is included in $\mathcal{B}(\widetilde{N})$ for any $\widetilde{N} \in [N_1, N]$ under Assumption 2. If $\widetilde{N}$ is set to N, $\mathcal{B}(N)$ coincides with the whole space, $\mathbb{R}^{T\times r}$. Whereas, if $\widetilde{N}$ is set to N_1, $\mathcal{B}(N_1)$ becomes as sparse as $\mathbf{B}^0$. The PC estimator always requires optimization in $\mathcal{B}(N)$ since it cannot be sparse, but the SOFAR estimator can allow sparse $\mathcal{B}(\widetilde{N})$ with $\widetilde{N} \in [N_1, N)$. An important consequence of permitting larger parameter space is that a wider class of the WF models can be consistently estimated; see the comments below and Uematsu and Yamagata (2020).

4.1 Theory on the asymptotic linear representation

Assume the following condition:

$$1 < \alpha_r + \tau. \tag{16}$$

Condition (16) guarantees divergence of λ_r. Under these conditions, the number of factors is correctly determined by the method of Onatski (2010). For more information, see Uematsu and Yamagata (2020). In what follows, suppose r is known. The theorems below show the asymptotic linear representation for the debiased SOFAR and PC estimators, respectively.

Theorem 1 (Debiased SOFAR)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$*. If Assumptions 1–4 with* (16) *and*

$$2\alpha_1 + \tilde{\alpha} \vee \tau < 2\alpha_r + 2(\alpha_r \wedge \tau) \tag{17}$$

hold, then the debiased SOFAR estimator has the asymptotic linear representation

$$\sqrt{T}\left(\widehat{\mathbf{b}}_i^{\mathsf{d}} - \mathbf{b}_i^0\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_{ti}\mathbf{f}_t^0 + \mathbf{r}_i, \tag{18}$$

where $\mathbf{r}_i$ *has the following bound with probability at least* $1 - O((N \vee T)^{-\nu})$*:*

$$\max_{i\in[N]} \|\mathbf{r}_i\|_{\max} \lesssim \frac{N_1^{3/2}\log(N \vee T)}{N_r(N_r \wedge T)} =: \delta_1.$$

The convergence of δ_1 *to zero is guaranteed under condition* (17).

Condition (17) is necessary to derive a nontrivial estimation error bound of the SOFAR estimator; see Uematsu and Yamagata (2020) for details. When we set $\tilde{\alpha} = \alpha_1$ in Assumption 4, condition (17) allows the widest class of $\{\alpha_1, \alpha_r\}$.

Theorem 2 (PC)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$*. If Assumptions 1–4 with* $\tilde{\alpha} = 1$*,* (16)*, and*

$$2\alpha_1 + 1 \vee \tau < 2\alpha_r + 2(\alpha_r \wedge \tau) \tag{19}$$

hold, then the PC estimator has an asymptotic linear representation

$$\sqrt{T}\left(\widehat{\mathbf{b}}_i^{\mathsf{PC}} - \mathbf{b}_i^0\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_{ti}\mathbf{f}_t^0 + \mathbf{r}_i^{\mathsf{PC}}, \tag{20}$$

where $\mathbf{r}_i^{\mathsf{PC}}$ *has the following bound with probability at least* $1 - O((N \vee T)^{-\nu})$*:*

$$\max_{i\in[N]} \|\mathbf{r}_i^{\mathsf{PC}}\|_{\max} \lesssim \delta_1\sqrt{\frac{N}{N_1}}.$$

The convergence of $\delta_1\sqrt{N/N_1}$ *to zero is guaranteed under condition* (19).

Remark 4. On condition $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$ a.s. in Theorems above (and below), it has been supposed only for technical simplicity and clear of presentation. In fact, this is not necessary to derive similar results since Assumption 1 guarantees $\mathbb{E}\,\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$ and the law of large numbers is applied. Without this condition, however, additional restrictions on $\{\alpha_1, \alpha_r\}$ will be required, which would render the results hereafter unnecessarily complicated. Indeed, this assumption is widely accepted in the literature on approximate factor models; see Bai and Ng (2013), Bai and Li (2014), and Ando and Bai (2017), among many others.

The upper bound of the estimation error $\mathbf{r}_i$ of the debiased SOFAR disappears faster than that of the PC estimator. Moreover, Condition (17) allows a wider class of $\{\alpha_1, \alpha_r\}$ than that implied by condition (19). In fact, the minimum value of α_r under (17) can achieve $1/3$ while (19) allows $\alpha_r > 1/2$. Even under condition (19) with $\alpha_1 < 1$, normal approximation of the debiased SOFAR estimator is expected to be more accurate than that of the PC estimator due to the behavior of the remainder terms. Hence, the finite sample normal approximation of the SOFAR estimator can be more accurate. This behavior is also confirmed by numerical simulations in Section 5. Of course a precise discussion requires a lower bound, but this is beyond the scope of this paper and is left for a future study.

In many cases, $T^{-1/2}\sum_{t=1}^{T} e_{ti}\mathbf{f}_t^0$ in (18) and (20) converges weakly to a normal distribution, $N(\mathbf{0}, \mathbf{\Gamma}_i)$, where $\mathbf{\Gamma}_i = \lim_{T\to\infty} T^{-1}\sum_{s,t=1}^{T}\mathbb{E}[\mathbf{f}_s^0\mathbf{f}_t^{0\prime}e_{si}e_{ti}]$, as shown in Bai (2003), for instance. The following subsection deals with such a case with simpler assumptions on $\{\mathbf{f}_t^0\}$ and $\{e_{ti}\}$.

4.2 Theory on the global inference for the loadings

Next we establish the theoretical results for the FDR control and power guarantee explored in Section 3.3. Although we focus on the case with the debiased SOFAR estimator here, we may establish a similar result with the PC estimator, as mentioned in Remark 3. We begin by strengthening the conditions.

Assumption 5. The factor matrix $\mathbf{F}^0 = (\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0)'$ is specified as i.i.d. vector process $\{\mathbf{f}_t^0\}$ with the elements f_{tk}^0 being subG(σ_ζ^2) and $\mathbb{E}\,\mathbf{f}_t^0 \mathbf{f}_t^{0\prime} = \mathbf{I}$. The error matrix $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_T)'$ is specified as i.i.d. vector process $\{\mathbf{e}_t\}$ with the elements e_{ti} being subG(σ_ε^2).

Assumption 6. There exist positive constants c, γ, and $\rho \in (0,1)$ and set $\Gamma \subset [N] \times [N]$ such that $|\Gamma| = O(N)$ and

$$|\mathrm{Corr}\,(e_{ti}, e_{tj})| \begin{cases} \in \big[0, c/\log^{2+\gamma}(Nr)\big) & \text{for } \ i \neq j \text{ and } (i,j) \in \Gamma^c, \\ \in \big[c/\log^{2+\gamma}(Nr), \rho\big] & \text{for } \ i \neq j \text{ and } (i,j) \in \Gamma, \\ = 1 & \text{for } \ i = j. \end{cases}$$

The independence of Assumption 5 is necessary for a technical reason. Assumption 6 permits moderate cross-sectional correlation among idiosyncratic errors. First we have the result of the FDR control of $\widehat{\mathcal{S}}$.

Theorem 3 (FDR control)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$. *If Assumptions 2 and 4–6 with* (16) *and* (17) *hold, then for any fixed* $q \in [0,1]$, *the FDR of* $\widehat{\mathcal{S}}$ *obtained by Procedure 1 is asymptotically controlled to be less than or equal to* q.

Next we derive the result of power analysis. For this purpose, it is common to suppose that the minimum signal does not decay too fast as N and T rise.

Assumption 7. For $\mathcal{S} = \mathrm{supp}(\mathbf{B}^0)$, the minimum signal is lower bounded as

$$\min_{(i,k)\in\mathcal{S}} |b_{ik}^0| \gtrsim \sqrt{\frac{2\log(Nr)}{T}}.$$

Theorem 4 (Power guarantee)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$. *If Assumptions 1–5 and 7 with* (16) *and* (17) *hold, and if* $s/N = o(1/\log N)$, *then the power of* $\widehat{\mathcal{S}}$ *obtained by Procedure 1 tends to unity.*

Theorems 3 and 4 have revealed that the factor selection procedure (Procedure 1) possesses statistically desirable properties. That is, the FDR of $\widehat{\mathcal{S}}$ will be asymptotically controlled less than or equal to pre-specified value $q \in [0,1]$, yet the power tends to unity. These properties are apparently inherited by the re-sparsified SOFAR estimator defined in (15). Moreover, it satisfies the following result:

Theorem 5 (Re-sparsified SOFAR)**.** *Suppose all the conditions in Theorems 3 and 4. If* $s^2/N = o(1/\log N)$, *then the re-sparsified estimator defined in* (15) *satisfies* $\|\widehat{\mathbf{B}}^{\mathrm{r}} - \mathbf{B}^0\|_{\max} \to_p 0$ *and* $\sqrt{T}(\hat{b}_{ik}^{\mathrm{r}} - b_{ik}^0) \to_d N(0, \sigma_i^2)$ *for any* $(i,k) \in \widehat{\mathcal{S}}$.

5 Monte Carlo Experiments

In this section we investigate the finite sample behavior of the debiased SOFAR estimator and the associated inferential procedure, comparing with those of the PC estimator by means of Monte Carlo experiments. First, we examine the quality of the standard normal approximation of t-statistics for the factor loadings. Next, we investigate the quality of the proposed FDR controlled global inferential procedure. Finally, we check the efficiency of the re-sparsified SOFAR and sparsified PC estimators.

We consider the following Data Generating Process (DGP):

$$x_{ti} = \sum_{k=1}^{r} b_{ik} f_{tk} + \sqrt{\theta} e_{ti}, \quad (t, i) \in [T] \times [N]. \tag{21}$$

The factor loadings b_{ik} and factors f_{tk} are formed such that $N^{-1} \sum_{i=1}^{N} b_{ik} b_{i\ell} = 1\{k = \ell\}$ and $T^{-1} \sum_{t=1}^{T} f_{tk} f_{t\ell} = 1\{k = \ell\}$, by applying Gram–Schmidt orthonormalization to b^*_{ik} and f^*_{tk}, respectively, which are constructed as follows. Non-zero factor loadings are computed as $b^*_{ik} = s_{ik} w_{ik}$, where s_{ik} is drawn from Rademacher distribution, $w_{ik} \sim U(\underline{b}, \bar{b})$, $\underline{b} = 0.103$ and $\bar{b}$ is chosen so that $\mathrm{Var}(b^*_{ik}) = 1$.[1] The first $N_k = \lfloor N^{\alpha_k} \rfloor$ elements of b^*_{ik} for $k = 1, 3, \ldots$ are non-zero, and the last N_k elements for $k = 2, 4, \ldots$ are non-zero. Let

$$f^*_{tk} = \rho_{fk} f^*_{t-1,k} + v_{tk} \tag{22}$$

for $t \in [T]$ and $k \in [r]$ with $v_{kt} \sim$ i.i.d.$N(0, 1 - \rho^2_{fk})$ and $f^*_{0k} \sim$ i.i.d.$N(0, 1)$. b_{ik} for $(i, k) \in [N] \times [r]$ are fixed over the replications. The idiosyncratic errors e_{ti} are generated by

$$e_{ti} = \rho_e e_{t-1,i} + \varepsilon_{ti}, \tag{23}$$

where $\varepsilon_{ti} \sim$ i.i.d.$N(0, 1 - \rho^2_e)$.

For all the experiments we set $r = 2$ and $\theta = 0.5$. We examine the performance of the proposed methods across different values of exponents $\{\alpha_1, \alpha_2\}$. In particular, we consider the combinations $\{0.9, 0.8\}$, $\{0.7, 0.6\}$, and $\{0.5, 0.4\}$ with $T, N \in \{100, 200, 500\}$.

We consider three different t-statistics for the inference on each factor loading and the proposed FDR controlled multiple testing procedure. First, a t-statistic which is the ratio of $\hat{b}_{ik}$ and its population standard deviation, denoted by (dropping the subscripts i and k for simplicity) $\mathtt{T}_0$. The other two are $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$, which are the t-statistics based on $\widehat{\mathbf{\Gamma}}_0$ and $\widehat{\mathbf{\Gamma}}$, respectively. To economize the space in what follows we report the results for the DGP with i.i.d. factors and i.i.d. errors only (by setting $\rho_{fk} = \rho_\varepsilon = 0$ for all $k \in [r]$). The results for serially correlated cases with $\mathtt{T}_{NW}$ are qualitatively similar, and are reported in the Online Appendix.

5.1 Normal approximation of t-statistics

We examine the quality of the normal approximation of the various t-statistics defined above. To evaluate the theoretical results in the earlier sections, we first inspect the distribution of $\hat{b}_{ik}$ for null $(i, k) \in \mathcal{S}^c$, scaled by its true standard deviation, $\mathtt{T}_0$, and refer to $N(0, 1)$, so that the assessment is exempted from the quality of the estimation of the variance of $\hat{b}_{ik}$. For the same purpose, we employ i.i.d. factors and errors, by setting $\rho_{fk} = \rho_e = 0$ for all $k \in [r]$.

Figures 1–6 report the Q-Q plots of $\mathtt{T}_0$ against $N(0, 1)$. The plots are based on 40,000 replications for the sample size $N = T = 100$. The left column shows the Q-Q plots of the debiased SOFAR estimator, and the right column shows the Q-Q plots of the PC estimator.

[1] The value of $\underline{b}$ is chosen using $g\sqrt{2\log(Nr)/T}$ with $g = 1$, $N = 100$, $r = 2$ an $T = 1000$.

As can be seen, when the factors are relatively strong, with $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$, both $\mathtt{T}_0$ based on the debiased SOFAR and PC estimators are virtually standard normally distributed. However, the distribution of $\mathtt{T}_0$ using the PC estimator deviates from the standard normal further as the factor loadings become weaker, while that of the debiased SOFAR estimator remains standard normally distributed, as weak as $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$. This supports our earlier theoretical results in Theorems 1 and 2. Qualitatively similar results are obtained with $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$, which are summarized in Online Appendix.

