Article:

Uematsu, Yoshimasa and Yamagata, Takashi orcid.org/0000-0001-5949-8833 (2021) Inference in Sparsity-Induced Weak Factor Models. Journal of Business and Economic Statistics. ISSN 0735-0015

https://doi.org/10.1080/07350015.2021.2003203

Inference in Sparsity-Induced 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

October 28, 2021

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 debiased version of the SOFAR estimator of Uematsu and Yamagata (2021), but is also applicable to the principal component (PC) estimator. After the factor selection, the *re-sparsified* SOFAR and *sparsified* PC estimators are proposed and their consistency is established. Finite sample evidence supports the theoretical results. We apply our method to the FRED-MD dataset of macroeconomic variables and the monthly firm-level excess returns which constitute the S&P 500 index. The results give very strong statistical evidence of sparse factor loadings under the identification restrictions and exhibit clear associations of factors and categories of the variables. Furthermore, our method uncovers a very weak but statistically significant factor in the residuals of Fama-French five factor regression.

*Correspondence: Yoshimasa Uematsu, Department of Economics and Management, Tohoku University, 27-1 Kawauchi, Aobaku, Sendai 980-8576, Japan (E-mail: yoshimasa.uematsu.e7@tohoku.ac.jp).

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

1 Introduction

Factor models have become increasingly important tools 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.

Recently, Uematsu and Yamagata (2021) have considered estimation of the *sparsity-induced weak factor (sWF) models* by extending the *sparse orthogonal factor regression* (SOFAR) of Uematsu et al. (2019). The key assumption of the model is that the loading matrix is sparse under the PCA restriction. If it is true, the SOFAR can more efficiently estimate the sWF models than the PC. In this paper, we will propose an inferential method and its theory for testing sparsity of the loading matrix.

1.1 Sparsity-induced weak factor models

Suppose that a vector of zero-mean stationary time series $\mathbf{x}_t \in \mathbb{R}^N$, $t = 1, \dots, T$, is generated from the factor model $\mathbf{x}_t = \mathbf{B}^*\mathbf{f}_t^* + \mathbf{e}_t$, where $\mathbf{B}^* = (\mathbf{b}_1^*, \dots, \mathbf{b}_r^*) \in \mathbb{R}^{N\times r}$ with $\mathbf{b}_k^* \in \mathbb{R}^N$ is a matrix of deterministic factor loadings which has full column rank, $\mathbf{f}_t^* \in \mathbb{R}^r$ is a vector of zero-mean latent factors, and $\mathbf{e}_t \in \mathbb{R}^N$ is an idiosyncratic error vector independent of $\mathbf{f}_t^*$. For a while suppose r is given. Let $\mathbf{\Sigma}_x = \mathbb{E}[\mathbf{x}_t\mathbf{x}_t']$, $\mathbf{\Sigma}_f^* = \mathbb{E}[\mathbf{f}_t^*\mathbf{f}_t^{*\prime}]$, and $\mathbf{\Sigma}_e = \mathbb{E}[\mathbf{e}_t\mathbf{e}_t']$ with assuming $\mathbf{\Sigma}_f^*$ is positive definite and all the eigenvalues of $\mathbf{\Sigma}_e$ are bounded away from zero and from above (uniformly in N). Then, the covariance matrix of data is expressed as $\mathbf{\Sigma}_x = \mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime} + \mathbf{\Sigma}_e$ and $\lambda_k(\mathbf{\Sigma}_x) \asymp \lambda_k(\mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime})$ for each $k = 1, \dots, r$, where $\lambda_k(\cdot)$ denotes the kth largest eigenvalue.

In the studies on high-dimensional factor models which employ the principal components (PC) estimator, including Connor and Korajczyk (1986, 1993), Stock and Watson (2002a,b), Bai and Ng (2002, 2006, 2013), Bai (2003) and Fan et al. (2018), it is typically assumed

$\lambda_k(\mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime}) \asymp N$ for all $k = 1, \ldots, r$, which is fairly restrictive. We call the models with this condition the *strong factor (SF) models*. The SF model suggests a large gap between the rth and the $(r+1)$th eigenvalues. However, in representative financial and economic data sets such a gap is often missing; for example, see the discussion for Figure 11 in Fan et al. (2013) by Onatski. Inspired by the original *approximate factor model* proposed by Chamberlain and Rothschild (1983), we characterize the "weakness" of the factor model by $\lambda_k(\mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime}) \asymp N^{\alpha_k}$ with $0 < \alpha_k \leq 1$ for each $k = 1, \ldots, r$. Following De Mol et al. (2008), (Onatski, 2012, p.246) and Uematsu and Yamagata (2021), we call it the *weak factor (WF) model* in this paper. The WF models allow different divergence rates of the signal eigenvalues possibly slower than N. This structure is thought to be more appropriate for modeling the financial and economic variables. Very recently, the weak factor models have started to be used in the finance literature; see Anatolyev and Mikusheva (2021), Giglio et al. (2021), and Kleibergen et al. (2021), for example.

To separately identify factors and factor loadings, we employ a specific rotation that is also used for the PC estimator: $\mathbf{B}^*\mathbf{f}_t^* = \mathbf{B}^*\mathbf{H}^{-1}\mathbf{H}\mathbf{f}_t^* = \mathbf{B}^0\mathbf{f}_t^0$, where $\mathbf{B}^{0\prime}\mathbf{B}^0 = \text{diag}(\mathbf{b}_1^{0\prime}\mathbf{b}_1^0, \ldots, \mathbf{b}_r^{0\prime}\mathbf{b}_r^0)$ and $\mathbb{E}[\mathbf{f}_t^0\mathbf{f}_t^{0\prime}] = \mathbf{I}_r$. Then, we obtain the model of our interest:

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

with which we have $\lambda_k(\mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime}) = \lambda_k(\mathbf{B}^0\mathbf{B}^{0\prime}) = \lambda_k(\mathbf{B}^{0\prime}\mathbf{B}^0) = \mathbf{b}_k^{0\prime}\mathbf{b}_k^0$ for $k = 1, \ldots, r$. The last equality is due to the specific choice of the rotation matrix, $\mathbf{H}$. Our approach links the degree of sparsity in $\mathbf{b}_k^0$ to the divergence rate of $\lambda_k(\mathbf{\Sigma}_x)$. Namely, suppose that $\mathbf{b}_k^0$ contains only $\lfloor N^{\alpha_k} \rfloor$ nonzero elements, which naturally entails $\mathbf{b}_k^{0\prime}\mathbf{b}_k^0 \asymp N^{\alpha_k}$, and hence $\lambda_k(\mathbf{B}^*\mathbf{\Sigma}_f^*\mathbf{B}^{*\prime}) \asymp N^{\alpha_k}$. Consequently we obtain $\lambda_k(\mathbf{\Sigma}_x) \asymp N^{\alpha_k}$. It is called the *sparsity-induced weak factor (sWF) model*. Freyaldenhoven (2021) considers a similar model and propose a method of determining the number of relevant factors using the PC estimator. Note that "sparse factor models" are not new in the literature. For instance, several authors, including Wang (2008), Cheng et al. (2016), Choi et al. (2018), and Choi et al. (2021), have considered sparse factor loadings, but their focus is different from ours and their modeling frameworks do not necessarily connect with the sWF structure.

Three remarks may be worth noting. First, we have $\mathbf{b}_k^{*\prime}\mathbf{b}_k^* \neq \lambda_k(\mathbf{\Sigma}_x)$ in general, which means that the degree of sparsity in $\mathbf{B}^*$ cannot be straightforwardly linked to the divergence rate of $\lambda_k(\mathbf{\Sigma}_x)$. Second, since the sparsity is rotation variant (i.e., it must generally mean that a rotation can raise or reduce the degree of sparsity), starting from a sparse $\mathbf{B}^*$ does not make much sense for our purpose. Third, we are cautioned that the definition of "weak factor" varies in the literature. For example, Onatski (2012), Bryzgalova (2016), Lettau and Pelger (2020) assume non-diverging factors (i.e. $\alpha_r = 0$), which Chamberlain and Rothschild (1983) and we exclude. See also Chudik et al. (2011) for categorizing the factors according to the values of the exponents.