5.2 The global inference for the loadings

Given the high quality normal approximation of the debiased SOFAR estimator, we are ready to investigate the finite sample properties of the proposed procedure for *global* inference. Recall that our interest is in testing whether *each* factor loading is zero or not, by controlling the FDR to be less than or equal to a predetermined level, $q \in [0, 1]$, while achieving high power.

In this set of experiments, q is fixed at 10%. We employ the DGP with i.i.d. factors and errors as before. To assess the efficacy of the proposed method to control the FDR, we report the FDR as well as the power, based on $\mathtt{T}_{iid}$. The corresponding results based on $\mathtt{T}_0$ and $\mathtt{T}_{NW}$ are qualitatively similar, which are available in the the Online Appendix. All the combinations of $N, T \in \{100, 200, 500\}$ are considered. All the results are based on 1000 replications. Three models with different exponents, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$, are examined.

The FDR and the power of the proposed procedure are represented as surface plots in Figures 7–12. The left column shows the FDR, and the right column shows the power. The results of the debiased SOFAR estimator are shown by the pink surface, and those of the PC estimator are reported by the blue surface. It is apparent that the proposed procedure based on the debiased SOFAR estimator successfully controls the FDR for all the models by keeping it less than or equal to $q = 0.1$ with sufficiently large T, whereas that based on the PC estimator deviates from the pre-assigned level as the model becomes weaker. Their power properties are very similar. Given the model, the power quickly rises towards unity as T increases. In general, it is less powerful for the models with weaker factors, since the overall signal-to-noise ratio becomes weaker in our design.

[INSERT Figures 1–6]

[INSERT Figures 7–12]

5.3 Re-sparsified SOFAR and sparsified PC estimators

We have seen that the proposed procedure successfully controls the FDR to be less than or equal to pre-specified level q, while achieving high power. With this encouraging result, we also examine the efficacy of the re-sparsified SOFAR estimator, along with other relevant estimators. In particular we consider the *sparsified* PC estimator,

$$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}} = (\hat{b}^{\mathsf{r}}_{ik}) \;\text{ with }\; \hat{b}^{\mathsf{r}}_{ik} = \hat{b}^{\mathsf{PC}}_{ik} 1\{(i,k) \in \widehat{\mathcal{S}}^{\mathsf{PC}}\},$$

where $\widehat{\mathcal{S}}^{\mathsf{PC}}$ is obtained by Procedure 1 with using $\mathtt{T}_{iid}$ constructed using the PC estimator. We employ the same DGP and set-up used for Figures 7–12 and compare the norm loss $\|N_1^{-1/2} \sum_{k=1}^{r} \{\text{abs}(\widehat{\mathbf{b}}_k) - \text{abs}(\mathbf{b}_k^0)\}\|$. Observe that this norm loss is immune to the consequences of SOFAR and PC estimators being up to rotation (i.e., sign indeterminacy and changes to the order of the factor components).

In Table 1, we report the norm loss of the re-sparsified debiased SOFAR estimator ($\widehat{\mathbf{B}}^{\mathsf{r}}$) and the sparsified PC estimator ($\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$), along with the SOFAR ($\widehat{\mathbf{B}}$), debiased SOFAR ($\widehat{\mathbf{B}}^{\mathsf{d}}$), and the PC estimator ($\widehat{\mathbf{B}}^{\mathsf{PC}}$). As can be seen, the proposed re-sparsified debiased SOFAR estimator performs best, followed by the sparsified PC estimator and the SOFAR estimator. In view of the popularity of the PC estimator, this is a very encouraging result. The debiased SOFAR estimator dominates the PC estimator in terms of the norm loss.

6 Empirical Applications

In this section we consider the empirical applications of the FDR controlled global inference on the factor selection. We extract factors by the SOFAR method from a large number of macroeconomic (prediction) variables, in line with the analyses of Ludvigson and Ng (2009) and McCracken and Ng (2016). The proposed global inferential procedure permits us to statistically analyze the information content of common factors in each variable.

Specifically, the FRED-MD macroeconomic and financial data file of May 2019 is obtained from McCracken's website and the variables are transformed as instructed by McCracken and Ng (2016). The data consists of a balanced panel of 128 monthly series spanning the period from June 1999 to May 2019. All series are standardized before the analysis. Following McCracken and Ng (2016), the series are categorised into eight groups (note that the group order is different from McCracken and Ng (2016)): **G1**. Output and Income; **G2**. Labour Market; **G3**. Consumption, Orders and Inventories; **G4**. Housing; **G5**. Interest and Exchange Rate; **G6**. Prices; **G7**. Money and Credit; **G8**. Stock Market.

The number of factors is estimated by the ED method of Onatski (2010), which suggests it most probably contains five factors. Given the number of factors, the re-sparsified SOFAR estimate is computed. The t-statistics for the procedure are computed using the serial correlation robust variance covariance estimator, $\mathtt{T}_{NW}$. We report the result for $q = 10\%$.

To assess the contribution of each of the 128 series to these five common factors, we report the value of factor loadings of each of the 128 series as a bar-chart in Figure 13. The variables are ordered by its eight groups. Note that the larger the absolute values of the factor loading, the higher the influence of the associated common factor to the variable. Just casting a glance at Figure 13 gives very strong evidence of sparse factor loadings under the identification restrictions and exhibits a clear association of factors (loadings) and groups of macroeconomic variables. The first factor is associated with five variable groups, G1-G5, and can be seen as a semi-global factor. Each of the remaining four factors is associated with just one or two dominating groups. Specifically, we may identify the second to the fifth factor as a price factor, housing factor, output and income factor, and a money, credit and stock market factor, respectively. [**INSERT Figure 13**]

7 Conclusion

In this paper, we have considered statistical inference for high-dimensional approximate factor models. We have supposed the weak factor (WF) structure, in which the factor loading matrix can be sparse and the signal eigenvalues may diverge more slowly than the cross-sectional dimension, N. The central theme of this paper is the global inference for factor selection, specifically whether each element of the factor loadings is zero or not, which is new in the literature. Initially we have proposed the *debiased* version of the SOFAR estimator (see Uematsu and Yamagata, 2020) of the sparse loadings in the WF models, and established its asymptotic normality. In addition, we have shown that the PC estimator is asymptotically normal even for the WF models. Building upon the asymptotic normality of the factor loading estimators, we have proposed a procedure in the multiple testing framework to decide whether each of the factor loadings is significantly zero or not, and have proved that this controls the false discovery rate (FDR) below a pre-assigned level, while the power tends to unity. Although the theory is established for the debiased SOFAR estimator, the method works with any asymptotically normal estimators, such as the PC estimator; whereas the latter can be less efficient as it cannot effectively utilize the sparseness of the loadings. Furthermore, we have proposed a new estimator of the factor loading matrix called the *re-sparsified* SOFAR estimator, which is defined as the debiased SOFAR estimator, with its insignificant elements being replaced with zeros. Similarly, we have proposed a *sparsified* PC estimator, which is obtained after the global inference based on the PC estimator in the same manner. We have also established its consistency. The finite sample performance has revealed that these estimators are superior to the SOFAR, the debiased SOFAR and the PC estimators in terms of the norm loss.

We also provide a coherent estimation-inference procedure for high-dimensional approximate factor models. Since the proposed method can be based upon any asymptotically normal estimator, such as the PC estimator, its applicability is very wide. The empirical application has provided firm statistical evidence of sparse factor loadings, which suggests that our approach can shed light on uncovered features in the factor models of macroeconomic data, as analyzed by Stock and Watson (2002), Ludvigson and Ng (2009), and McCracken and Ng (2016), among many others. In the recent finance literature, there have been increasing interest in selection of factors in high-dimensional environments; see Feng et al. (2019) and Kozak et al. (2020), for example. The proposed methods are well suited to address such issues.

Table 1: Norm Loss (×1000) of SOFAR ($\widehat{\mathbf{B}}$), debiased-SOFAR ($\widehat{\mathbf{B}}^{\mathsf{d}}$), PC ($\widehat{\mathbf{B}}^{\mathsf{PC}}$), re-sparsified SOFAR ($\widehat{\mathbf{B}}^{\mathsf{r}}$) and sparsified PC ($\widehat{\mathbf{B}}^{\mathsf{d}}_{\mathsf{PC}}$) estimators.

	$\{\alpha_1, \alpha_2\}$	$\{0.9, 0.8\}$			$\{0.7, 0.6\}$			$\{0.5, 0.4\}$		
	Est.$\backslash N$	100	200	500	100	200	500	100	200	500
$T = 100$										
	$\widehat{\mathbf{B}}$	160.0	167.2	173.6	200.3	222.0	232.7	207.8	217.7	236.9
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	149.9	156.4	165.2	248.5	280.4	321.5	404.5	482.1	606.9
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	189.6	166.1	166.5	270.7	308.3	327.1	459.7	526.1	636.6
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	137.1	138.4	136.5	153.5	157.3	159.3	189.3	183.3	180.1
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	180.0	150.1	139.4	178.4	193.8	166.2	230.0	211.9	203.0
$T = 200$										
	$\widehat{\mathbf{B}}$	116.0	120.5	124.6	140.5	153.0	164.5	146.3	154.3	167.5
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	106.8	112.6	117.3	177.8	200.1	227.6	291.6	343.2	430.5
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	132.4	116.2	117.5	191.0	213.9	230.7	329.9	374.7	450.7
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	95.3	97.0	95.5	106.6	107.7	107.4	132.6	125.2	123.0
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	123.1	101.4	96.5	120.7	125.4	110.5	161.5	144.8	135.7
$T = 500$										
	$\widehat{\mathbf{B}}$	71.7	78.1	81.4	85.3	95.7	100.6	91.1	96.7	100.9
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	69.7	71.3	74.9	114.5	126.8	144.4	191.6	221.2	273.2
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	80.3	72.6	75.0	122.0	133.2	146.3	216.6	241.1	286.1
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	59.8	59.8	59.1	65.1	65.0	64.8	89.0	80.3	74.0
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	71.7	61.4	59.6	73.1	73.1	66.5	109.7	93.8	81.7

Notes: For the re-sparsified estimator, the target FDR level is set $q = 0.1$.

Figures 1–6 show the Q-Q plot of the distribution of a t-statistic based on the debiased SOFAR estimator and the PC estimator against $N(0,1)$ for the models with $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$, $\{0.7, 0.6\}$, $\{0.5, 0.4\}$.

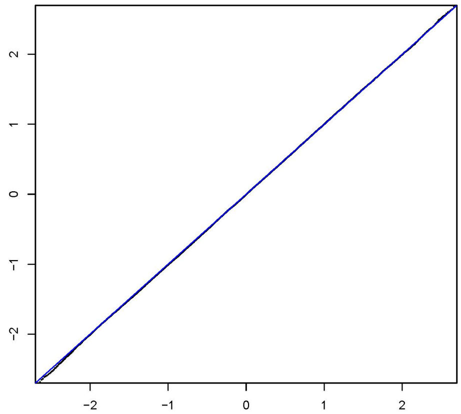

Figure 1: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

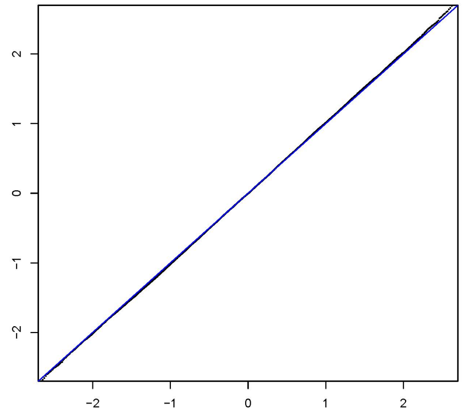

Figure 2: PC, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

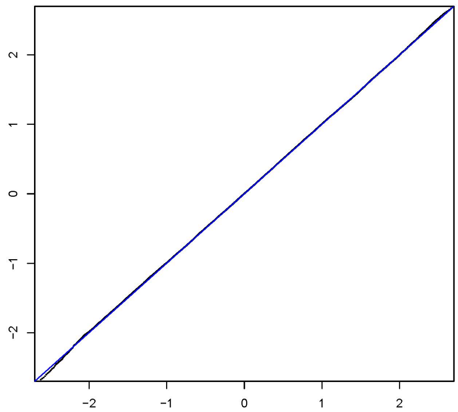

Figure 3: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

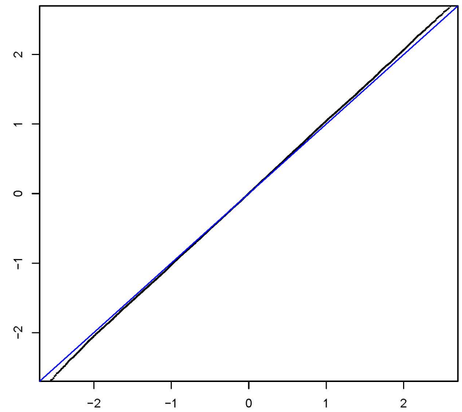

Figure 4: PC, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

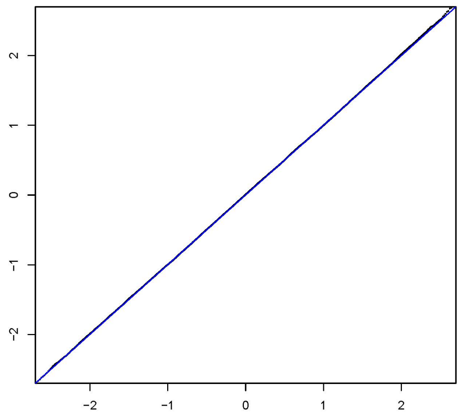

Figure 5: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

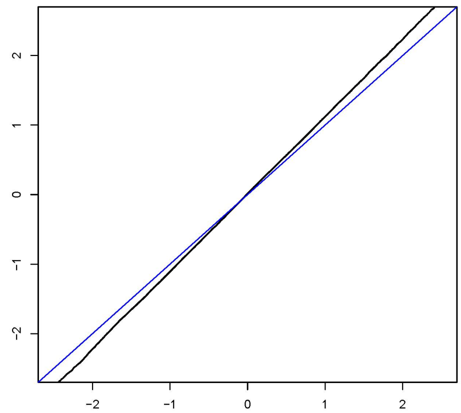

Figure 6: PC, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

Figures 7–12 show the FDR and power with $q = 0.1$ for the models with $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$, $\{0.7, 0.6\}$, $\{0.5, 0.4\}$.

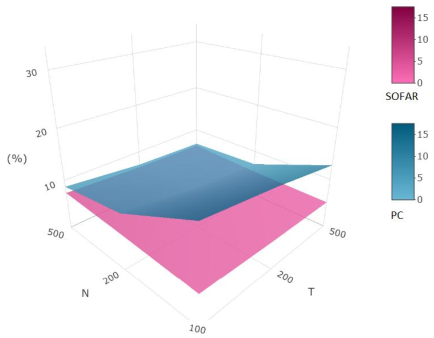

Figure 7: FDR, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$ with $q = 0.1$

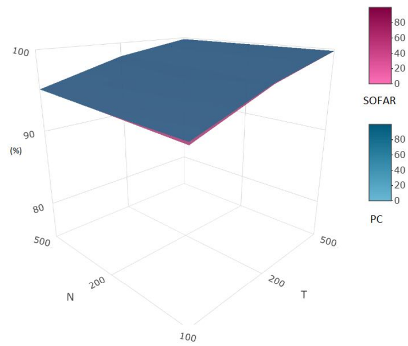

Figure 8: Power, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

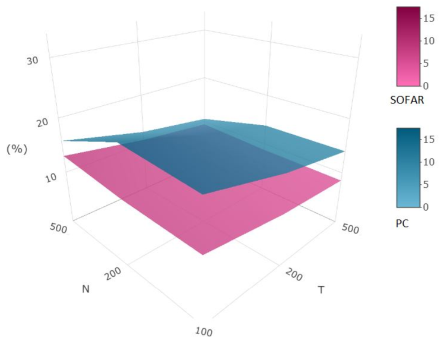

Figure 9: FDR, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$ with $q = 0.1$

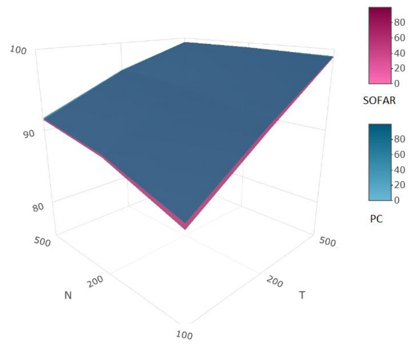

Figure 10: Power, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

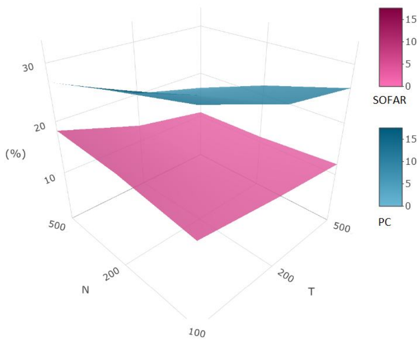

Figure 11: FDR, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$ with $q = 0.1$

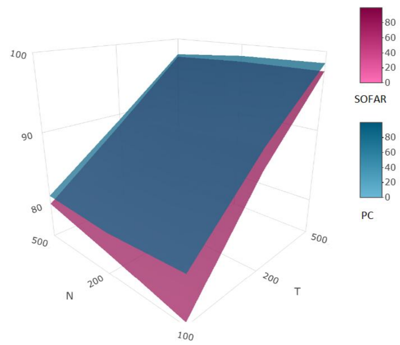

Figure 12: Power, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

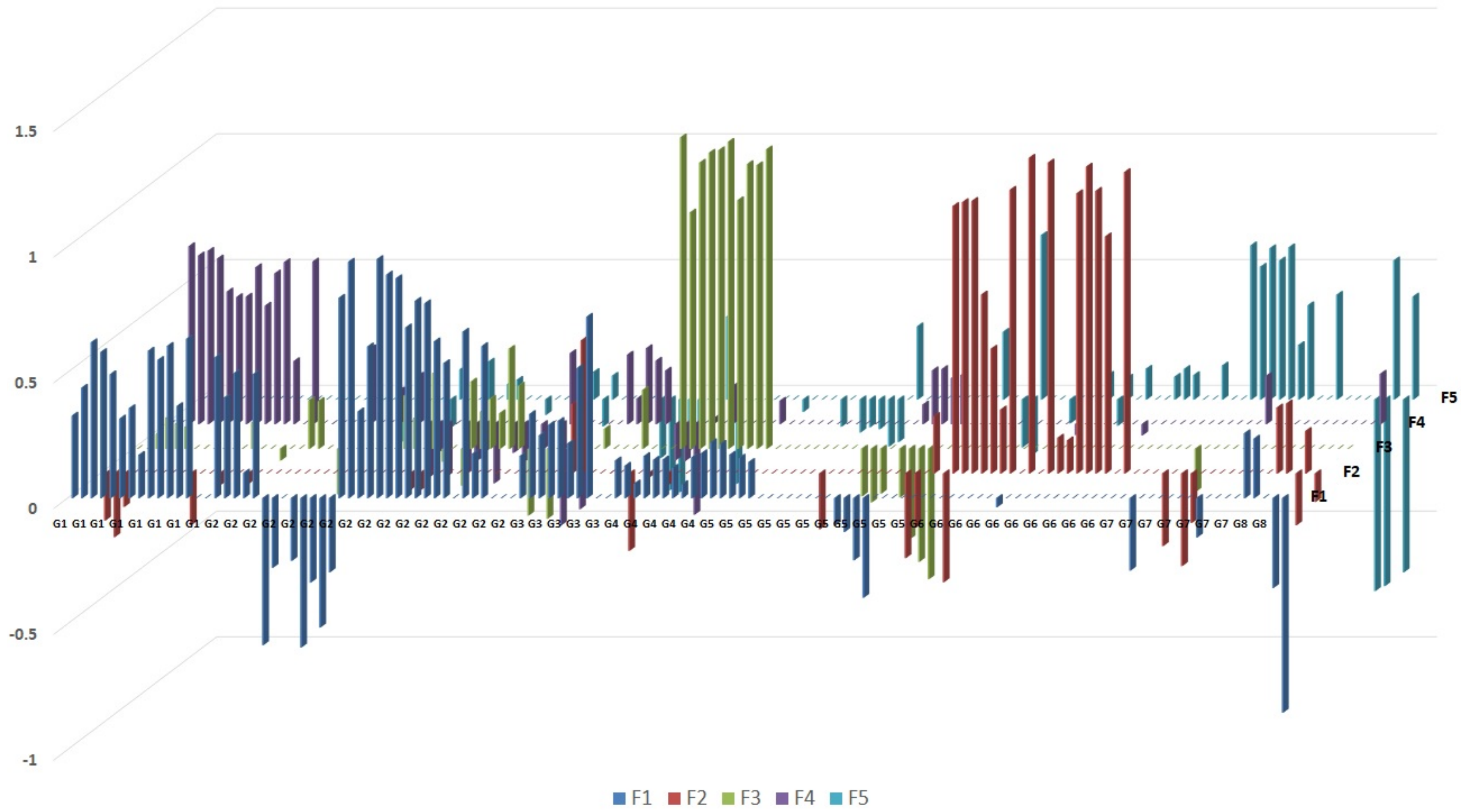

Figure 13: Bar-chart of the factor loadings estimates for each of 128 variables with the target FDR level 0.1

Appendix

A Proofs of the Main Results

We first fix a finite number $\nu > 0$ and use it throughout all the proofs. Since the choice is arbitrary and ν can always be replaced by a larger one at the first stage, we may write $N^a T^b O((N \vee T)^{-\nu}) = O((N \vee T)^{-\nu})$ with abuse of notation even for positive (but finite) numbers a and b, unless a precise order is required.

A.1 Proof of Theorem 1

Proof. Define $\widehat{\mathbf{\Delta}} = \widehat{\mathbf{F}} - \mathbf{F}^0$ and $\mathcal{F} = \left\{\mathbf{\Delta} \in \mathbb{R}^{T\times r} : \|\mathbf{\Delta}\|_{\mathrm{F}} \leq C r_n\right\}$, where C is some positive constant and

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

Then under the assumed conditions, $\widehat{\mathbf{\Delta}} \in \mathcal{F}$ holds with probability at least $1 - O((N \vee T)^{-\nu})$ by Uematsu and Yamagata (2020). By the definition of the debiased SOFAR estimator, we have the decomposition

$$T^{1/2}(\widehat{\mathbf{B}}^* - \mathbf{B}^0) = \mathbf{Z} + \mathbf{R}^{(1)}(\widehat{\mathbf{\Delta}}) + \mathbf{R}^{(2)}(\widehat{\mathbf{\Delta}}), \tag{A.1}$$

where $\mathbf{Z} = T^{-1/2}\mathbf{E}'\mathbf{F}^0$, $\mathbf{R}^{(1)}(\widehat{\mathbf{\Delta}}) = T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}\widehat{\mathbf{\Delta}}$, and $\mathbf{R}^{(2)}(\widehat{\mathbf{\Delta}}) = T^{-1/2}\mathbf{E}'\widehat{\mathbf{\Delta}}$. Therefore, to obtain the asymptotic linear representation, it is enough to show that $\mathbf{R}^{(1)}(\widehat{\mathbf{\Delta}})$ and $\mathbf{R}^{(2)}(\widehat{\mathbf{\Delta}})$ are negligible in the max-norm. From the proof of Lemma 9 in Uematsu and Yamagata (2020), the first term is bounded as

$$\begin{aligned}\|\mathbf{R}^{(1)}(\widehat{\mathbf{\Delta}})\|_{\max} &\leq \sup_{\mathbf{\Delta}\in\mathcal{F}} \|T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}\mathbf{\Delta}\|_{\max} \\ &\leq r\|\mathbf{B}^0\|_{\max} \sup_{\mathbf{\Delta}\in\mathcal{F}} \|T^{-1/2}\mathbf{F}^{0\prime}\mathbf{\Delta}\|_{\max} \lesssim T^{-1/2} r_n \log^{1/2}(N \vee T) = \delta_1\end{aligned}$$

with probability at least $1 - O((N \vee T)^{-\nu})$. Similarly, the second term is bounded as

$$\|\mathbf{R}^{(2)}(\widehat{\mathbf{\Delta}})\|_{\max} \leq \sup_{\mathbf{\Delta}\in\mathcal{F}} \|T^{-1/2}\mathbf{E}'\mathbf{\Delta}\|_{\max} \lesssim T^{-1/2} r_n \log^{1/2}(N \vee T) = \delta_1$$

with probability at least $1 - O((N \vee T)^{-\nu})$. Thus the desired upper bound is obtained in view of the triangle inequality. Its convergence is easily verified by condition 17. This completes the proof. □

A.2 Proof of Theorem 2

Proof. The proof is basically the same as that of Theorem 1 except for the convergence rate r_n replaced by r_n^{PC} for the PC estimator. Let $\widehat{\mathbf{\Delta}}^{PC} = \widehat{\mathbf{F}}^{PC} - \mathbf{F}^0$ and define $\mathcal{F}_{PC} =$

$\{\mathbf{\Delta} \in \mathbb{R}^{T\times r} : \|\mathbf{\Delta}\|_{\mathrm{F}} \le C r_n^{PC}\}$, where C is some positive constant and

$$r_n^{PC} = \frac{N_1 N^{1/2} T^{1/2} \log^{1/2}(N \vee T)}{N_r(N_r \wedge T)}.$$

Then under the assumed conditions, $\widehat{\mathbf{\Delta}}^{PC} \in \mathcal{F}$ holds with probability at least $1 - O((N \vee T)^{-\nu})$ by Uematsu and Yamagata (2020). By the definition of the PC estimator, we have the decomposition

$$T^{1/2}(\widehat{\mathbf{B}}^{PC} - \mathbf{B}^0) = \mathbf{Z} + \mathbf{R}_{PC}^{(1)}(\widehat{\mathbf{\Delta}}) + \mathbf{R}_{PC}^{(2)}(\widehat{\mathbf{\Delta}}), \tag{A.2}$$

where $\mathbf{Z} = T^{-1/2}\mathbf{E}'\mathbf{F}^0$, $\mathbf{R}_{PC}^{(1)}(\widehat{\mathbf{\Delta}}_{PC}) = T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}\widehat{\mathbf{\Delta}}_{PC}$, and $\mathbf{R}_{PC}^{(2)}(\widehat{\mathbf{\Delta}}_{PC}) = T^{-1/2}\mathbf{E}'\widehat{\mathbf{\Delta}}_{PC}$. The rest of the proof is the same as the proof of Theorem 1 and is omitted. □

A.3 Proof of Theorem 3

Proof. Let $G(\mathtt{t}) = 2(1-\Phi(\mathtt{t}))$. Consider two cases; Case 1 deals with the case when (14) does not exist and $\mathtt{t}_0 = (2\log N)^{1/2}$, and Case 2 when $\mathtt{t}_0$ is given by (14). Write $Z_{ik}^* := Z_{ik}/\sigma_i$ and $e_{ti}^* = e_{ti}/\sigma_i$, where $Z_{ik} = T^{-1/2}\sum_{t=1}^T e_{ti} f_{tk}^0$.