1.2 Empirical evidence of the sWF models

Influential empirical studies often give implicit yet strong evidence of sparse $\mathbf{B}^0$ in the sWF models. For example, in order to analyze the characteristics of the factor $\mathbf{f}_t^0$, Stock and Watson (2002b) and Ludvigson and Ng (2009) investigate the association of f_{tk}^0 to groups of the variables for each $k = 1, \dots, r$. Specifically, they extract the PC factors from standardized N macroeconomic variables, then run N time-series regressions of the variables on each of the extracted PC factors to report N values of R^2s. They find high R^2 values for a small number of variables while the rest are very close to zero. Using similar macroeconomic variables, Uematsu and Yamagata (2021) directly estimate the sparse $\mathbf{B}^0$ by the adaptive SOFAR method, which embodies characteristics of factors. Observe that these studies provide point estimation of a sparse loading matrix, but formal statistical inference on zeros in $\mathbf{B}^0$ has not been considered.

1.3 Testing sparsity in sWF models

In this paper, we consider statistical inference on the sparsity in $\mathbf{B}^0$, regarding the sWF model (with $\alpha_k < 1$ for some k) as the model under the null hypothesis of sparse $\mathbf{B}^0$. This is important, because the WF structure, $\lambda_k(\mathbf{\Sigma}_x) \asymp N^{\alpha_k}$ with $\alpha_k < 1$ for some $k \leq r$, can be induced by *non-sparse* factor loadings, as the earlier discussion implies. For instance, it is the case when a factor affects all the variables at similar strengths thinly. In a conventional setting, we may consider a test for the null hypothesis of $b_{ik}^0 = 0$ for all $(i, k) \in \mathcal{H}$ for given subset $\mathcal{H} \subset [N] \times [r]$, where $[N] := \{1, \ldots, N\}$. A standard testing procedure will work when $\mathcal{H}$ is small. For our purpose, however, it is interesting to see whether $b_{ik}^0 = 0$ for each $(i, k) \in [N] \times [r]$, which makes it difficult to apply the standard testing procedure.

1.4 Toward global inferences

Let $\mathcal{S} \subset [N] \times [r]$ denote the support (i.e., the index set of nonzero elements) of $\mathbf{B}^0$. Given $\mathcal{H}$ as above, the hypothesis of Section 1.3 is formally rewritten as

$$H_0 : (i, k) \in \mathcal{S}^c \text{ for all } (i, k) \in \mathcal{H} \text{ vs. } H_1 : (i, k) \in \mathcal{S} \text{ for some } (i, k) \in \mathcal{H}. \tag{2}$$

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 zeros, especially when $\mathcal{H}$ is very large. Alternatively, it is more interesting to investigate whether *each* entry in $\mathbf{B}^0$ is significantly null or not. To this end, we consider a *multiple testing* for a sequence of pairs of hypotheses

$$H_0^{(i,k)} : (i, k) \in \mathcal{S}^c \text{ vs. } H_1^{(i,k)} : (i, k) \in \mathcal{S} \text{ for each } (i, k) \in [N] \times [r]. \tag{3}$$

In such multiple testing problems, it is important to control the type I error 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, the FWER control is too stringent and will lead to a very conservative variable selection, especially in high dimensions. Instead of the FWER, it is more suitable to control the *false discovery rate* (FDR) as another measure of type I errors. The FDR was first introduced by Benjamini and Hochberg (1995) and is defined as the expectation of the *false discovery proportion* (FDP):

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

where $\widehat{\mathcal{S}} \subset \{1, \ldots, N\} \times \{1, \ldots, r\}$ is a set of indexes discovered 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 for (3) can be called a *global* inference, in contrast with the local inference for (2).

1.5 Contributions

In light of the recent development of global inferences described above, we first propose the *debiased* SOFAR estimator of the sparse loadings $\mathbf{B}^0$, learned from Javanmard and Montanari (2014), van de Geer et al. (2014), and Zhang and Zhang (2014) in a linear regression context, and establish its asymptotic normality. In addition, we show that the PC estimator is asymptotically normal even for the sWF models, which is an extension of Bai (2003).

Building upon the asymptotic normality of the loading estimators, we consider multiple testing (3) for the sequence, $H_0^{(i,k)} : b_{ik}^0 = 0$ vs. $H_1^{(i,k)} : b_{ik}^0 \neq 0$ for $i = 1, \ldots, N$ and $k = 1, \ldots, r$, 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 finite sample distribution of the debiased SOFAR estimator is approximated by normal distribution very well while that of 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 their consistency. Since these estimators inherit the asymptotic normality of the debiased SOFAR and PC estimators, they can be attractive alternatives to the adaptive SOFAR estimator.

We apply our procedure to the FRED-MD dataset of macroeconomic variables and the firm-level excess returns consisting of the S&P500 index. The results give very strong statistical evidence of sparse $\mathbf{B}^0$ under the identification restrictions and exhibit statistically significant associations of factors and categories of the variables. In addition, applying our method to the residuals of the regressions of the firm-level excess returns on the Fama and French (2015) five factors uncovers a hidden, very weak but statistically significant factor.

1.6 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 sequences 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 sWF 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 data set and firm-level excess returns. Section 7 concludes. All the proofs of our theoretical results and additional experimental and empirical results are collected in Supplementary Material.

2 Sparsity-Induced Weak Factor Models

Consider the factor model in (1) more precisely. Stacking the vectors vertically like $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)'$, $\mathbf{F}^0 = (\mathbf{f}_1^0, \ldots, \mathbf{f}_T^0)'$, and $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_T)'$, we rewrite it as the matrix form

$$\mathbf{X} = \mathbf{F}^0\mathbf{B}^{0\prime} + \mathbf{E} = \mathbf{C}^0 + \mathbf{E}, \tag{4}$$

where $\mathbf{C}^0$ is called the matrix of common components. By the construction, the model satisfies the restrictions: $\mathbb{E}\,\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$ and $\mathbf{B}^{0\prime}\mathbf{B}^0$ is a diagonal matrix. Then the covariance matrix reduces to

$$\mathbf{\Sigma}_x = \mathbf{B}^0\mathbf{B}^{0\prime} + \mathbf{\Sigma}_e.$$

As discussed in Introduction (Section 1.1), we consider the sWF models. Specifically, we assume 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 $N_k := \lfloor N^{\alpha_k} \rfloor$ for $k \in [r]$, where $1 \geq \alpha_1 \geq \cdots \geq \alpha_r > 0$ and exponents α_k's are unknown. Note that N_r must diverge since $\alpha_r > 0$ and $N \to \infty$.

By the sparsity assumption and the diagonality of $\mathbf{B}^{0\prime}\mathbf{B}^0$, there exist some positive constants $d_1, \ldots, d_r$ such that

$$\mathbf{B}^{0\prime}\mathbf{B}^0 = \operatorname{diag}\left(d_1^2 N_1, \ldots, d_r^2 N_r\right)$$

and $d_1^2 N_1 \geq \cdots \geq d_r^2 N_r > 0$. Then, under the assumption $\max_N \lambda_1(\mathbf{\Sigma}_e) < \infty$, we have

$$\lambda_k(\mathbf{\Sigma}_x) \begin{cases} \asymp \lambda_k(\mathbf{B}^0\mathbf{B}^{0\prime}) = \lambda_k(\mathbf{B}^{0\prime}\mathbf{B}^0) = d_k^2 N_k & \text{for } \ k \in [r], \\ = O(1) & \text{for } \ k \in [N]\backslash[r]. \end{cases}$$

This specification fulfills the requirement of the WF structure, $\lambda_k(\mathbf{\Sigma}_x) \asymp N^{\alpha_k}$ for all $k = 1, \ldots, N$ with $\alpha_{r+1} = \cdots = \alpha_N = 0$. Finally define $\mathcal{S} := \operatorname{supp}(\mathbf{B}^0) \subset [N] \times [r]$ and $s := |\mathcal{S}| = \sum_{k=1}^r N_k$. Thus $|\mathcal{S}^c| = Nr - s$.

3 Inferential Methodology

We develop a new inferential framework for the sWF models defined in Sections 1.1 and 2. 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 Section 4.

For the sWF models, we first need to estimate the number of factors, r. Uematsu and Yamagata (2021) show that the method of Onatski (2010) asymptotically works well under some conditions (see Section 4 for a theoretical summary). Namely, for given $\delta > 0$ and $k_{\max} \in \mathbb{N}$, define

$$\hat{r}(\delta) = \max\left\{k = 1, \ldots, k_{\max} - 1 : \lambda_k - \lambda_{k+1} \geq \delta\right\}, \tag{5}$$

where λ_k is the kth largest eigenvalue of $(N \vee T)^{-1}\mathbf{X}\mathbf{X}'$. In practice, δ should appropriately be predetermined. Onatski (2010) suggests the *edge distribution* (ED) method based on a calibration; see the paper for full details.