Case 1. The FDR is defined as

$$\mathrm{FDR}(\mathtt{t}_0) = \mathbb{E}\,\mathrm{FDP}(\mathtt{t}_0) = \mathbb{E}\left[\frac{\sum_{(i,k)\in\mathcal{S}^c} 1\{|\mathtt{T}_{ik}| \ge \mathtt{t}_0\}}{R(\mathtt{t}_0) \vee 1}\right].$$

Set $\delta \asymp \delta_1 \log^{1/2}(N \vee T)$, where δ_1 has been defined in Theorem 1. In view of the law of iterated expectations, $\mathrm{FDR}(\mathtt{t}_0)$ is bounded by the probability that at least one variable is falsely discovered. Thus, using the notation in the proof of Lemma 5 together with the law of total probability and union bound, we have

$$\begin{aligned}
\mathrm{FDR}(\mathtt{t}_0) &\le \mathbb{P}\left(\sum_{(i,k)\in\mathcal{S}^c} 1\{|\mathtt{T}_{ik}| \ge \mathtt{t}_0\} \ge 1\right) \le \mathbb{P}\left(\sum_{(i,k)\in\mathcal{S}^c} 1\{|Z_{ik}^*| + |W_{ik}| \ge \mathtt{t}_0\} \ge 1\right)\\
&\le \mathbb{P}\left(\sum_{(i,k)\in\mathcal{S}^c} 1\{|Z_{ik}^*| \ge \mathtt{t}_0 - \delta\} \ge 1\right) + \mathbb{P}\left(\max_{(i,k)\in\mathcal{S}^c} |W_{ik}| > \delta\right)\\
&\le Nr \max_{(i,k)\in\mathcal{S}\cup\mathcal{S}^c} \mathbb{P}\left(|Z_{ik}^*| \ge \mathtt{t}_0 - \delta\right) + |\mathcal{S}^c| \max_{(i,k)\in\mathcal{S}^c} \mathbb{P}\left(|W_{ik}| > \delta\right).
\end{aligned}$$

Because δ_1 converges to zero polynomially under the assumed conditions, we have $\delta = o(\mathtt{t}_0)$, where $\mathtt{t}_0 = (2\log Nr)^{1/2}$. Thus the last two terms tend to zero by Lemma 5. This entails the asymptotic FDR control for any predetermined level $q \in [0, 1]$.

Case 2. Consider the case when $\mathtt{t}_0$ is given by (14). Define

$$A = \sup_{\mathtt{t}\in[0,\bar{\mathtt{t}}]} \left|\frac{\sum_{(i,k)\in\mathcal{S}^c}[1\{|\mathtt{T}_{ik}| \ge \mathtt{t}\} - G(\mathtt{t})]}{NrG(\mathtt{t})}\right|.$$

Then the FDP computed with threshold $\mathtt{t}_0$ is bounded as

$$\begin{aligned}\mathrm{FDP}(\mathtt{t}_0) &= \frac{\sum_{(i,k)\in\mathcal{S}^c}[1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_0\} - G(\mathtt{t}_0)] + |\mathcal{S}^c|G(\mathtt{t}_0)}{R(\mathtt{t}_0)\vee 1}\\ &\leq \frac{NrG(\mathtt{t}_0)A + NrG(\mathtt{t}_0)}{R(\mathtt{t}_0)\vee 1} \leq q(1+A),\end{aligned}$$

where the last inequality holds by (14). Taking the expectation, we have $\mathrm{FDR}(\mathtt{t}_0)\leq q\,\mathbb{E}[1+A]$. Therefore, it is sufficient to show $A = o_p(1)$ because this entails $\mathbb{E}[A] = o(1)$ by the reverse Fatou lemma and the result follows.

In order to show $A = o_p(1)$, we consider discretization of A. That is, we partition $[0,\bar{\mathtt{t}}]$ into small intervals, $0 = \mathtt{t}_0 < \mathtt{t}_1 < \cdots < \mathtt{t}_b = \bar{\mathtt{t}} = (2\log(Nr) - a\log\log(Nr))^{1/2}$, such that $\mathtt{t}_m - \mathtt{t}_{m-1} = v_N$ for $m\in\{1,\ldots,b-1\}$ and $\mathtt{t}_b - \mathtt{t}_{b-1}\leq v_N$, where $v_N = (\log\log(Nr))^{-1}$. Note that $b \asymp \bar{\mathtt{t}}/v_N \asymp \log^{1/2}(Nr)\log\log(Nr)$. Fix $m\in\{1,\ldots,b\}$ arbitrary. For any $\mathtt{t}\in[\mathtt{t}_{m-1},\mathtt{t}_m]$, we have

$$\frac{\sum_{(i,k)\in\mathcal{S}^c}1\{|\mathtt{T}_{ik}|\geq \mathtt{t}\}}{NrG(\mathtt{t})} \leq \frac{\sum_{(i,k)\in\mathcal{S}^c}1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_{m-1}\}}{NrG(\mathtt{t}_{m-1})}\cdot\frac{G(\mathtt{t}_{m-1})}{G(\mathtt{t}_m)} \tag{A.3}$$

and

$$\frac{\sum_{(i,k)\in\mathcal{S}^c}1\{|\mathtt{T}_{ik}|\geq \mathtt{t}\}}{NrG(\mathtt{t})} \geq \frac{\sum_{(i,k)\in\mathcal{S}^c}1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_{m}\}}{NrG(\mathtt{t}_{m})}\cdot\frac{G(\mathtt{t}_{m})}{G(\mathtt{t}_{m-1})}. \tag{A.4}$$

Because of (A.3), (A.4), and the fact that $G(\mathtt{t}_{m-1})/G(\mathtt{t}_m) = 1 + o(1)$ uniformly in $m\in\{1,\ldots,b\}$, the proof completes if the following is verified:

$$A^* := \max_{m\in\{1,\ldots,b\}}\left|\frac{\sum_{(i,k)\in\mathcal{S}^c}[1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_m\} - G(\mathtt{t}_m)]}{NrG(\mathtt{t}_m)}\right| = o_p(1). \tag{A.5}$$

Fix $\varepsilon > 0$ arbitrary. The union bound and Chebyshev's inequality yield

$$\begin{aligned}\mathbb{P}(A^* > \varepsilon) &\leq b\max_{m\in\{1,\ldots,b\}}\mathbb{P}\left(\left|\frac{\sum_{(i,k)\in\mathcal{S}^c}[1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_m\} - G(\mathtt{t}_m)]}{NrG(\mathtt{t}_m)}\right| > \varepsilon\right)\\ &\lesssim \ell_N\max_{m\in\{1,\ldots,b\}}\mathbb{E}\left[\left|\frac{\sum_{(i,k)\in\mathcal{S}^c}[1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_m\} - G(\mathtt{t}_m)]}{NrG(\mathtt{t}_m)}\right|^2\right]/\varepsilon^2,\end{aligned}$$

where $\ell_N = \log^{1/2}(Nr)\log\log(Nr)$. Expanding the expectation and collecting terms with using Lemma 5, we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\left[1\{|\mathtt{T}_{ik}|\geq \mathtt{t}_m\}-G(\mathtt{t}_m)\right]\left[1\{|\mathtt{T}_{j\ell}|\geq \mathtt{t}_m\}-G(\mathtt{t}_m)\right]}{N^2r^2G(\mathtt{t}_m)^2}\right]\\
&\leq \frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\mathbb{P}\left(|\mathtt{T}_{ik}|\geq \mathtt{t}_m, |\mathtt{T}_{j\ell}|\geq \mathtt{t}_m\right)\\
&\quad -\frac{2}{NrG(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\mathbb{P}\left(|\mathtt{T}_{ik}|\geq \mathtt{t}_m\right)+1\\
&\leq \frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)+\frac{O((N\vee T)^{-\nu})}{G(\mathtt{t}_m)^2}\\
&\quad -\frac{2}{NrG(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)+\frac{O((N\vee T)^{-\nu})}{G(\mathtt{t}_m)}+1. \qquad (A.6)
\end{aligned}
$$

We evaluate each term to conclude that the upper bound of (A.6) is $o(\log^{-1} N)$. First consider the second and fourth terms of (A.6). Note that

$$G(\mathtt{t}_m)>\frac{2\phi(\mathtt{t}_m)}{\mathtt{t}_m+1/\mathtt{t}_m}\asymp\frac{e^{-\mathtt{t}_m^2/2}}{\mathtt{t}_m+1/\mathtt{t}_m}\gtrsim\frac{e^{-\log(Nr)+(a/2)\log\log(Nr)}}{\log^{1/2}(Nr)}=\frac{\log^{a/2-1/2}(Nr)}{Nr}$$

uniformly in $m\in\{1,\dots,b\}$. Thus we have

$$\frac{O((N\vee T)^{-\nu})}{G(\mathtt{t}_m)^2}+\frac{O((N\vee T)^{-\nu})}{G(\mathtt{t}_m)}\lesssim\frac{N^2r^2}{\log^{a-1}(Nr)}O((N\vee T)^{-\nu})=O((N\vee T)^{-\nu})$$

uniformly in $m\in\{1,\dots,b\}$. Next consider the third term of (A.6). By the triangle inequality, we have

$$
\begin{aligned}
&-\frac{2}{NrG(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)\\
&=-\frac{2G(\mathtt{t}_m+\delta)}{NrG(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\left(\frac{\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)}{G(\mathtt{t}_m+\delta)}-1\right)-\frac{2|\mathcal{S}^c|G(\mathtt{t}_m+\delta)}{NrG(\mathtt{t}_m)}\\
&\leq\frac{2G(\mathtt{t}_m+\delta)}{G(\mathtt{t}_m)}\max_{(i,k)\in\mathcal{S}^c}\left|\frac{\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)}{G(\mathtt{t}_m+\delta)}-1\right|-\frac{2G(\mathtt{t}_m+\delta)}{G(\mathtt{t}_m)}.
\end{aligned}
$$

Lemma 7.2 of Javanmard and Javadi (2019) gives

$$\frac{G(\mathtt{t}_m+\delta)}{G(\mathtt{t}_m)}\leq 1+8(\delta+\mathtt{t}_m\delta)=1+O(\mathtt{t}_m\delta)=1+o\left(\ell_N\right),$$

where the last equality holds since δ polynomially decreases while $\mathtt{t}_m$ is a logarithmic function. Lemma 6.1 of Liu (2013) yields

$$\max_{(i,k)\in\mathcal{S}^c}\left|\frac{\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)}{G(\mathtt{t}_m+\delta)}-1\right|\lesssim\log^{-3/2}(Nr).$$

Consequently, the third term of (A.6) is bounded as

$$-\frac{2}{NrG(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\mathbb{P}(|Z^*_{ik}|\geq \mathtt{t}_m+\delta)\lesssim \log^{-3/2}(Nr)-2+o\left(\ell_N\right)=-2+o\left(\ell_N\right).$$

Finally we consider the first term of (A.6). In order to tightly bound the joint probability, we divide the summand into two parts based on the strength of their correlations. Recall that $Z^*_{ik}=T^{-1/2}\sum_{t=1}^T e^*_{ti}f^0_{tk}$ with $\mathbb{E}\, {e^*_{ti}}^2=\mathbb{E}\, e^2_{ti}/\sigma_i^2=1$ and $\mathbb{E}\, f^0_{tk}f^0_{t\ell}=1\{k=\ell\}$. We have

$$\mathbb{E}\left[e^*_{ti}e^*_{tj}f^0_{tk}f^0_{t\ell}\right]=\mathbb{E}\left[e^*_{ti}e^*_{tj}\right]1\{k=\ell\},$$

where

$$\left|\mathbb{E}\left[e^*_{ti}e^*_{tj}\right]\right|\begin{cases}\in\left[0,c\log^{-2-\gamma}(Nr)\right) & \text{for}\;\; i\neq j \text{ s.t. } (i,j)\in\Gamma^c,\\ \in\left[c\log^{-2-\gamma}(Nr),\rho\right] & \text{for}\;\; i\neq j \text{ s.t. } (i,j)\in\Gamma,\\ =1 & \text{for}\;\; i=j,\end{cases}$$

for some constants $c>0$, $\gamma>0$, and $\rho\in(0,1)$ introduced in Assumption 6. Define

$$\begin{aligned}\mathcal{A}_1&=\{(i,j)\in[N]\times[N],(k,\ell)\in[r]\times[r]: k\neq \ell\}\cap\{(i,k),(j,\ell)\in\mathcal{S}^c\},\\ \mathcal{A}_2&=\{(i,j)\in[N]\times[N],(k,\ell)\in[r]\times[r]: i\neq j \text{ and } k=\ell\}\cap\{(i,k),(j,\ell)\in\mathcal{S}^c\},\\ \mathcal{A}_3&=\{(i,j)\in[N]\times[N],(k,\ell)\in[r]\times[r]: i=j \text{ and } k=\ell\}\cap\{(i,k),(j,\ell)\in\mathcal{S}^c\},\end{aligned}$$

and partition $\mathcal{A}_2$ into $\mathcal{A}_2^W=\mathcal{A}_2\cap\Gamma^c$ and $\mathcal{A}_2^S=\mathcal{A}_2\cap\Gamma$, where $\mathcal{A}_2^W$ and $\mathcal{A}_2^S$ are sets whose components have weak and strong correlations, respectively. Note that $|\mathcal{A}_1|=N^2(r^2-r)$, $|\mathcal{A}_2|=(N^2-N)r$, $|\mathcal{A}_3|=Nr$, $|\mathcal{A}_2^W|=|\mathcal{A}_2|-|\mathcal{A}_2^S|$, and $|\mathcal{A}_2^S|=O(N)$. Based on these sets, the first term of (A.6) is partitioned as

$$\begin{aligned}&\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\mathbb{P}\left(|Z^*_{ik}|\geq\mathtt{t}_m-\delta,|Z^*_{j\ell}|\geq\mathtt{t}_m-\delta\right)\\ &=\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_1\cup\mathcal{A}_2^W}\mathbb{P}\left(|Z^*_{ik}|\geq\mathtt{t}_m-\delta,|Z^*_{j\ell}|\geq\mathtt{t}_m-\delta\right)\\ &\quad+\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_2^S}\mathbb{P}\left(|Z^*_{ik}|\geq\mathtt{t}_m-\delta,|Z^*_{j\ell}|\geq\mathtt{t}_m-\delta\right)\\ &\quad+\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_3}\mathbb{P}\left(|Z^*_{ik}|\geq\mathtt{t}_m-\delta,|Z^*_{j\ell}|\geq\mathtt{t}_m-\delta\right).\end{aligned}\tag{A.7}$$

These terms are bounded by a similar way to the third term of (A.6). The first term of

(A.7) (weakly correlated variables) can be evaluated by Lemma 6.1 of Liu (2013):

$$
\begin{aligned}
&\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_1\cup\mathcal{A}_2^W}\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)\\
&=\frac{1}{N^2r^2}\cdot\frac{G(\mathtt{t}_m-\delta)^2}{G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_1\cup\mathcal{A}_2^W}\frac{\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)}{G(\mathtt{t}_m-\delta)^2}\\
&\lesssim\frac{1}{N^2r^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_1\cup\mathcal{A}_2^W}\left|\frac{\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)}{G(\mathtt{t}_m-\delta)^2}-1\right|+\frac{|\mathcal{A}_1\cup\mathcal{A}_2^W|}{N^2r^2}\\
&\leq\max_{(i,j,k,\ell)\in\mathcal{A}_1\cup\mathcal{A}_2^W}\left|\frac{\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)}{G(\mathtt{t}_m-\delta)^2}-1\right|+1\\
&=O\left(\log^{-1-(\gamma\wedge 0.5)}N\right)+1=o\left(\ell_N\right)+1.
\end{aligned}
$$

The second term of (A.7) (strongly correlated variables) can be evaluated by Lemma 6.2 of Liu (2013):

$$
\begin{aligned}
&\sum_{(i,j,k,\ell)\in\mathcal{A}_2^S}\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)\\
&\lesssim|\mathcal{A}_2^S|\frac{\exp\left\{-(\mathtt{t}_m-\delta)^2/(1+\rho)\right\}}{(\mathtt{t}_m-\delta+1)^2}\lesssim N\frac{(Nr)^{-2/(1+\rho)}\log^2(Nr)}{\log(Nr)}(1+o(1))\\
&=O\left(N^{1-2/(1+\rho)}\log N\right)=o\left(\ell_N\right).
\end{aligned}
$$

The third term of (A.7) becomes

$$
\begin{aligned}
&\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,j,k,\ell)\in\mathcal{A}_3}\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta, |Z^*_{j\ell}|\geq \mathtt{t}_m-\delta\right)\\
&\lesssim\frac{1}{N^2r^2G(\mathtt{t}_m)}\cdot\frac{G(\mathtt{t}_m-\delta)}{G(\mathtt{t}_m)}\sum_{(i,k)\in\mathcal{S}^c}\left|\frac{\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta\right)}{G(\mathtt{t}_m-\delta)}-1\right|+\frac{1}{NrG(\mathtt{t}_m)}\cdot\frac{G(\mathtt{t}_m-\delta)}{G(\mathtt{t}_m)}\\
&\lesssim\frac{1}{NrG(\mathtt{t}_m)}\max_{(i,k)\in\mathcal{S}^c}\left|\frac{\mathbb{P}\left(|Z^*_{ik}|\geq \mathtt{t}_m-\delta\right)}{G(\mathtt{t}_m-\delta)}-1\right|+\frac{1}{NrG(\mathtt{t}_m)}\\
&\lesssim\frac{1}{\log^{a/2-1/2}(Nr)}\frac{1}{\log^{1+\gamma\wedge 0.5}(Nr)}+\frac{1}{\log^{a/2-1/2}(Nr)}=o\left(\ell_N\right).
\end{aligned}
$$

Combining the obtained results reveal that (A.6) is $o(\ell_N)$. Therefore, (A.5) holds. This completes the proof. □

A.4 Proof of Theorem 4

Proof. Define

$$
\mathtt{t}_*=\Phi^{-1}\left(1-\frac{qs}{2Nr}(1-x_N)\right)\quad\text{with}\quad x_N=\frac{1}{\log N}. \tag{A.8}
$$

A direct use of Lemma 6 with condition $s/N = o(1/\log N)$ establishes that

$$\mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) \le \max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) = O(s/N) = o(1/\log N).$$

Furthermore, Lemma 7 gives

$$\mathbb{P}\left(|\mathtt{T}_{ik}| \ge \mathtt{t}_0\right) \ge \mathbb{P}\left(|\mathtt{T}_{ik}| \ge \mathtt{t}_0 \mid \mathtt{t}_0 \le \mathtt{t}_*\right) \mathbb{P}\left(\mathtt{t}_0 \le \mathtt{t}_*\right) \ge \mathbb{P}\left(|\mathtt{T}_{ik}| \ge \mathtt{t}_*\right)(1 + o(1)).$$

Using these results yield

$$\begin{aligned}
\text{Power} &= \frac{1}{s}\mathbb{E}\left[\sum_{(i,k)\in\mathcal{S}} 1\{|\mathtt{T}_{ik}| \ge \mathtt{t}_0\}\right] = \frac{1}{s}\sum_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \ge \mathtt{t}_0\right) \\
&\ge \frac{1}{s}\sum_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \ge \mathtt{t}_*\right)(1 + o(1)) = 1 - \frac{1}{s}\sum_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) + o(1) \\
&\ge 1 - \max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) + o(1) \ge 1 + o(1).
\end{aligned}$$

This completes the proof. □

A.5 Proof of Theorem 5

Proof. By the sparseness of $\mathbf{B}^0$, we have $b_{ik}^0 = b_{ik}^0 1\{(i,k)\in\mathcal{S}\} = b_{ik}^0 1\{(i,k)\in\widehat{\mathcal{S}}\}$ as long as $\mathcal{S}\subseteq\widehat{\mathcal{S}}$. Thus for any $\varepsilon > 0$, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\max_{i,k} |\hat{b}_{ik}^{\mathsf{d}} 1\{(i,k)\in\widehat{\mathcal{S}}\} - b_{ik}^0| > \varepsilon\right) \\
&\le \mathbb{P}\left(\max_{i,k} |\hat{b}_{ik}^{\mathsf{d}} 1\{(i,k)\in\widehat{\mathcal{S}}\} - b_{ik}^0 1\{(i,k)\in\mathcal{S}\}| > \varepsilon \mid \mathcal{S}\subseteq\widehat{\mathcal{S}}\right) + \mathbb{P}\left(\mathcal{S}\supsetneq\widehat{\mathcal{S}}\right) \\
&= \mathbb{P}\left(\max_{i,k} |\hat{b}_{ik}^{\mathsf{d}} - b_{ik}^0| 1\{(i,k)\in\widehat{\mathcal{S}}\} > \varepsilon\right) + \mathbb{P}\left(\mathcal{S}\supsetneq\widehat{\mathcal{S}}\right) \\
&\le \mathbb{P}\left(\max_{i,k} |\hat{b}_{ik}^{\mathsf{d}} - b_{ik}^0| > \varepsilon\right) + \mathbb{P}\left(\mathcal{S}\supsetneq\widehat{\mathcal{S}}\right).
\end{aligned}$$

Consider the first probability. By Theorem 1, it follows with high probability that

$$\begin{aligned}
\max_{i,k} |\hat{b}_{ik}^{\mathsf{d}} - b_{ik}^0| &\le \max_i \left\|\frac{1}{T}\sum_{t=1}^T e_{ti}\mathbf{f}_t^0\right\|_{\max} + \max_i \left\|\frac{1}{T^{1/2}}\mathbf{r}_i\right\|_{\max} \\
&\lesssim \frac{\log^{1/2}(N\vee T)}{T^{1/2}} + \frac{N_1^{3/2}\log(N\vee T)}{T^{1/2}N_r(N_r\wedge T)},
\end{aligned}$$

where the upper bound converges to zero under the assumed conditions. Next prove that the second probability goes to zero. For any $\delta\in(0,1)$, we have

$$\begin{aligned}
\mathbb{P}\left(\mathcal{S}\supsetneq\widehat{\mathcal{S}}\right) &\le \mathbb{P}\left(|\mathcal{S}| > |\widehat{\mathcal{S}}|\right) = \mathbb{P}\left(|\mathcal{S}| > |\widehat{\mathcal{S}}| + \delta\right) \\
&\le \mathbb{P}\left(1 - |\mathcal{S}\cap\widehat{\mathcal{S}}|/|\mathcal{S}| > \delta/|\mathcal{S}|\right) \le |\mathcal{S}|\left(1 - \mathbb{E}\,|\mathcal{S}\cap\widehat{\mathcal{S}}|/|\mathcal{S}|\right)/\delta,
\end{aligned}$$

where the last inequality holds by the Markov inequality along with the fact that $|\mathcal{S}\cap\widehat{\mathcal{S}}|/|\mathcal{S}| \leq 1$ a.s. From the proof of Theorem 4 and Lemma 6(ii), one minus the power is bounded as

$$1 - \mathbb{E}\,|\mathcal{S}\cap\widehat{\mathcal{S}}|/|\mathcal{S}| \leq \max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \leq \mathtt{t}_*\right) = O(s/N).$$

Therefore, since $s^2/N = o(1)$, the upper bound converges to zero. This completes the proof. □

B Lemmas and their Proofs

Lemma 1. *If Assumptions 1–3 are satisfied, then for any matrix (vector) norm $\|\cdot\|$, the inequalities (i)–(iii) simultaneously hold with probability at least $1 - O((N\vee T)^{-\nu})$:*

$$(i) \qquad \left\|T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}\right)\right\| \lesssim T^{-1/2}\log^{1/2}(N\vee T),$$

$$(ii) \qquad \left|T^{-1}\sum_{t=1}^{T}\left(e_{ti}^2 - \mathbb{E}\,e_{ti}^2\right)\right| \lesssim T^{-1/2}\log^{1/2}(N\vee T),$$

$$(iii) \qquad \left\|T^{-1}\sum_{t=1}^{T} e_{ti}\mathbf{f}_t^0\right\| \lesssim T^{-1/2}\log^{1/2}(N\vee T).$$

Moreover, if additionally Assumptions 5 and 6 are satisfied, then for any matrix norm $\|\cdot\|$, the inequality (iv) holds with probability at least $1 - O((N\vee T)^{-\nu})$:

$$(iv) \qquad \left\|T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}\right)\left(e_{ti}^2 - \mathbb{E}\,e_{ti}^2\right)\right\| \lesssim T^{-1/2}\log^{1/2}(N\vee T).$$

Proof. The proofs of (ii) and (iii) are found in Uematsu and Yamagata (2020). For (i), note that

$$\left\|T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}\right)\right\| \lesssim \max_{k\in[r]}\left|T^{-1}\sum_{t=1}^{T}\left({f_{tk}^0}^2 - 1\right)\right| + \max_{k\neq\ell}\left|T^{-1}\sum_{t=1}^{T} f_{tk}^0 f_{t\ell}^0\right|. \qquad \text{(A.9)}$$

On the first term of the upper bound, the summand has the same distributional structure as that of (ii). Therefore we can apply the same bound, $T^{-1/2}\log^{1/2}(N\vee T)$, up to a positive constant factor. The second term can be evaluated by the same way and is omitted.

Prove (iv). By the same decomposition as (A.9), we obtain

$$\begin{aligned}
&\left\|T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}\right)\left(e_{ti}^2 - \mathbb{E}\,e_{ti}^2\right)\right\| \\
&\lesssim \max_{k\in[r]}\left|T^{-1}\sum_{t=1}^{T}\left({f_{tk}^0}^2 - 1\right)\left(e_{ti}^2 - \mathbb{E}\,e_{ti}^2\right)\right| + \max_{k\neq\ell}\left|T^{-1}\sum_{t=1}^{T} f_{tk}^0 f_{t\ell}^0\left(e_{ti}^2 - \mathbb{E}\,e_{ti}^2\right)\right|. \qquad \text{(A.10)}
\end{aligned}$$

Consider the first term. Using Assumptions 5 and 6 along with the argument of Vershynin (2018), we first note that ${f_{tk}^0}^2 - 1$, $f_{tk}^0 f_{t\ell}^0$, and $e_{ti}^2 - \mathbb{E}\,e_{ti}^2$ are sub-exponential random variables. Furthermore, by Theorem 2.1 of Vladimirova and Arbel (2019), the product of two i.i.d. sub-exponential random variables is semi-exponential (sub-Weibull) with parameter 1/2. Therefore, by the Bernstein type inequality for semi-exponential random variables of Merlevède et al. (2011), there exist some constants $c_1, c_2 > 0$ such that

$$\mathbb{P}\left(\max_{k\in[r]}\left|T^{-1}\sum_{t=1}^{T}\left({f_{tk}^{0}}^{2}-1\right)\left(e_{ti}^{2}-\mathbb{E}\,e_{ti}^{2}\right)\right|>u\right)$$
$$\leq r\exp\left(-c_1Tu^2\right)+rT\exp\left(-c_2T^{1/2}u^{1/2}\right).$$

Setting $u \asymp T^{-1/2}\log^{1/2}(N\vee T)$ leads to the desired upper bound, which holds with probability at least

$$1-r\exp\left(-c_1\log(N\vee T)\right)-rT\exp\left(-c_2T^{1/4}\log^{1/4}(N\vee T)\right)$$
$$=1-O((N\vee T)^{-\nu}).$$

The second term in (A.10) is bounded by the same way. This completes the proof. □

Lemma 2. *If all the conditions in Theorem 1 are satisfied, then for any vector norm $\|\cdot\|$, the following inequalities simultaneously hold with probability at least $1-O((N\vee T)^{-\nu})$:*

$$(i)\quad T^{-1/2}\left\|\widehat{\mathbf{F}}-\mathbf{F}^0\right\|_{\mathrm{F}}\lesssim\frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_r(N_r\wedge T)},$$

$$(ii)\quad \max_{i\in[N]}\left\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\right\|\lesssim\frac{\log^{1/2}(N\vee T)}{T^{1/2}}\leq\frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_r(N_r\wedge T)}.$$

In particular, the upper bound converges to zero under (17).