For such given $\hat{r}$, the SOFAR estimator proposed by Uematsu et al. (2019) and Uematsu and Yamagata (2021) is defined as

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

where $\eta > 0$ is a regularization coefficient. Setting $\eta = 0$ eventuates in the PC estimator. The SOFAR estimator can be more efficient than the PC estimator for the sWF models because it provides sparse estimates, whereas the PC does not.

Remark 1. The SOFAR problem (6) is nonconvex due to two constraints, but is numerically stable. In fact, the SOFAR procedure is composed of two steps: first an initial consistent estimator is obtained by a convex optimization, and then the SOFAR estimator is computed by the nonconvex optimization in the shrinking neighborhood of the initial estimator. As a result, it is stable and asymptotically globally optimal. For further information, see Uematsu et al. (2019).

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) have proposed a 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

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

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

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

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 an r-dimensional 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 2. Unlike the debiased lasso for high-dimensional linear models, the debiased SOFAR for the sWF 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.

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 $\alpha_k < 1$ for some $k = 1, \dots, r$. In Section 4.1, we will derive the asymptotic normality and consider the theoretical properties through comparison with the debiased SOFAR.

3.2 Asymptotic t-test

Under some conditions, each row of the debiased SOFAR estimator (9) can admit asymptotic normality:

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

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 $O(T^{1/3})$ (Andrews, 1991), for instance. Once the consistent estimator is obtained, a conventional asymptotic 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 Introduction, we consider a *multiple testing* of a sequence of hypotheses (3), which is rephrased as

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

For each (i, k), the t-statistic is defined as

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

where $\hat{\sigma}^2_{ik}$ is the kth diagonal element of $\widehat{\mathbf{\Gamma}}_i$ introduced in (11). Repeating the t-test with a "conventional" critical value, like 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) with critical value $\mathtt{t}$.

1. For any target FDR level $q \in [0, 1]$, define

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

   where $\bar{\mathtt{t}} \leq \sqrt{2\log(Nr)}$ is some positive value and $G(\mathtt{t}) = 2(1 - \Phi(\mathtt{t}))$ 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 $\widehat{\mathcal{S}}$ 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

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

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 $\hat{b}^{\mathsf{d}}_{ik}$ in (13). The associated *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.

For the sake of convenience, define $n = N \wedge T$. Then we have $N = N(n) \to \infty$ and $T = T(n) \to \infty$ as $n \to \infty$. Throughout this section, set $\eta_n \asymp T^{1/2} \log^{1/2}(N \vee T)$ in optimization (6). Furthermore, following Vershynin (2018), we introduce a *sub-Gaussian* random variable: a random variable $Y \in \mathbb{R}$ is said to be sub-Gaussian and denoted as $Y \sim \mathrm{subG}$ if there exists some constant $c > 0$ such that $\mathbb{P}(|Y| \geq y) \leq 2\exp(-y^2/c)$ for all $y \geq 0$. Throughout the paper, including all the proofs in Supplementary Material, $\nu > 0$ is a fixed large constant, and n is sufficiently large.

Assumption 1 (Latent factors)**.** The factor matrix $\mathbf{F}^0 = (\mathbf{f}^0_1, \ldots, \mathbf{f}^0_T)'$ is specified as the vector linear process $\mathbf{f}^0_t = \sum_{\ell=0}^{\infty} \mathbf{\Psi}_\ell \boldsymbol{\zeta}_{t-\ell}$, where $\boldsymbol{\zeta}_t = (\zeta_{t1}, \ldots, \zeta_{tr})'$ with $\{\zeta_{tk}\}_{t,k}$ are i.i.d. subG that has $\mathbb{E}\,\zeta^2_{tk} = 1$ and $\sum_{\ell=0}^{\infty} \mathbf{\Psi}_\ell \mathbf{\Psi}'_\ell = \mathbf{I}_r$. Moreover, there exist constants $C_f > 0$ and $\ell_f \in \mathbb{N}$ such that $\|\mathbf{\Psi}_\ell\|_2 \leq C_f \ell^{-(\nu+2)}$ for all $\ell \geq \ell_f$.

Assumption 2 (Factor loadings)**.** Each column $\mathbf{b}^0_k$ of $\mathbf{B}^0$ has the sparsity $N_k = \lfloor N^{\alpha_k} \rfloor$ with $0 < \alpha_r \leq \cdots \leq \alpha_1 \leq 1$ and $\mathbf{B}^{0\prime}\mathbf{B}^0 = \mathrm{diag}\{d_1^2 N_1, \ldots, d_r^2 N_r\}$ with $0 < d_r N_r^{1/2} \leq \cdots \leq d_1 N_1^{1/2}$. For k such that $\alpha_k = \alpha_{k-1}$, it holds that $d^2_{k-1} - d^2_k \geq \kappa^{1/2} d^2_{k-1}$ for some constant $\kappa > 0$.

Assumption 3 (Idiosyncratic errors)**.** The error matrix $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_T)'$ is independent of $\mathbf{F}^0$ and is specified as the vector linear process $\mathbf{e}_t = \sum_{\ell=0}^{\infty} \mathbf{\Phi}_\ell \boldsymbol{\varepsilon}_{t-\ell}$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{t1}, \ldots, \varepsilon_{tN})'$

with $\{\varepsilon_{ti}\}_{t,i}$ are i.i.d. subG and $\mathbf{\Phi}_0$ is a nonsingular, lower triangular matrix. Moreover, there exist constants $C_e > 0$ and $\ell_e \in \mathbb{N}$ such that $\|\mathbf{\Phi}_\ell\|_2 \le C_e \ell^{-(\nu+2)}$ for all $\ell \ge \ell_e$.

Assumptions 1–3 are the same as in Uematsu and Yamagata (2021). Assumptions 1 and 3 specify the stochastic processes $\{\mathbf{f}_t^0\}$ and $\{\mathbf{e}_t\}$, respectively, to be the stationary vector linear processes satisfying the summability condition $\sum_{\ell=0}^{\infty}(\|\mathbf{\Psi}_\ell\|_2 + \|\mathbf{\Phi}_\ell\|_2) < \infty$. The decaying rates are at most polynomial, which includes a wide range of multivariate weakly dependent processes. Under this condition we can achieve the concentration inequalities. Assumption 2 is key to our analysis and provides the sWF models. The sparsity in $\mathbf{B}^0$ makes the divergence rate of $\lambda_k(\mathbf{B}^{0\prime}\mathbf{B}^0)$ possibly slower than N for each $k = 1, \dots, r$.

We summarize some preliminary results obtained by Uematsu and Yamagata (2021), which will be a basis for the inferential theory developed in the following subsections. In addition to Assumptions, we impose the condition that restricts the class of sWF models:

$$\alpha_1 + (1 \vee \tau)/2 < 3\alpha_r/2 + \tau/2, \tag{16}$$

where τ is such that $T = \lfloor N^\tau \rfloor$. Roughly, this condition excludes the sWF models with α_1 and α_r being far away from each other, especially when τ is small. Under Assumptions 1–3 with condition (16), the rth largest eigenvalue of $(N \vee T)^{-1}\mathbf{X}\mathbf{X}'$ diverges while the $(r+1)$th largest bounded with high probability. This property justifies the asymptotic validity of (5) with any fixed positive constant δ; we have $\hat{r}(\delta) = r$ eventually with high probability. Moreover, the rates of convergence of the SOFAR estimator $(\widehat{\mathbf{F}}, \widehat{\mathbf{B}})$ and the PC estimator $(\widehat{\mathbf{F}}^{\mathsf{PC}}, \widehat{\mathbf{B}}^{\mathsf{PC}})$ are derived for the sWF models under the same assumptions: $\|\widehat{\mathbf{F}} - \mathbf{F}^0\|_{\mathrm{F}} + \|\widehat{\mathbf{B}} - \mathbf{B}^0\|_{\mathrm{F}} \lesssim R_n$ and $\|\widehat{\mathbf{F}}^{\mathsf{PC}} - \mathbf{F}^0\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}^{\mathsf{PC}} - \mathbf{B}^0\|_{\mathrm{F}} \lesssim R_n^{\mathsf{PC}}$ with high probability, where

$$R_n = \frac{N_1^{3/2} T^{1/2} \log^{1/2}(N \vee T)}{N_r(N_r \wedge T)}, \quad R_n^{\mathsf{PC}} = R_n(1 + \gamma_n), \quad \gamma_n = \frac{N^{1/2}(N_r \wedge T)^{1/2}}{N_1^{1/2} T^{1/2}}.$$

4.1 Theory on the asymptotic linear representation

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 1–3 and (16) hold,* *then the debiased SOFAR estimator has the asymptotic linear representation*

$$\sqrt{T}\left(\widehat{\mathbf{b}}_i^{\mathsf{d}} - \mathbf{b}_i^0\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_{ti}\mathbf{f}_t^0 + \mathbf{r}_i, \tag{17}$$

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

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

Theorem 2 (PC)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$*. If Assumptions 1–3 and (16) hold, then the PC estimator has an asymptotic linear representation*

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

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

$$\max_{i\in[N]} \|\mathbf{r}_i^{\mathsf{PC}}\|_{\max} \lesssim \delta_1(1+\gamma_n).$$

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 clarity 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.