Proof. Result (i) follows from Uematsu and Yamagata (2020). Prove (ii). From (8) with the triangle inequality, we have

$$\max_{i\in[N]}\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\|_{\max}\leq T^{-1}\eta_n+T^{-1/2}\|\mathbf{R}\|_{\max}+T^{-1/2}\|\mathbf{Z}\|_{\max},$$

where $\mathbf{Z}=T^{-1/2}\mathbf{E}'\mathbf{F}^0$ and $\mathbf{R}=\mathbf{R}^{(1)}+\mathbf{R}^{(2)}$ with $\mathbf{R}^{(1)}=T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}(\widehat{\mathbf{F}}-\mathbf{F}^0)$ and $\mathbf{R}^{(2)}=T^{-1/2}\mathbf{E}'(\widehat{\mathbf{F}}-\mathbf{F}^0)$. From Theorem 1, the definition of η_n, and Lemma 1, we have

$$T^{-1/2}\|\mathbf{R}\|_{\max}\lesssim T^{-1/2}\frac{N_1^{3/2}\log(N\vee T)}{N_r(N_r\wedge T)}$$

and

$$T^{-1}\eta_n+T^{-1/2}\|\mathbf{Z}\|_{\max}\lesssim T^{-1/2}\log^{1/2}(N\vee T),$$

which hold with probability at least $1-O((N\vee T)^{-\nu})$. Thus the first inequality follows by the equivalence of norms for finite dimensional vectors. The second inequality is true since

$$\frac{N_1^{3/2}T^{1/2}}{N_r(N_r\wedge T)}\geq\frac{N_1^{1/2}T^{1/2}}{N_r\wedge T}=\frac{(N_1\vee T)^{1/2}}{(N_r\wedge T)^{1/2}}\geq 1. \tag{A.11}$$

Convergence of the bounds is easily verified from (17). This completes the proof. □

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

$$(i) \qquad \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|\hat{e}_{ti}^2 - e_{ti}^2\right| \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_r(N_r \wedge T)},$$

$$(ii) \qquad \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|\hat{e}_{ti}^2 - e_{ti}^2\right|^2 \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_r(N_r \wedge T)}.$$

Proof. First note that

$$\begin{aligned}
\hat{e}_{ti}^2 - e_{ti}^2 &= (x_{ti} - \hat{c}_{ti})^2 - e_{ti}^2 \\
&= \left(e_{ti} - \left(\hat{c}_{ti} - c_{ti}^0\right)\right)^2 - e_{ti}^2 = -2e_{ti}\left(\hat{c}_{ti} - c_{ti}^0\right) + \left(\hat{c}_{ti} - c_{ti}^0\right)^2
\end{aligned}$$

and

$$\hat{c}_{ti} - c_{ti}^0 = (\hat{\mathbf{f}}_t - \mathbf{f}_t^0)'\mathbf{b}_i^0 + \hat{\mathbf{f}}_t'(\widehat{\mathbf{b}}_i - \mathbf{b}_i^0).$$

Prove (i). We have

$$\begin{aligned}
&\max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|\hat{e}_{ti}^2 - e_{ti}^2\right| \\
&\lesssim \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} |e_{ti}| \left|\hat{c}_{ti} - c_{ti}^0\right| + \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|\hat{c}_{ti} - c_{ti}^0\right|^2 \\
&\lesssim \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} |e_{ti}| \left|(\hat{\mathbf{f}}_t - \mathbf{f}_t^0)'\mathbf{b}_i^0\right| + \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} |e_{ti}| \left|\hat{\mathbf{f}}_t'(\widehat{\mathbf{b}}_i - \mathbf{b}_i^0)\right| \\
&\qquad + \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|(\hat{\mathbf{f}}_t - \mathbf{f}_t^0)'\mathbf{b}_i^0\right|^2 + \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left|\hat{\mathbf{f}}_t'(\widehat{\mathbf{b}}_i - \mathbf{b}_i^0)\right|^2 \\
&=: A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

Consider each term. In the following, we use $\max_{i\in[N]} \|\mathbf{b}_i^0\|_2 < \infty$. First A_1 is bounded as

$$\begin{aligned}
A_1 &\leq \max_{i\in[N]} \|\mathbf{b}_i^0\|_2 T^{-1} \sum_{t=1}^{T} |e_{ti}| \|\hat{\mathbf{f}}_t - \mathbf{f}_t^0\|_2 \\
&\leq \max_{i\in[N]} \|\mathbf{b}_i^0\|_2 \left(T^{-1} \sum_{t=1}^{T} |e_{ti}|^2\right)^{1/2} T^{-1/2} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} \\
&\lesssim T^{-1/2} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}}.
\end{aligned}$$

Similarly, we obtain

$$
\begin{aligned}
A_2 &\leq \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2 T^{-1} \sum_{t=1}^{T} |e_{ti}| \|\hat{\mathbf{f}}_t\|_2 \\
&\leq \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2 \left(T^{-1} \sum_{t=1}^{T} |e_{ti}|^2 \right)^{1/2} T^{-1/2} \|\widehat{\mathbf{F}}\|_{\mathrm{F}} \\
&\lesssim \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2.
\end{aligned}
$$

Next, we see that

$$
\begin{aligned}
A_3 &= \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left| (\hat{\mathbf{f}}_t - \mathbf{f}_t^0)' \mathbf{b}_i^0 \right|^2 \\
&\leq \max_{i\in[N]} \|\mathbf{b}_i^0\|_2^2 T^{-1} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}}^2 \lesssim T^{-1} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}}^2.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
A_4 &= \max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left| \hat{\mathbf{f}}_t' (\widehat{\mathbf{b}}_i - \mathbf{b}_i^0) \right|^2 \\
&\leq \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^2 T^{-1} \|\widehat{\mathbf{F}}\|_{\mathrm{F}}^2 \lesssim \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^2.
\end{aligned}
$$

From the argument so far with Lemma 2, we conclude that

$$
\begin{aligned}
&\max_{i\in[N]} T^{-1} \sum_{t=1}^{T} \left| \hat{e}_{ti}^2 - e_{ti}^2 \right| \\
&\lesssim T^{-1/2} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} + \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2 + T^{-1} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}}^2 + \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^2 \\
&\lesssim T^{-1/2} \|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_r (N_r \wedge T)},
\end{aligned}
$$

which gives the proof of (i).

Prove (ii). We have

$$
\begin{aligned}
&\max_{i\in[N]} T^{-1}\sum_{t=1}^{T}\left|\hat{e}_{ti}^2-e_{ti}^2\right|^2 \\
&\lesssim \max_{i\in[N]} T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\left|\hat{c}_{ti}-c_{ti}^0\right|^2+\max_{i\in[N]} T^{-1}\sum_{t=1}^{T}\left|\hat{c}_{ti}-c_{ti}^0\right|^4 \\
&\lesssim \max_{i\in[N]} T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\left|(\hat{\mathbf{f}}_t-\mathbf{f}_t^0)'\mathbf{b}_i^0\right|^2+\max_{i\in[N]} T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\left|\hat{\mathbf{f}}_t'(\widehat{\mathbf{b}}_i-\mathbf{b}_i^0)\right|^2 \\
&\quad+\max_{i\in[N]} T^{-1}\sum_{t=1}^{T}\left|(\hat{\mathbf{f}}_t-\mathbf{f}_t^0)'\mathbf{b}_i^0\right|^4+\max_{i\in[N]} T^{-1}\sum_{t=1}^{T}\left|\hat{\mathbf{f}}_t'(\widehat{\mathbf{b}}_i-\mathbf{b}_i^0)\right|^4 \\
&=: A_5+A_6+A_7+A_8.
\end{aligned}
$$

Consider each term. In the following, we use $\max_{i\in[N]}\|\mathbf{b}_i^0\|_2<\infty$. First A_5 is bounded as

$$
\begin{aligned}
A_5 &\le \max_{i\in[N]}\|\mathbf{b}_i^0\|_2^2 T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\|\hat{\mathbf{f}}_t-\mathbf{f}_t^0\|_2^2 \\
&\le \max_{i\in[N]}\|\mathbf{b}_i^0\|_2^2\left(T^{-1}\sum_{t=1}^{T}|e_{ti}|^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\|\hat{\mathbf{f}}_t-\mathbf{f}_t^0\|_2^4\right)^{1/2} \\
&\le \max_{i\in[N]}\|\mathbf{b}_i^0\|_2^2\left(\mathbb{E}\,|e_{ti}|^4+o(1)\right)^{1/2}\left\{2\max_t\left(\|\hat{\mathbf{f}}_t\|_2^2+\|\mathbf{f}_t^0\|_2^2\right)T^{-1}\|\widehat{\mathbf{F}}-\mathbf{F}^0\|_{\mathrm{F}}^2\right\}^{1/2} \\
&\lesssim T^{-1/2}\|\widehat{\mathbf{F}}-\mathbf{F}^0\|_{\mathrm{F}}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
A_6 &\le \max_{i\in[N]}\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\|_2^2 T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\|\hat{\mathbf{f}}_t\|_2^2 \\
&\le \max_{i\in[N]}\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\|_2^2\max_t\|\hat{\mathbf{f}}_t\|_2^2\left(T^{-1}\sum_{t=1}^{T}|e_{ti}|^2\right)^{1/2} \\
&\le \max_{i\in[N]}\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\|_2^2\max_t\|\hat{\mathbf{f}}_t\|_2^2\left(\mathbb{E}\,|e_{ti}|^2+o(1)\right)^{1/2}\lesssim\max_{i\in[N]}\|\widehat{\mathbf{b}}_i-\mathbf{b}_i^0\|_2^2
\end{aligned}
$$

Next,

$$
\begin{aligned}
A_7 &\le \max_{i\in[N]}\|\mathbf{b}_i^0\|_2^4 T^{-1}\sum_{t=1}^{T}\left\|\hat{\mathbf{f}}_t-\mathbf{f}_t^0\right\|_2^4 \\
&\le \max_{i\in[N]}\|\mathbf{b}_i^0\|_2^4\max_t\left(2\|\hat{\mathbf{f}}_t\|_2^2+2\|\mathbf{f}_t^0\|_2^2\right)T^{-1}\|\widehat{\mathbf{F}}-\mathbf{F}^0\|_{\mathrm{F}}^2\lesssim T^{-1}\|\widehat{\mathbf{F}}-\mathbf{F}^0\|_{\mathrm{F}}^2.
\end{aligned}
$$

Similarly,

$$A_8 \leq \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^4 T^{-1} \sum_{t=1}^{T} \left\|\hat{\mathbf{f}}_t\right\|_2^4$$
$$\leq \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^4 \max_t \left\|\hat{\mathbf{f}}_t\right\|_2^2 T^{-1}\|\widehat{\mathbf{F}}\|_{\mathrm{F}}^2 \lesssim \max_{i\in[N]} \|\widehat{\mathbf{b}}_i - \mathbf{b}_i^0\|_2^4.$$

By the same reason as the proof of (i), the result follows. This completes the proof. □

Lemma 4. *If all the conditions of Theorem 3 are satisfied, then the following inequality holds with probability at least* $1 - O((N \vee T)^{-\nu})$*:*

$$\left\|\widehat{\mathbf{\Gamma}}_i - \sigma_i^2 \mathbf{I}_r\right\|_{\max} \lesssim \frac{N_1^{3/2} \log^{1/2}(N \vee T)}{N_r(N_r \wedge T)}.$$

Proof. Under Assumptions 5 and 6, we have $\mathbf{\Gamma}_i = \mathbb{E}[\mathbf{f}_t\mathbf{f}_t' e_{ti}^2] = \sigma_i^2 \mathbf{I}_r$ and $\widehat{\mathbf{\Gamma}}_i = \widehat{\mathbf{\Gamma}}_{0i}$. Then it follows that

$$\left\|\widehat{\mathbf{\Gamma}}_i - \sigma_i^2 \mathbf{I}_r\right\|_{\max} \leq \left\|T^{-1}\sum_{t=1}^{T}(\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t' - \mathbf{f}_t\mathbf{f}_t')\hat{e}_{ti}^2\right\|_{\max} + \left\|T^{-1}\sum_{t=1}^{T}(\mathbf{f}_t\mathbf{f}_t' - \mathbf{I}_r)\hat{e}_{ti}^2\right\|_{\max}$$
$$+ \max_{i\in[N]}\left|T^{-1}\sum_{t=1}^{T}(\hat{e}_{ti}^2 - e_{ti}^2)\right| + \left|T^{-1}\sum_{t=1}^{T}(e_{ti}^2 - \mathbb{E}\, e_{ti}^2)\right|$$
$$=: A_1 + A_2 + A_3 + A_4.$$

We first see that A_3 and A_4 are directly bounded from Lemmas 3(i) and 1(ii), respectively. Next we bound A_1. By the triangle inequality and the Cauchy–Schwarz inequality, we have

$$A_1 \leq \left(T^{-1}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t' - \mathbf{f}_t\mathbf{f}_t'\|_{\max}^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\hat{e}_{ti}^4\right)^{1/2}.$$