The upper bound of the estimation error $\mathbf{r}_i$ of the debiased SOFAR can decay faster than that of the PC estimator. 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. 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 (17) and (18) 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 a stronger assumption on $\{e_{ti}\}$.

4.2 Theory on the global inference for the loadings

We establish the theoretical results for the FDR control and power guarantee explored in Section 3.3. Although we focus on the procedure 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 condition on the error term.

Assumption 4. 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 sub-Gaussian with its variance σ_i^2 and covariance $\sigma_{ij} = \mathbb{E}[e_{ti}e_{tj}]$.

Assumption 5. For the correlation of e_{ti} and e_{tj}, $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$ for all $i \neq j$, there exists a partition $\{P_1, P_2\}$ of set $\{(i,j) \in [N] \times [N] : i \neq j\}$ such that

$$|\rho_{ij}| \in \begin{cases} \left[0, c/\log^{\xi} N\right] & \text{for } (i,j) \in P_1 \text{ with } |P_1| = N^2 - N - |P_2|, \\ \left(c/\log^{\xi} N, \bar{\rho}\right] & \text{for } (i,j) \in P_2 \text{ with } |P_2| = O(N^2/\log^{\xi} N) \end{cases}$$

for some fixed constants $c > 0$, $\xi > 1$, and $\bar{\rho} \in (0,1)$.

The independence of Assumption 4 is required for technical reasons. Assumptions 1 and 4 make $\{e_{ti} f_{tk}\}$ form a martingale difference sequence (MDS) for each (i,k). This enables us to apply a strong approximation of sum of a vector MDS to a Gaussian random vector. Furthermore, since $\mathbf{\Gamma}_i = \sigma_i^2 \mathbf{I}_r$ in (10) under these assumptions, we can avoid dealing with a complicated estimation error bound for the HAC estimator. Assumption 5 stipulates the correlation structure of e_{ti} and e_{tj} over all $i \neq j$.

First we have the result of the FDR control of $\widehat{\mathcal{S}}$.

Theorem 3 (FDR control)**.** *Suppose* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$ *and* $\delta_1 = O(n^{-c})$ *for some* $c > 0$*. If Assumptions 1, 2, 4, and 5 with (16) hold, then for any fixed* $q \in [0,1]$*, the FDR of* $\widehat{\mathcal{S}}$ *obtained by Procedure 1 with setting* $\bar{\mathtt{t}} = \sqrt{2\log(Nr) - a_1 \log\log(Nr) + a_2}$*, where* $a_1 > 3$ *and* $a_2 > 0$ *are arbitrary constants, 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 rises.

Assumption 6. For $\mathcal{S} = \text{supp}(\mathbf{B}^0)$, 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* $\mathbf{F}^{0\prime}\mathbf{F}^0/T = \mathbf{I}_r$ *and* $\delta_1 = O(n^{-c})$ *for some* $c > 0$*. If Assumptions 1, 2, 4, and 6 with (16) 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.

Remark 5. In line with the literature on the adaptive lasso, the asymptotic normality of the adaptive SOFAR estimator could also be established for the nonzero elements. However, we do not pursue this direction, given the critique by, e.g. Leeb and Pötscher (2008) and Pötscher and Leeb (2009), that the property requires a perfect model selection, which is implied by a "beta-min" condition, and is very sensitive to a violation of the condition (see Chernozhukov et al., 2015, Ch. 6). The same criticism could apply to the adaptive SOFAR estimator. Meanwhile, our procedure based on the debiased SOFAR does not include any model selection step. Note that Assumption 6 is used only for a power guarantee.

These properties are apparently inherited by the re-sparsified SOFAR estimator defined in (15). The next theorem states additional properties of this estimator.

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* $\|\widehat{\mathbf{B}}^{\mathsf{r}} - \mathbf{B}^0\|_{\max} \to_p 0$ *and* $\sqrt{T}(\hat{b}_{ik}^{\mathsf{r}} - b_{ik}^0) \to_d N(0, \sigma_i^2)$ *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, by comparing it with that of the PC estimator by means of Monte Carlo experiments. First, we examine the quality of the standard normal approximation of the distribution of a t-statistic for a factor loading. 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):

$$x_{ti} = \sum_{k=1}^{r} b_{ik}^0 f_{tk}^0 + \sqrt{\theta} e_{ti}, \quad (t,i) \in [T] \times [N].$$

The factor loadings b_{ik}^0 and factors f_{tk}^0 are formed such that $N^{-1}\sum_{i=1}^N b_{ik}^0 b_{i\ell}^0 = 1\{k=\ell\}$ and $T^{-1}\sum_{t=1}^T f_{tk}^0 f_{t\ell}^0 = 1\{k=\ell\}$, 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 $\mathrm{Var}(b_{ik}^*) = 1$. 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

$$f_{tk}^* = \rho_{fk} f_{t-1,k}^* + v_{tk}$$

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

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

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

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 for the proposed FDR controlled multiple testing procedure. First, $\mathtt{T}_0$ denotes the t-statistic which is the ratio of $\hat{b}_{ik}$ and its population standard deviation (dropping the subscripts i and k for simplicity). 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; see (11). To economize the space in what follows we report the results for the DGP with serially correlated factors and i.i.d. errors only, unless otherwise stated. The results for serially correlated errors with $\mathtt{T}_{NW}$ are qualitatively similar, and are reported in

Section C in Supplementary Material.

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 $\mathtt{T}_0$, which is $\hat{b}_{ik}$ for null $(i,k) \in \mathcal{S}^c$, scaled by its true standard deviation, referring 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. 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 model becomes 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, established as Theorems 1 and 2. Qualitatively similar results are obtained with $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$ and for serially correlated errors, which are summarized in Section C.1 in Supplementary Material.

5.2 Global inference for the loadings

Having seen the accuracy of the 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 serially correlated factors and i.i.d. errors ($\rho_{fk} = 1/4$ and $\rho_e = 0$). To assess the efficacy of the proposed method to control the FDR, we report the FDR as well as the power, based on $\mathtt{T}_{iid}$. All the combinations of $N, T \in \{100, 200, 500\}$ are considered and 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 further 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. Qualitatively similar results are obtained with $\mathtt{T}_0$ and $\mathtt{T}_{NW}$ and for serially correlated errors, which are summarized in Section C.2 in Supplementary Material.

[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,

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

where $\widehat{\mathcal{S}}^{\mathsf{PC}}$ is obtained by Procedure 1 based on $\mathtt{T}_{iid}$, which is constructed using the PC estimator. We employ the same DGP and set-up used in Section 5.2 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. The experimental results for the DGP with serially correlated errors ($\rho_{fk} = 1/4$ and $\rho_e = 1/4$) based on $\mathtt{T}_{NW}$ are qualitatively similar, which are reported in Section C.3 in Supplementary Material.

6 Empirical Applications

In this section we consider the empirical applications of the FDR controlled global inference on the factor selection, based on the SOFAR estimates. Section 6.1 considers the FRED-MD macroeconomic and financial variables. In Section 6.2 a large number of excess stock returns are analyzed. Section 6.3 investigates the residuals of the Fama and French (2015) five-factor regressions. The associated results based on the PC estimates are provided in Section D in Supplementary Material.

6.1 Macroeconomic and financial variables

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 most probably it contains five factors. The re-sparsified SOFAR loading estimate is computed, following the global inference based on the debiased BIC chosen SOFAR estimate. The t-statistics for the procedure are computed using the serial correlation robust variance covariance estimator, $\mathtt{T}_{NW}$, and we report the result for $q = 10\%$. The procedure has chosen the value of the FDR controlling threshold, $\mathtt{t}_0 = 2.05$ for the upper bound $\bar{\mathtt{t}} = \sqrt{2\log(Nr)} = 3.59$. For example, the Bonferroni correction, which aims to control the FWER below 10% gives the threshold value 3.78 for each of the $\mathtt{T}_{ik}$ tests. The large difference in values between $\mathtt{t}_0$ and the Bonferroni threshold suggests that the inferential procedure controlling FDR enjoys substantial power gain compared to the procedure controlling the FWER.

To investigate the characteristics of five common factors, we report the value of re-sparsified SOFAR loadings for the 128 series as a bar-chart in Figure 13. The variables are ordered according to their characteristics of 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 is sufficient to recognize very strong statistical evidence of sparse factor loadings under the identification restrictions and it exhibits clear associations of factors (loadings) and groups of macroeconomic variables. The first factor is significantly associated with five variable groups, G1-G5, and can be seen as a semi-global factor. Each of the remaining four factors is significantly associated with just one or two dominating groups. Specifically, we may identify the second to the fifth factor as a price factor, a housing factor, an output and income factor, and a money, credit and stock market factor, respectively. [**INSERT Figure 13**]

6.2 S&P500 firm-level excess returns

We consider firm-level monthly excess returns, which constitute the Standard & Poor's 500 Stock Index (S&P 500) on April 2018 over the period from May 1998 to April 2018, which left 376 securities. The monthly excess returns of security i for month t are computed as $r_{e,ti} = 100 \times (P_{ti} - P_{t-1,i})/P_{t-1,i} + DY_{ti}/12 - r_{ft}$, where P_{ti} is the end-of-the-month price,

DY_{ti} is the percent per annum dividend yield, and r_{ft} is the one-month US treasury bill rate chosen as the risk-free rate. P_{ti} and DY_{ti} are obtained from Datastream, and r_{ft} is obtained from Ken French's data library web page. We standardize the excess returns and denote them as $r^*_{e,ti}$.

The model with three factors is estimated by the BIC-SOFAR method, then the loading estimate is debiased and the re-sparsified SOFAR loading estimate is obtained, following the proposed procedure for the global inference. The procedure is based on $\mathtt{T}_{iid}$ and $q = 10\%$, which chooses $\mathtt{t}_0 = 1.85$, with the upper bound $\bar{\mathtt{t}} = 3.75$. Note that the Bonferroni correction to control the FWER below 10% gives the threshold value 3.92, which suggests the substantial power gain of our procedure over the methods controlling the FWER.

For a characteristics analysis of the estimated factors, all the firms are assigned to ten industrial sectors based on Industry Classification Benchmark (ICB): (0) *Oil & Gas*; (1) *Basic Materials*; (2) *Industrials*; (3) *Consumer Goods*; (4) *Health Care*; (5) *Consumer Services*; (6) *Telecommunications*; (7) *Utilities*; (8) *Financials*; (9) *Technology*. We refer to FTSE Russell for more details about ICB.

As in the previous subsection, to analyze the three common factors, we report the value of the re-sparsified SOFAR loadings for the 376 excess returns as a bar-chart in Figure 14. It shows very strong statistical evidence of sparse loadings under the identification restrictions and exhibits interesting associations of factors and industrial sectors. The first factor is the market factor, significantly affecting virtually all the firm securities. The second factor is significant for most securities of industrial sectors 5-9, while it is largely insignificant for the securities of industrial sectors 0-4. The third factor is highly significant for the securities in *Oil & Gas* sector and to lesser extent for those in industrial sectors 5-8. [**INSERT Figure 14**]

6.3 Residuals of Fama-French five-factor regressions

In this subsection, we examine regression residuals of the celebrated Fama and French (2015) five-factor model in order to see whether any additional common factors were left out. We consider firm-level monthly excess returns, which constitute the S&P500 index on April 2018, with 500 months observations back, leaving 194 securities. The firm-level excess returns are obtained and computed as explained in Section 6.2, and the five factors are obtained from the Kenneth R. French Data Library. See Fama and French (2015) for more details of the data and the regression. Specifically, we run the time series regression $r_{ti} - r_{ft} = a_i + b_i (r_{mt} - r_{ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{ti}$, where r_{ti} is the i-th security monthly return at the month t, r_{ft} is the one-month treasury bill rate, r_{mt} is the market return, SMB_t is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low B/M stocks, RMW_t is the difference between the returns on 13 diversified portfolios of stocks with robust and weak profitability, and CMA_t is the difference between the returns on diversified portfolios of the stocks of low and high investment firms, which is called conservative and aggressive.

We have applied our method to the obtained residual. Onatski's (2010) ED procedure suggests that the residuals contain only one factor. As in the previous subsection, we obtain the BIC-SOFAR estimate then compute the re-sparsified SOFAR loading estimate, using $\mathtt{T}_{iid}$ and $q = 10\%$. Only two securities are significant (with the same sign), and interestingly both of which belong to *Technology* industry.

7 Conclusion

In this paper, we have studied statistical inference for high-dimensional approximate factor models. We have considered the sparsity-induced weak factor (sWF) 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, extending Uematsu and Yamagata (2021), we have proposed the debiased SOFAR estimator of the sparse loadings in the sWF models, and established its asymptotic normality. In addition, we have shown that the PC estimator is asymptotically normal even for the sWF 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 a similar manner. We have also established its consistency. The finite sample experiment has revealed that the performance of these estimators can be superior to the SOFAR, the debiased SOFAR and the PC estimators in terms of the norm loss.

The proposed methods 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 (2002b), Ludvigson and Ng (2009), and McCracken and Ng (2016), among many others. In recent finance literature, there has 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 addressing such issues.

Supplementary Materials

The Supplementary Material consists of four sections. Section A contains the proofs of the main results. Sections B contains lemmas and their proofs. Section C contains additional experimental results. Section D contains additional empirical results.

Acknowledgments

An earlier version of this article was circulated under the title, "Inference in weak factor models". The authors thank the editor Jianqing Fan, the associate editor, and the three anonymous referees for their valuable comments and suggestions. The authors appreciate

Kun Chen giving helpful suggestions and modification of the R package, rrpack.

Funding

This work was supported by JSPS KAKENHI Grant Numbers 19K13665, 20H01484, 20H05631, 21H00700, and 21H04397.

References

Anatolyev, S. and A. Mikusheva (2021). Factor factor models with many assets: Strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics, forthcoming*.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

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

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.

Bryzgalova, S. (2016). Spurious factors in linear asset pricing models. *mimeo*.

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

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. *Review of Economic Studies 83*(4), 1511–1543.

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.

Choi, I., D. Kim, Y. J. Kim, and N.-S. Kwark (2018). A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics 33*, 355–377.

Choi, I., R. Lin, and Y. Shin (2021). Canonical correlation-based model selection for the multilevel factors. *SSRN: https://ssrn.com/abstract=3590109*.

Chudik, A., , H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

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.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*, 318–328.

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

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.

Fan, J., K. Wang, Y. Zhong, and Z. Zhu (2018). Robust high-dimensional factor models with applications to statistical machine learning. *arXiv:1808.03889v1*.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications Vol. 1 (3rd ed.)*. Wiley.

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

Freyaldenhoven, S. (2021). Factor models with local factors - determining the number of relevant factors. *Journal of Econometrics, forthcoming*.

Giglio, S., D. Xiu, and D. Zhang (2021). Identification robust testing of risk premia in finite samples. *SSRN: https://ssrn.com/abstract=3768081*.

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.

Kifer, Y. (2013). Strong approximations for nonconventional sums and almost sure limit theorems. *Stochastic Processes and their Applications 123*, 2286–2302.

Kleibergen, F., L. Kong, and Z. Zhan (2021). Identification robust testing of risk premia in finite samples. *Journal of Financial Econometrics, forthcoming*.

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.

Lettau, M. and M. Pelger (2020). Estimating latent asset-pricing factors. *Journal of Econometrics 218*, 1–31.

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.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics 168*, 244–258.

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.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). 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 (2021). Estimation of sparsity-induced 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., S. Girard, H. Nguyen, and J. Arbel (2020). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat 9*, e318.

Wang, P. (2008). Large dimensional factor models with a multi-level factor structure: Identification, estimation and inference. *Unpublished manuscript, New York University*.

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.