By Lemma 3, the second parentheses can be bounded as

$$
\begin{aligned}
\Big(T^{-1}\sum_{t=1}^{T}\hat{e}_{ti}^4\Big)^{1/2} &\leq \Big(T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^4 - e_{ti}^4\big|\Big)^{1/2} + \Big(T^{-1}\sum_{t=1}^{T}e_{ti}^4\Big)^{1/2} \\
&= \Big(T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|\,\big|\hat{e}_{ti}^2 - e_{ti}^2 + 2e_{ti}^2\big|\Big)^{1/2} + \big(\mathbb{E}\, e_{ti}^4 + o(1)\big)^{1/2} \\
&\leq \Big(T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|^2\Big)^{1/2} + \Big(2T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|\,\big|e_{ti}^2\big|\Big)^{1/2} + \big(\mathbb{E}\, e_{ti}^4 + o(1)\big)^{1/2} \\
&\leq \Big(T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|^2\Big)^{1/2} + \Big(2T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|^2\Big)^{1/4}\Big(2T^{-1}\sum_{t=1}^{T}e_{ti}^4\Big)^{1/4} + \big(\mathbb{E}\, e_{ti}^4 + o(1)\big)^{1/2} \\
&= \Big(T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|^2\Big)^{1/2} + \Big(2T^{-1}\sum_{t=1}^{T}\big|\hat{e}_{ti}^2 - e_{ti}^2\big|^2\Big)^{1/4}\big(2\,\mathbb{E}\, e_{ti}^4 + o(1)\big)^{1/4} + \big(\mathbb{E}\, e_{ti}^4 + o(1)\big)^{1/2} \\
&\lesssim (\mathbb{E}\, e_{ti}^4)^{1/2} + o(1).
\end{aligned}
$$

Therefore we eventually have

$$
\begin{aligned}
A_1 &\lesssim \Big(T^{-1}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t(\widehat{\mathbf{f}}_t - \mathbf{f}_t^0)'\|_{\max}^2 + T^{-1}\sum_{t=1}^{T}\|(\widehat{\mathbf{f}}_t - \mathbf{f}_t^0)\mathbf{f}_t^{0\prime}\|_{\max}^2\Big)^{1/2} \\
&\lesssim \Big(T^{-1}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t - \mathbf{f}_t^0\|_2^2 + T^{-1}\sum_{t=1}^{T}\|\widehat{\mathbf{f}}_t - \mathbf{f}_t^0\|_2^2\Big)^{1/2} \\
&\lesssim T^{-1/2}\|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} \lesssim \frac{N_1^{3/2}\log^{1/2}(N \vee T)}{N_r(N_r \wedge T)},
\end{aligned}
$$

where the last inequality follows from Lemma 2(i). Finally bound A_2. We further expand the terms by the triangle inequality:

$$
\begin{aligned}
A_2 &\leq \Big\|T^{-1}\sum_{t=1}^{T}(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}_r)\hat{e}_{ti}^2\Big\|_{\max} \\
&\leq \Big\|T^{-1}\sum_{t=1}^{T}(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}_r)(\hat{e}_{ti}^2 - e_{ti}^2)\Big\|_{\max} + \Big\|T^{-1}\sum_{t=1}^{T}(\mathbf{f}_t^0\mathbf{f}_t^{0\prime} - \mathbf{I}_r)(e_{ti}^2 - \mathbb{E}\, e_{ti}^2)\Big\|_{\max},
\end{aligned}
$$

where we have used the condition $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}$. The second term of this upper bound is directly evaluated by Lemma 1(iv). By Lemma 3(i), the first term is further bounded as

$$\begin{aligned}
&\left\| T^{-1}\sum_{t=1}^{T}(\mathbf{f}_t^0\mathbf{f}_t^{0\prime}-\mathbf{I}_r)(\hat{e}_{ti}^2-e_{ti}^2)\right\|_{\max} \\
&\leq \max_t \left\|\mathbf{f}_t^0\mathbf{f}_t^{0\prime}-\mathbf{I}_r\right\|_2 T^{-1}\sum_{t=1}^{T}\left|\hat{e}_{ti}^2-e_{ti}^2\right| \leq \max_t \left(\|\mathbf{f}_t^0\|_2^2+1\right) T^{-1}\sum_{t=1}^{T}\left|\hat{e}_{ti}^2-e_{ti}^2\right| \\
&\lesssim \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_r(N_r\wedge T)}.
\end{aligned}$$

Consequently, we obtain

$$\left\|\widehat{\mathbf{\Gamma}}_i-\sigma_i^2\mathbf{I}_r\right\|_{\max} \lesssim \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_r(N_r\wedge T)}+\frac{\log^{1/2}(N\vee T)}{T^{1/2}} \lesssim \frac{N_1^{3/2}\log^{1/2}(N\vee T)}{N_r(N_r\wedge T)},$$

where we have used (A.11) in the last inequality. Note that all the bounds hold with probability at least $1-O((N\vee T)^{-\nu})$. This completes the proof. □

Lemma 5. *Define $\delta \asymp \delta_1 \log^{1/2}(N\vee T)$, where*

$$\delta_1 = \frac{N_1^{3/2}\log(N\vee T)}{N_r(N_r\wedge T)}$$

has been defined in Theorem 1. If all the conditions of Theorem 3 are satisfied, then for any $\mathtt{t}>0$ the following results simultaneously hold:

$$\begin{aligned}
&(i) \qquad \max_{i,k}\mathbb{P}\left(|W_{ik}|\geq\delta\right)=O((N\vee T)^{-\nu}),\\
&(ii) \qquad \mathbb{P}\left(|\mathtt{T}_{ik}|\geq\mathtt{t}\right)\geq\mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}}\geq\mathtt{t}+\delta\right)+O((N\vee T)^{-\nu}),\\
&(iii) \qquad \mathbb{P}\left(|\mathtt{T}_{ik}|\geq\mathtt{t},|\mathtt{T}_{j\ell}|\geq\mathtt{t}\right)\leq\mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}}\geq\mathtt{t}-\delta,\frac{|Z_{j\ell}|}{\sigma_{j\ell}}\geq\mathtt{t}-\delta\right)+O((N\vee T)^{-\nu}).
\end{aligned}$$

Proof. For $(i,k)\in\mathcal{S}^c$, the t-statistic is written as

$$\begin{aligned}
\mathtt{T}_{ik}=\frac{T^{1/2}\hat{b}_{ik}^*}{\hat{\sigma}_{ik}}&=\frac{Z_{ik}+R_{ik}}{\hat{\sigma}_{ik}}\\
&=\frac{Z_{ik}}{\sigma_{ik}}+\frac{R_{ik}}{\sigma_{ik}}+\left(\frac{\sigma_{ik}}{\hat{\sigma}_{ik}}-1\right)\left(\frac{Z_{ik}+R_{ik}}{\sigma_{ik}}\right)=:\frac{Z_{ik}}{\sigma_{ik}}+W_{ik}.
\end{aligned}$$

Consider (ii) and (iii) first. For any $\mathtt{t}>0$ and δ given in the statement, we have

$$\mathbb{P}\left(|\mathtt{T}_{ik}|\geq\mathtt{t}\right)\geq\mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}}-|W_{ik}|\geq\mathtt{t}\right)\geq\mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}}\geq\mathtt{t}+\delta\right)-\mathbb{P}\left(|W_{ik}|\geq\delta\right)$$

and

$$
\begin{aligned}
&\mathbb{P}\left(|\mathtt{T}_{ik}| \geq \mathtt{t}, |\mathtt{T}_{j\ell}| \geq \mathtt{t}\right) \\
&\leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} + |W_{ik}| \geq \mathtt{t}, \frac{|Z_{j\ell}|}{\sigma_{j\ell}} + |W_{j\ell}| \geq \mathtt{t}\right) \\
&\leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} \geq \mathtt{t} - \delta, \frac{|Z_{j\ell}|}{\sigma_{j\ell}} \geq \mathtt{t} - \delta\right) + \mathbb{P}\left(\max_{i,k} |W_{ik}| > \delta\right) \\
&\leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} \geq \mathtt{t} - \delta, \frac{|Z_{j\ell}|}{\sigma_{j\ell}} \geq \mathtt{t} - \delta\right) + |\mathcal{S}^c| \max_{i,k} \mathbb{P}\left(|W_{ik}| > \delta\right).
\end{aligned}
$$

Thus the proof completes if (i) is true. We have

$$
\begin{aligned}
&\mathbb{P}\left(|W_{ik}| \geq \delta\right) \\
&\leq \mathbb{P}\left(\frac{|R_{ik}|}{\sigma_{ik}} + \left|\frac{\sigma_{ik}}{\hat{\sigma}_{ik}} - 1\right| \left(\frac{|Z_{ik}|}{\sigma_{ik}} + \frac{|R_{ik}|}{\sigma_{ik}}\right) > \delta\right) \\
&\leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} > \frac{\delta - \delta_1}{\delta_1} - \delta_1\right) + \mathbb{P}\left(\left|\frac{\sigma_{ik}}{\hat{\sigma}_{ik}} - 1\right| > \delta_1\right) + \mathbb{P}\left(\frac{|R_{ik}|}{\sigma_{ik}} > \delta_1\right). \qquad \text{(A.12)}
\end{aligned}
$$

Then the third term in the upper bound of (A.12) is evaluated by the proof of Theorem 1:

$$
\mathbb{P}\left(\frac{|R_{ik}|}{\sigma_{ik}} \gtrsim \delta_1\right) = O((N \vee T)^{-\nu}).
$$

Consider the second term of (A.12). By Lemma 4 with a simple calculation, we have

$$
\begin{aligned}
\mathbb{P}\left(\left|\frac{\sigma_{ik}}{\hat{\sigma}_{ik}} - 1\right| > \delta_1\right) &= \mathbb{P}\left(\left|\hat{\sigma}_{ik}^2 - \sigma_{ik}^2\right| > \delta_1 \left|\hat{\sigma}_{ik}\right| \left|\hat{\sigma}_{ik} + \sigma_{ik}\right|\right) \\
&\lesssim \mathbb{P}\left(\left|\hat{\sigma}_{ik}^2 - \sigma_{ik}^2\right| \gtrsim \delta_1\right) = O((N \vee T)^{-\nu}).
\end{aligned}
$$

Finally, consider the first term of (A.12). Note that $\delta_1 = o(\delta)$ and $\delta/\delta_1 \asymp \log^{1/2}(N \vee T)$. Therefore, we have

$$
\begin{aligned}
\mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} > \frac{\delta - \delta_1}{\delta_1} - \delta_1\right) &= \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} > \frac{\delta + o(\delta)}{\delta_1} + o(\delta)\right) \\
&= \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} > \frac{\delta}{\delta_1}(1 + o(1))\right) \\
&\lesssim \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_{ik}} \gtrsim \log^{1/2}(N \vee T)\right) = O((N \vee T)^{-\nu}),
\end{aligned}
$$

where the last equation is due to Lemma 1(iii). Combining the results leads to the proof of (i). This completes the proof. □

Lemma 6. *If all the conditions of Theorem 4 are satisfied, then the following result holds:*

$$
\max_{(i,k) \in \mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \leq \mathtt{t}_*\right) = O(s/N).
$$

Proof. For any $(i,k) \in \mathcal{S}$, the t-statistic is decomposed as

$$\mathtt{T}_{ik} = \frac{T^{1/2}\hat{b}_{ik}^*}{\hat{\sigma}_{ik}} = \frac{\sigma_{ik}}{\hat{\sigma}_{ik}} \cdot \frac{T^{1/2}b_{ik}^0 + Z_{ik} + R_{ik}}{\sigma_{ik}} =: \frac{\sigma_{ik}}{\hat{\sigma}_{ik}} \left(T^{1/2}b_{ik}^* + Z_{ik}^* + R_{ik}^* \right),$$

where Z_{ik} and R_{ik} were defined in (8). Then for any $(i,k) \in \mathcal{S}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) &= \mathbb{P}\left(|T^{1/2}b_{ik}^* + Z_{ik}^* + R_{ik}^*| \le \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \right) \\
&\le \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le |Z_{ik}^*| \right) \\
&\le \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le |Z_{ik}^*| \;\middle|\; |T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* > \mathtt{t}_* \right) \\
&\quad + \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le \mathtt{t}_* \right) \\
&\le \mathbb{P}\left(\mathtt{t}_* \le |Z_{ik}^*|\right) + \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le \mathtt{t}_* \right). \qquad \text{(A.13)}
\end{aligned}$$

By the proof of Theorem 3, the first term of (A.13) is approximated by $G(\mathtt{t}_*)$:

$$\max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(\mathtt{t}_* \le |Z_{ik}^*|\right) = G(\mathtt{t}_*)(1 + o(1)).$$

Recall that $\mathtt{t}_* = \Phi^{-1}\left(1 - qs(1 - o(1))/(2Nr)\right)$. Then we obtain

$$G(\mathtt{t}_*) = 2(1 - \Phi(\mathtt{t}_*)) = 2 - 2\Phi\left(\Phi^{-1}\left(1 - \frac{qs}{2Nr}(1 - o(1)) \right) \right) = \frac{qs}{Nr}(1 - o(1)).$$

Therefore, the first term of (A.13) is evaluated as

$$\max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(\mathtt{t}_* \le |Z_{ik}^*|\right) = O(s/N). \qquad \text{(A.14)}$$

The second term of (A.13) is bounded as follows. Fix $c_1 > 1$ arbitrary. Then we have

$$\begin{aligned}
&\mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le \mathtt{t}_* \right) \\
&\le \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le \mathtt{t}_* \;\middle|\; \frac{\hat{\sigma}_{ik}}{\sigma_{ik}} \le c_1 \right) + \mathbb{P}\left(\frac{\hat{\sigma}_{ik}}{\sigma_{ik}} > c_1 \right) \\
&\le \mathbb{P}\left(|T^{1/2}b_{ik}^*| - (1 + c_1)\mathtt{t}_* \le |R_{ik}^*| \right) + \mathbb{P}\left(|\hat{\sigma}_{ik}^2 - \sigma_{ik}^2| > \sigma_{ik}^2(c_1^2 - 1) \right). \qquad \text{(A.15)}
\end{aligned}$$

Note that $\mathtt{t}_* < \sqrt{2\log(Nr)}$ by the construction. Hence by Assumption 7, we obtain

$$|T^{1/2}b_{ik}^*| - (1 + c_1)\mathtt{t}_* > \min_{(i,k)\in\mathcal{S}} |T^{1/2}b_{ik}^*| - (1 + c_1)\sqrt{2\log(Nr)} \gtrsim \sqrt{2\log(Nr)}.$$