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 ($\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$) 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
$T = 100$										
	$\widehat{\mathbf{B}}$	174.9	177.1	180.0	215.0	225.6	236.2	217.3	229.4	247.8
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	155.1	160.5	166.7	258.3	285.9	326.3	408.5	486.8	611.9
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	164.9	171.7	170.8	277.9	293.5	331.5	449.8	558.6	631.5
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	148.1	146.1	141.5	165.0	163.3	164.9	178.0	182.4	185.9
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	161.1	159.8	148.3	190.0	173.7	172.5	229.2	296.1	209.9
$T = 200$										
	$\widehat{\mathbf{B}}$	123.0	127.1	128.9	153.8	159.1	166.0	153.8	161.9	174.9
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	109.9	114.7	118.2	185.0	203.0	229.9	292.5	345.2	433.9
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	112.5	117.7	119.2	195.1	206.8	232.1	320.7	416.4	445.6
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	103.6	103.6	100.7	119.0	115.1	113.4	132.2	129.9	129.3
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	107.4	107.9	102.9	131.4	119.8	116.0	165.4	237.6	140.7
$T = 500$										
	$\widehat{\mathbf{B}}$	76.8	80.9	82.0	94.0	101.6	106.3	96.3	106.2	108.0
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	71.2	74.0	75.0	118.5	129.6	146.3	187.9	220.6	275.7
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	72.2	75.1	75.3	123.2	131.8	147.5	203.5	267.6	282.8
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	65.7	64.6	61.2	73.9	72.6	70.4	84.5	82.8	81.9
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	67.4	66.1	61.9	79.0	74.9	71.5	101.9	150.9	87.8

Notes: For the re-sparsified estimators, 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 $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$, $\{0.7, 0.6\}$, $\{0.5, 0.4\}$.

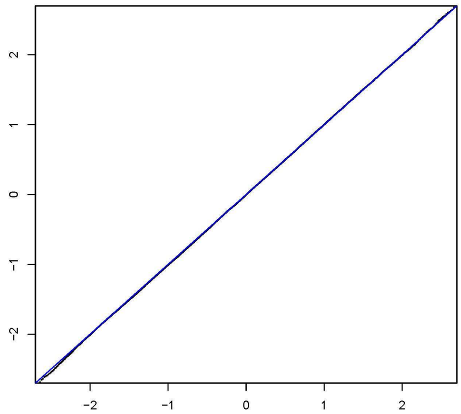

Figure 1: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

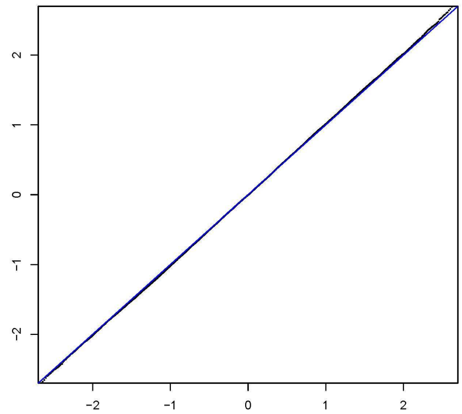

Figure 2: PC, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

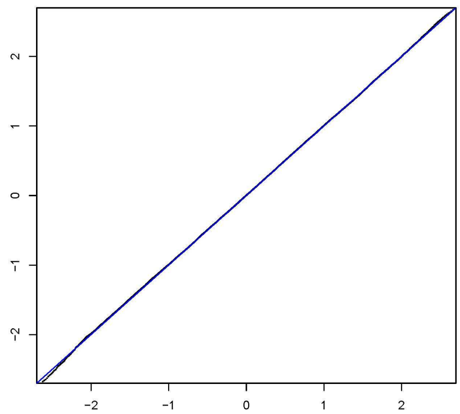

Figure 3: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

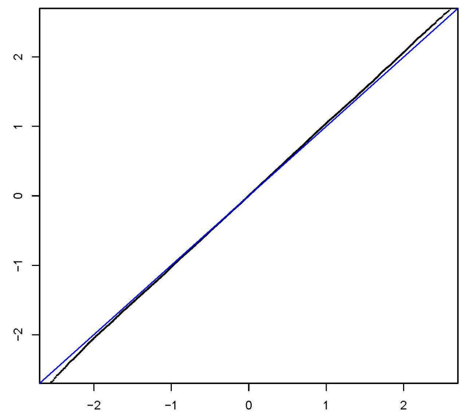

Figure 4: PC, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

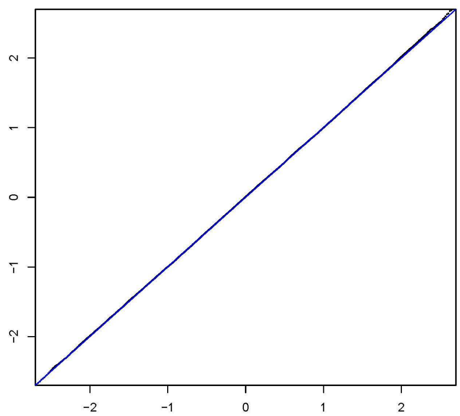

Figure 5: debiased SOFAR, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

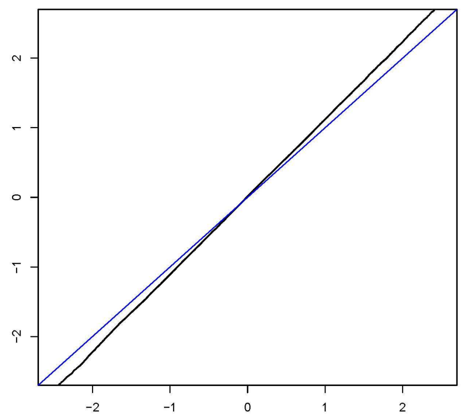

Figure 6: PC, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

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

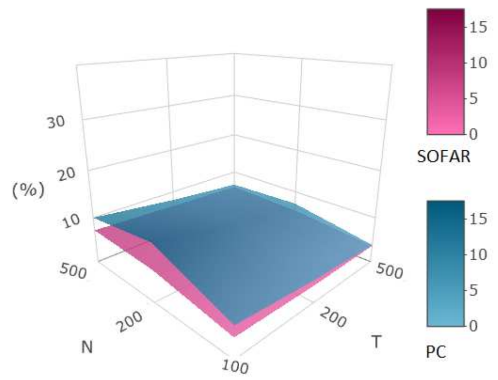

Figure 7: FDR, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

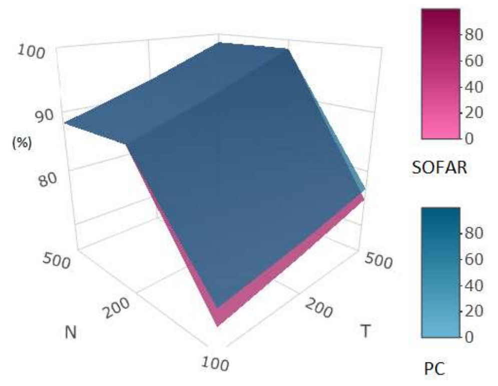

Figure 8: Power, $\{\alpha_1, \alpha_2\} = \{0.9, 0.8\}$

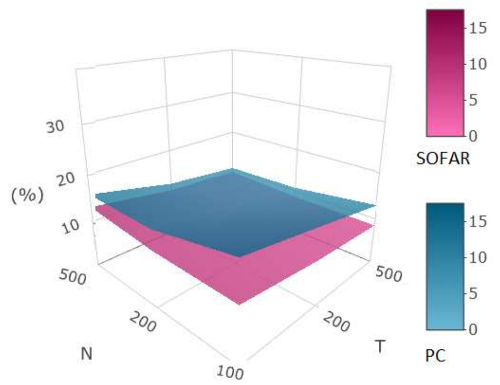

Figure 9: FDR, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

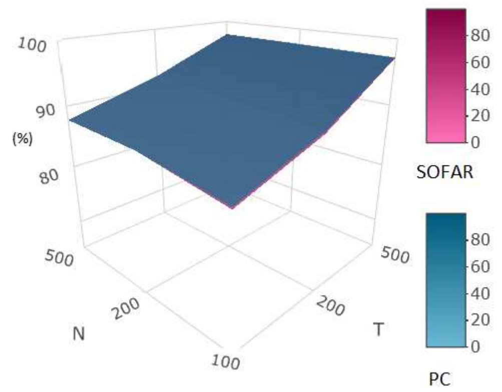

Figure 10: Power, $\{\alpha_1, \alpha_2\} = \{0.7, 0.6\}$

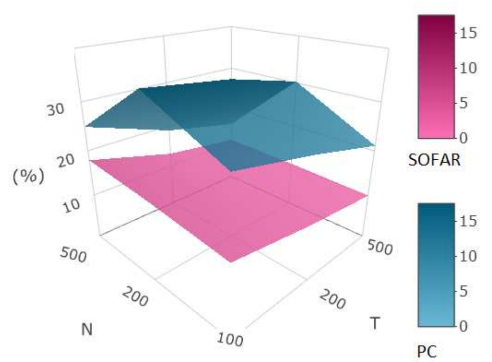

Figure 11: FDR, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

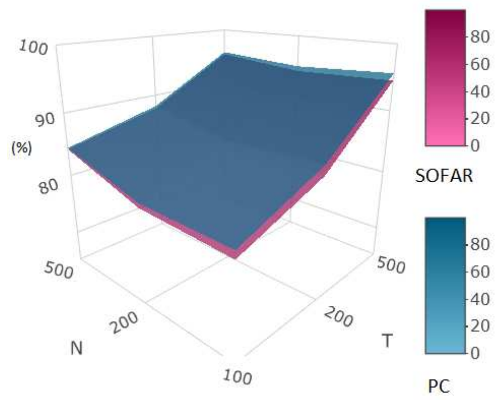

Figure 12: Power, $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$

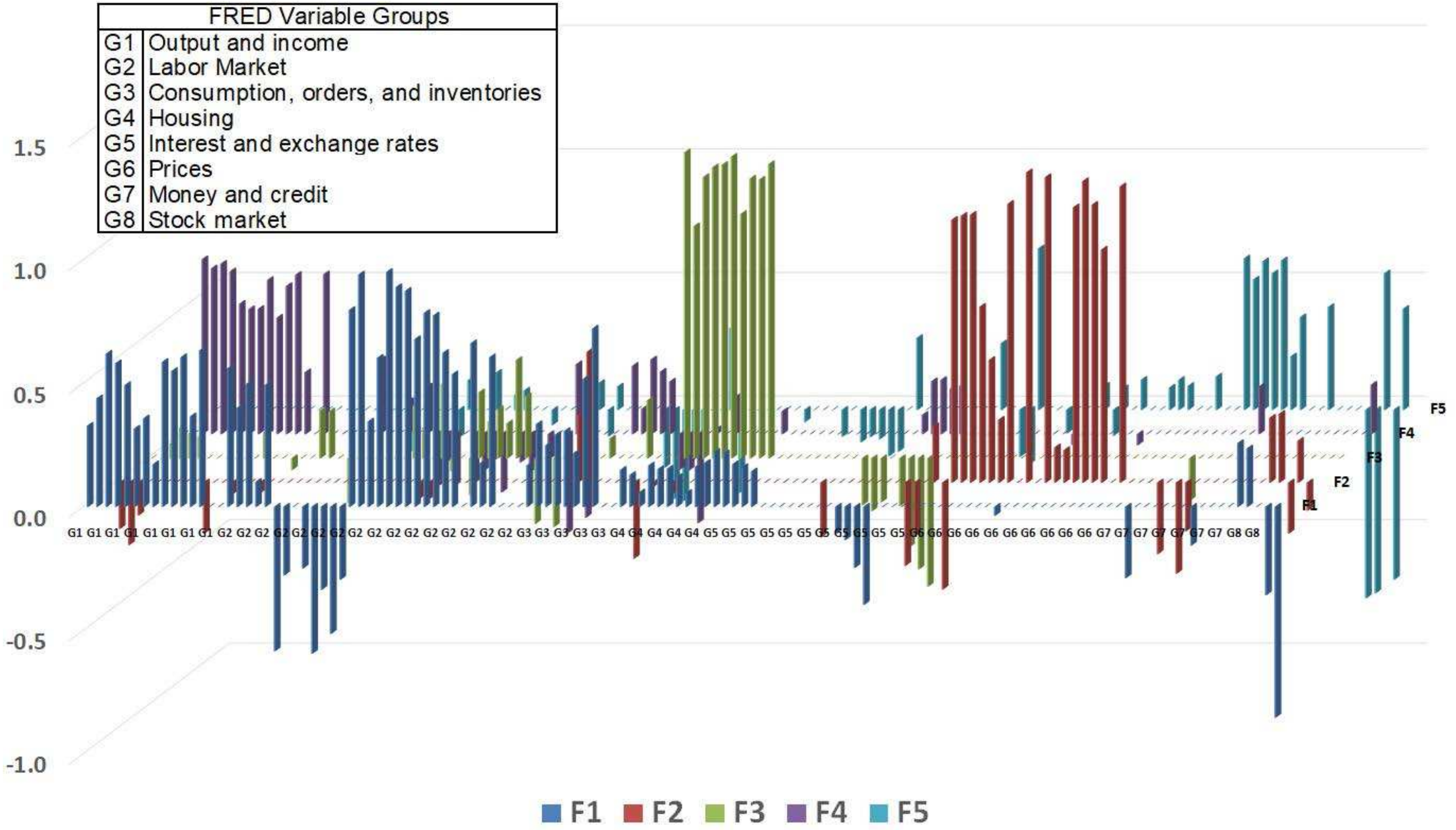

Figure 13: Bar-chart of the resparsified loadings estimates for each of 128 variables with the target FDR level 0.1

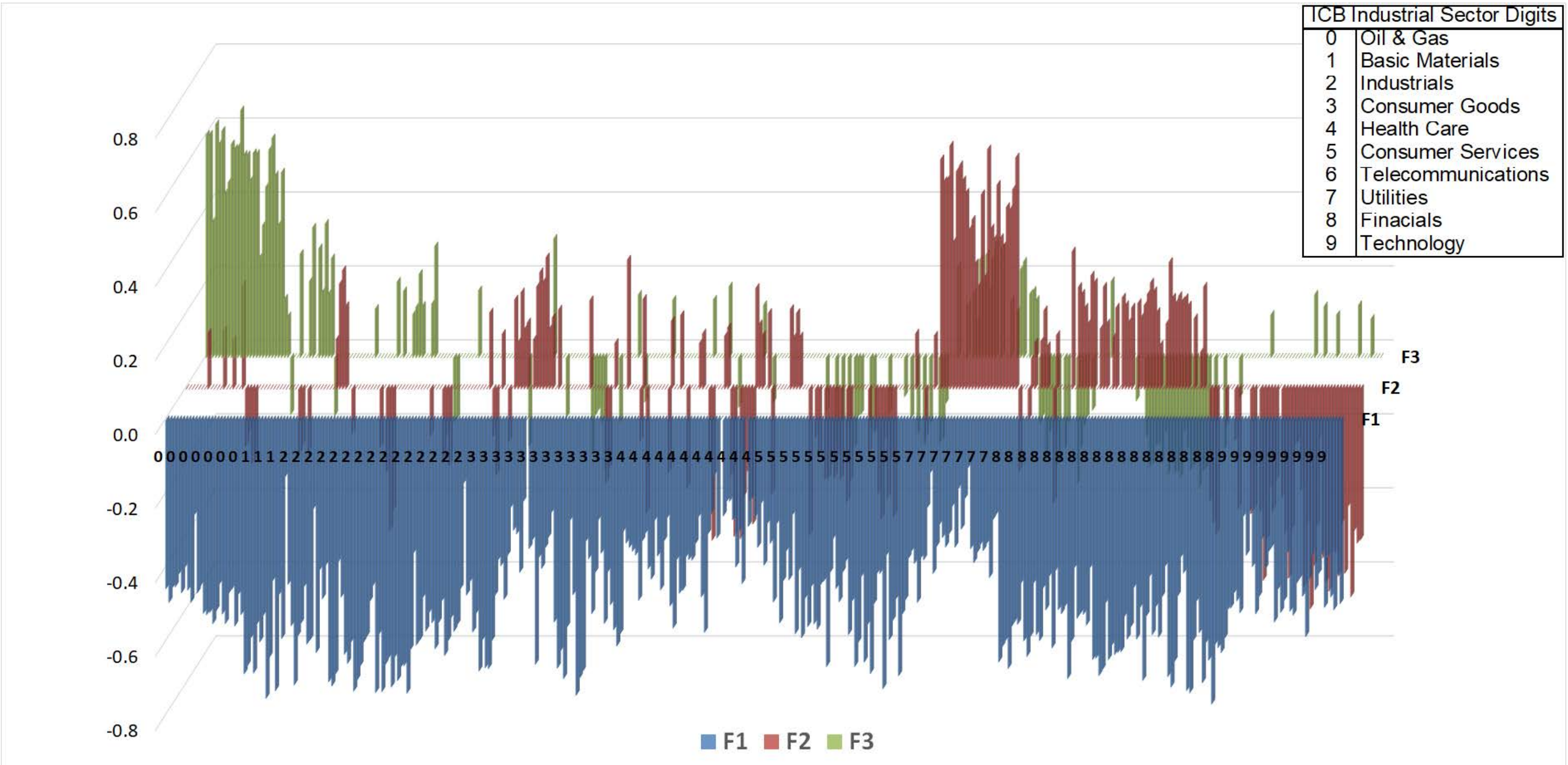

Figure 14: Bar-chart of the resparsified loadings estimates for each of 376 firm security excess returns with the target FDR level 0.1

Supplementary Material for

Inference in Sparsity-Induced 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

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 $\mathcal{F} = \left\{\mathbf{\Delta} \in \mathbb{R}^{T\times r} : \|\mathbf{\Delta}\|_{\mathrm{F}} \le CR_n\right\}$, 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 (2021). Conditional on $\widehat{\mathbf{\Delta}} \in \mathcal{F}$, we can write $\widehat{\mathbf{\Delta}} = R_n\mathbf{U}$ with $\|\mathbf{U}\|_{\mathrm{F}} \le C$ for some matrix $\mathbf{U}$.