Therefore, in view of the proof of Lemma 5(i), the upper bound of (A.15) is found to be $O((N \vee T)^{-\nu})$, which leads to

$$\max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|T^{1/2}b_{ik}^*| - |R_{ik}^*| - \frac{\hat{\sigma}_{ik}}{\sigma_{ik}}\mathtt{t}_* \le \mathtt{t}_* \right) = O((N \vee T)^{-\nu}). \qquad \text{(A.16)}$$

From (A.14) and (A.16), we bound (A.13) as

$$\max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) = O(s/N),$$

which completes the proof. □

Lemma 7. *If all the conditions of Theorem 4 are satisfied, then* $\mathtt{t}_0 \le \mathtt{t}_*$ *holds with high probability.*

Proof. Recall that $\mathtt{t}_* = \Phi^{-1}\left(1 - qs(1-x_N)/(2Nr)\right)$ with $x_N = 1/\log N$. Prove the statement by contradiction. Suppose $\mathtt{t}_0 > \mathtt{t}_*$ a.s. By the definition of $\mathtt{t}_0$, we have

$$\frac{2Nr(1-\Phi(\mathtt{t}_*))}{R(\mathtt{t}_*)} > q \tag{A.17}$$

with probability one. By the definition of $R(\mathtt{t}_*)$, it holds that

$$R(\mathtt{t}_*) = |\widehat{\mathcal{S}}(\mathtt{t}_*)| = |\mathcal{S} \cup \widehat{\mathcal{S}}(\mathtt{t}_*)| - |\mathcal{S} \cap \widehat{\mathcal{S}}(\mathtt{t}_*)^c| \ge s - |\mathcal{S} \cap \widehat{\mathcal{S}}(\mathtt{t}_*)^c|.$$

The Markov inequality gives

$$\begin{aligned}
&\mathbb{P}\left(|\mathcal{S} \cap \widehat{\mathcal{S}}(\mathtt{t}_*)^c| > s x_N\right) \le \frac{\log N}{s}\, \mathbb{E}\left|\mathcal{S} \cap \widehat{\mathcal{S}}(\mathtt{t}_*)^c\right| \\
&= \frac{\log N}{s}\, \mathbb{E}\left(\sum_{(i,k)\in\mathcal{S}} 1\{|\mathtt{T}_{ik}| \le \mathtt{t}_*\}\right) \le \log N \max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|\mathtt{T}_{ik}| \le \mathtt{t}_*\right) = o(1),
\end{aligned} \tag{A.18}$$

where the last equality holds by Lemma 6. Hence we have $R(\mathtt{t}_*) \ge s(1-x_N)$ with high probability. This lower bound with (A.17) entails that

$$1 - \Phi(\mathtt{t}_*) > \frac{qR(\mathtt{t}_*)}{2Nr} \ge \frac{qs(1-x_N)}{2Nr} \tag{A.19}$$

with high probability. On the other hand, by the definition of $\mathtt{t}_*$, we have

$$1 - \Phi(\mathtt{t}_*) = \frac{qs(1-x_N)}{2Nr},$$

but this equality contradicts (A.19). This completes the proof. □

References

Ando, T. and J. Bai (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association 112*, 1182–1198.

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

Bai, J. and K. Li (2014, 02). Theory and methods of panel data models with interactive effects. *Ann. Statist. 42*(1), 142–170.

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

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions. *Econometrica 74*, 1133–1150.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*, 18–29.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B 57*, 289–300.

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. *Studi in Onore del Professore Salvatore Ortu Carboni*, 13–60.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica 51*, 1281–1304.

Chernozhukov, V., C. Hansen, and M. Spindler (2015). Valid post-selection and postregularization inference: An elementary, general approach. *Annual Review of Economics 7*, 649–688.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*, 373–394.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor modela test for the number of factors in an approximate factor model. *Journal of Finance 48*, 1263–1291.

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

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

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

Feng, G., S. Giglio, and D. Xiu (2019). Taming the factor zoo: A test of new factors. *NBER Working Paper No. 25481*.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*, 65–70.

Javanmard, A. and H. Javadi (2019). False discovery rate control via debiased lasso. *Electronic Journal of Statistics 13*, 1212–1253.

Javanmard, A. and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research 15*, 2869–2909.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*, 271–292.

Leeb, H. and B. M. Pötscher (2008). Sparse estimators and the oracle property, or the return of hodges' estimator. *Journal of Econometrics 142*, 201–211.

Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Annals of Statistics 41*, 2948–2978.

Ludvigson, C. S. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*, 5027–5067.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Merlevède, F., M. Peligrad, and E. Rio (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*, 435–474.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

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

Pötscher, B. M. and H. Leeb (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis 100*, 2065–2082.

Rigollet, P. and J.-C. Hütter (2017). *High Dimensional Statistics.* Massachusetts Institute of Technology, MIT Open CourseWare.

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 30*, 147–162.

Uematsu, Y., Y. Fan, K. Chen, J. Lv, and W. Lin (2019). SOFAR: large-scale association network learning. *IEEE Transactions on Information Theory 65*, 4929–4939.

Uematsu, Y. and T. Yamagata (2020). Estimation of weak factor models. *SSRN: https://ssrn.com/abstract=3374750*.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics 42*, 1166–1202.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press.

Vladimirova, M. and J. Arbel (2019). Sub-weibull distributions: generalizing sub-gaussian and sub-exponential properties to heavier-tailed distributions. *arXiv:1905.04955v1*, 1–11.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B 76*, 217–242.

Supplementary Material for

Inference in Weak Factor Models

Yoshimasa Uematsu* and Takashi Yamagata†

*Department of Economics and Management, Tohoku University
†Department of Economics and Related Studies, University of York
†Institute of Social Economic Research, Osaka University

B.1 Empirical size of the t-test

Next, we investigate the finite sample behaviour of (feasible) t-statistics. Unlike the previous subsection, the t-statistics are based on estimated variances of $\hat{b}_{N2}$. To assess finite sample performance the t-tests in virtually practical situations, we focus on investigating empirical size of the t-tests of factor loadings based on debiased SOFAR and PC estimators. We consider two t-statistics which are based on two standard errors: $t_{N2,i.i.d.}$ is the t-statistic of $\hat{b}_{N2}$ based on the variance assuming no serial correlation; $t_{N2,NW}$ is based on the variance estimator permitting error serial correlation. From now on we denote the t-statistics without subscript ik to ease the notation, unless clarification is necessary. We consider two cases. For the first design the factors and errors are serially independent ($\rho_{fk} = 0$ and $\rho_e = 0$ for all k), and for the second design they are serially correlated ($\rho_{fk} = 1/4$ and $\rho_e = 1/4$ for all k). We conduct two-sided test at the five per cent significance level, by rejecting the null $H_0 : b_{N2} = 0$ when the absolute value of the t-statistic is greater than 1.96. We investigate all the combinations of $N = 100$ and $T = 100, 200, 500, 1000$. The results are based on 2000 replications.

Table 1 reports the estimated size of the test. Panel A reports the results for the case of i.i.d. factors and errors, and Panel B summarises the size of the tests for serially correlated errors. Let us look at Panel A. First, let us look at the size behaviour of t_{iid} since it is expected to be more efficient than t_{NW}. Even when $T = 100$, the t-test based on the debiased SOFAR estimatorhas satisfactory size across the models, exhibiting only minor size distortions. For $T = 100$, the size of t_{iid} for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$ are 6.6%, 5.7% and 7.7%, respectively. In contrast, the size of the t-test based on the PC estimator are more distorted, and the degree of its size distortion becomes severer as the model becomes weaker. For example, for $T = 100$, the associated size of t_{iid} for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$ are 7.5%, 6.6% and 10.6%, respectively. Furthermore, the size distortion does not seem to disappear when T rises for the models with weaker factors. For instance, for the model with exponents $\{0.5, 0.4\}$, when T rises from 500 to 1000, the sizes of the t_{iid} based on the debiased SOFAR estimator are 5.4% and 4.7%, while those based on the PC are 7.7% and 7.7%.

Now let turn our attention to the serial correlation robust test, t_{NW}. As can be seen in Panel A, for $T = 100$, the size of t_{NW} based on the debiased SOFAR estimator exhibits a moderate size distortion: for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$, the associated sizes are 7.6%, 6.7% and 7.8%, respectively. Whereas, such distortions disappear quickly as T rises: the corresponded sizes for $T = 200$ become 6.7%, 6.6% and 6.3%, respectively. In contrast, the observed size distortion pattern of t_{iid} based on the PC estimator is exaggerated for the serial correlation robust test. For $T = 100$, the sizes of t_{NW} based on the PC for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$ are 8.0%, 7.3% and 11.0%, respectively. Increasing T to 200 does not seem to make the distortion sufficiently reduced: the corresponding sizes become 7.0%, 7.4% and 9.1%, respectively.

Let us turn our attention to Panel B. First thing to note is that the size distortion of t_{iid} does not ease as T rises. This is expected since the variance estimator used for t_{iid} is not consistent. An interesting feature to note is that the size distortion of t_{iid} based on the PC is much larger than that based on the debiased SOFAR, and the distortion of the PC tests widen as the model becomes weaker. For example, the average of the sizes over $T = 100, ..., 1000$ of t_{iid} based on the debiased SOFAR for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$ are 7.4%, 7.6% and 6.8%, while that of t_{iid} based on the PC are 7.9%, 8.4% and 10.6%, respectively. For sufficiently large $T (\geq 200)$, t_{NW} based on the debiased SOFAR has satisfactory size, exhibiting minor distortions, while t_{NW} based on the PC can show serious size distortions. For example, the average of the sizes over $T = 200, 500, 1000$ of t_{NW} based on the debiased SOFAR for the exponents $\{0.9, 0.8\}$, $\{0.7, 0.6\}$ and $\{0.5, 0.4\}$ are 6.2%, 6.1% and 5.7%, while that of t_{NW} based on PC are 6.7%, 7.2% and 9.2%, respectively.

To conclude, unless all the factors of the model are relatively strong, for (the estimation and) the inference on factor loadings the t-statistic based on the debiased SOFAR is preferred to that of the PC. In particular, when the data are expected to be serially correlated, the serial correlation robust t-statistic, t_{NW}, based on the debiased SOFAR is recommended to use, since the associated test based on the PC estimator can exhibit serious size distortions.

Table 2: Size of t-tests for $\hat{b}_{N2}$

Panel A: i.i.d. factors and errors

$\{\alpha_1, \alpha_2\}$	$\{0.9, 0.8\}$		$\{0.7, 0.6\}$		$\{0.5, 0.4\}$	
T, *t-statistic*	t_{iid}	t_{NW}	t_{iid}	t_{NW}	t_{iid}	t_{NW}
Debiased SOFAR						
100	6.6	7.4	5.7	6.7	7.3	7.8
200	6.7	6.7	6.2	6.6	6.1	6.3
500	6.0	6.3	6.2	6.4	5.4	5.9
1000	5.4	5.6	6.0	6.0	4.7	5.1
PC						
100	7.5	8.0	6.6	7.3	10.6	11.0
200	6.8	7.0	7.3	7.4	8.6	9.1
500	6.4	6.6	7.0	7.3	7.7	8.6
1000	5.3	5.7	6.7	6.5	7.7	7.6

Panel B: Serially correlated factors and errors

$\{\alpha_1, \alpha_2\}$	$\{0.9, 0.8\}$		$\{0.7, 0.6\}$		$\{0.5, 0.4\}$	
T, *t-statistic*	t_{iid}	t_{NW}	t_{iid}	t_{NW}	t_{iid}	t_{NW}
Debiased SOFAR						
100	8.2	8.1	8.2	8.0	7.7	7.6
200	7.1	6.5	7.6	6.5	7.5	6.8
500	7.2	5.8	7.1	6.0	5.9	4.7
1000	7.4	6.4	7.7	6.0	6.4	5.6
PC						
100	8.2	8.1	8.5	9.0	11.1	10.1
200	7.5	6.9	8.2	7.6	10.7	9.9
500	7.9	6.7	8.2	6.9	9.9	8.8
1000	8.0	6.5	8.7	7.0	10.9	9.0

Notes: The data is generated as $x_{ti} = \sum_{k=1}^{r} b_{ik} f_{tk} + \sqrt{\theta} e_{ti}$, $t = 1, .., T, i = 1, .., N$. The factor loadings b_{ik} and factors f_{tk} are formed such that $N^{-1} \sum_{i=1}^{N} b_{ik} b_{i\ell} = 1\{k = \ell\}$ and $T^{-1} \sum_{t=1}^{T} f_{tk} f_{t\ell} = 1\{k = \ell\}$, by applying Gram–Schmidt orthonormalization to b_{ik}^* and f_{tk}^*, respectively, where $b_{ik}^* \sim i.i.d.N(0,1)$ for $i = 1, \ldots, N_k$ and $b_{ik}^* = 0$ for $i = N_k + 1, \ldots, N$ with $N_k = \lfloor N^{\alpha_k} \rfloor$, and $f_{tk}^* = \rho_{fk} f_{t-1,k}^* + v_{tk}$ with $v_{kt} \sim i.i.d.N(0, 1 - \rho_{fk}^2)$ and $f_{0k}^* \sim i.i.d.N(0,1)$. The idiosyncratic errors e_{ti} are generated by $e_{ti} = \rho_e e_{t-1,i} + \varepsilon_{ti}$, where $\varepsilon_{ti} \sim i.i.d.N(0, 1 - \rho_e^2)$. b_{ik} are drawn once and fixed over the replications. We set $r = 2$, $\theta = 0.5$. For Panel A, we set $\rho_{fk} = 0$ and $\rho_e = 0$ and $\rho_{fk} = 1/4$ and $\rho_e = 1/4$ for Panel B. The model is estimated by debiased SOFAR and PC methods. t_{iid} and t_{NW} are t-statistics for $H_0 : b_{2N} = 0$ assuming i.i.d. errors and serially correlated errors, respectively. The null hypothesis is rejected when the absolute value of the t-statistic exceeds 1.96. The reported size is based on 2000 replications.