By the definition of the debiased SOFAR estimator in Section 3.1, we have the decomposition

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

where $\mathbf{Z} = T^{-1/2}\mathbf{E}'\mathbf{F}^0$, $\mathbf{R}^{(1)}(\widehat{\boldsymbol{\Delta}}) = T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}\widehat{\boldsymbol{\Delta}}$, and $\mathbf{R}^{(2)}(\widehat{\boldsymbol{\Delta}}) = T^{-1/2}\mathbf{E}'\widehat{\boldsymbol{\Delta}}$. Therefore, to obtain the asymptotic linear representation, it is enough to show that $\mathbf{R}^{(1)}(\widehat{\boldsymbol{\Delta}})$ and $\mathbf{R}^{(2)}(\widehat{\boldsymbol{\Delta}})$ are negligible in the max-norm. For any $x > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\|\mathbf{R}^{(1)}(\widehat{\boldsymbol{\Delta}})\|_{\max} > x\right) \le \mathbb{P}\left(\|T^{-1/2}\mathbf{B}^0\mathbf{F}^{0\prime}\widehat{\boldsymbol{\Delta}}\|_{\max} > x \mid \widehat{\boldsymbol{\Delta}} \in \mathcal{F}\right) + \mathbb{P}\left(\widehat{\boldsymbol{\Delta}} \notin \mathcal{F}\right) \\
&\le \mathbb{P}\left(r\|\mathbf{B}^0\|_{\max}\|T^{-1/2}\mathbf{F}^{0\prime}\widehat{\boldsymbol{\Delta}}\|_{\max} > x \mid \widehat{\boldsymbol{\Delta}} \in \mathcal{F}\right) + O((N \vee T)^{-\nu}) \\
&\le \mathbb{P}\left(R_n\|T^{-1/2}\mathbf{F}^{0\prime}\mathbf{U}\|_{\max} \gtrsim x \mid \widehat{\boldsymbol{\Delta}} \in \mathcal{F}\right) + O((N \vee T)^{-\nu}).
\end{aligned}$$

Setting $x = \delta_1 := T^{-1/2}R_n \log^{1/2}(N \vee T)$ leads to the upper bound to be $O((N \vee T)^{-\nu})$ by Assumption 1 with an application of Lemma 1 in Uematsu and Yamagata (2021). Similarly, the second term is bounded as

$$\begin{aligned}
&\mathbb{P}\left(\|\mathbf{R}^{(2)}(\widehat{\boldsymbol{\Delta}})\|_{\max} > x\right) \le \mathbb{P}\left(\|T^{-1/2}\mathbf{E}'\widehat{\boldsymbol{\Delta}}\|_{\max} > x \mid \widehat{\boldsymbol{\Delta}} \in \mathcal{F}\right) + \mathbb{P}\left(\widehat{\boldsymbol{\Delta}} \notin \mathcal{F}\right) \\
&\le \mathbb{P}\left(R_n\|T^{-1/2}\mathbf{E}'\mathbf{U}\|_{\max} > x \mid \widehat{\boldsymbol{\Delta}} \in \mathcal{F}\right) + O((N \vee T)^{-\nu})
\end{aligned}$$

Setting $x = \delta_1$ gives the upper bound to be $O((N \vee T)^{-\nu})$ by Assumption 3 in the same manner. Thus the desired upper bound is obtained in view of the triangle inequality. 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, which is replaced by R_n^{PC} for the PC estimator. Let $\widehat{\boldsymbol{\Delta}}^{PC} = \widehat{\mathbf{F}}^{PC} - \mathbf{F}^0$ and define $\mathcal{F}_{PC} = \left\{\boldsymbol{\Delta} \in \mathbb{R}^{T\times r} : \|\boldsymbol{\Delta}\|_{\mathrm{F}} \le CR_n^{PC}\right\}$, where C is some positive constant and

$$R_n^{PC} = R_n(1 + \gamma_n), \quad \gamma_n = \frac{N^{1/2}(N \wedge T)^{1/2}}{N_1^{1/2}T^{1/2}}.$$

Then under the assumed conditions, $\widehat{\boldsymbol{\Delta}}^{PC} \in \mathcal{F}_{PC}$ holds with probability at least $1 - O((N \vee T)^{-\nu})$ by Uematsu and Yamagata (2021). 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 considers the case when $\mathtt{t}_0$ is given by (14). Write $Z_{ik}^* := Z_{ik}/\sigma_i$ and $e_{ti}^* = e_{ti}/\sigma_i$, where $Z_{ik} = T^{-1/2}\sum_{t=1}^T e_{ti} f_{tk}^0$ with $\sigma_i^2 = \mathbb{E}[e_{ti}^2]$.

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

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

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

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

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, $0 = \mathtt{t}_0 < \mathtt{t}_1 < \cdots < \mathtt{t}_b = \bar{\mathtt{t}} = (2\log(Nr) - a_1\log\log(Nr) + a_2)^{1/2}$ for $a_1 > 3$ and $a_2 > 0$, 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 $v_N = (\log(Nr)\log\log(Nr))^{-1/2}$. Note that $b \asymp \bar{\mathtt{t}}/v_N \asymp \log(Nr)(\log\log(Nr))^{1/2}$. 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 $m \in \{1,\ldots,b\}$ by Javanmard and Javadi (2019), the proof completes if the following is verified:

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

For arbitrary fixed $\varepsilon > 0$, the union bound and Chebyshev's inequality yield

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

Expanding the expectation and collecting terms with using Lemma 5, we obtain

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

where $(\mathcal{Z}^*_{ik}, \mathcal{Z}^*_{j\ell}) := (\mathcal{Z}_{ik}/\sigma_i, \mathcal{Z}_{j\ell}/\sigma_j)$ is the standard bivariate normal vector with the covariance (correlation) $\rho_{ijk\ell} := \rho_{ij}1\{k = \ell\}$ and $\tilde{\delta} \asymp \delta + O(T^{-\kappa})$. We evaluate each term and show that the upper bound of (A.6) is $o(1/b)$.

First derive a lower bound of $G(\mathtt{t}_m)$, and bound the second and fourth terms of (A.6). It is well-known that a standard normal random variable $\mathcal{Z}^*$ satisfies the bound

$$(2\pi)^{-1/2}(z^{-1} - z^{-3})\exp(-z^2/2) \le \mathbb{P}\left(\mathcal{Z}^* > z\right) \le (2\pi)^{-1/2}z^{-1}\exp(-z^2/2) \tag{A.7}$$

for all $z > 0$; see e.g., Feller (1968). Thus we obtain $G(\mathtt{t}_m) \gtrsim N^{-1}\log^{-(1-a_1)/2} N$ uniformly in $m \in \{1,\ldots,b\}$ for given $a_1 > 3$. Thus the second and fourth terms of (A.6) are found to be $o(1/b)$ uniformly in $m \in \{1,\ldots,b\}$.

Next consider the third term of (A.6). Since $\delta > 0$ is polynomially decreasing while $\mathtt{t}_m$

is a logarithmic function for every m, an application of (A.7) with some algebra yields

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

Finally we show that the first term of (A.6) is $1+o(1/b)$. As for the case when $i=j$ we have

$$
\begin{aligned}
&\frac{1}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\mathbb{P}\left(|\mathcal{Z}^*_{ik}|\geq \mathtt{t}_m-\tilde{\delta}, |\mathcal{Z}^*_{j\ell}|\geq \mathtt{t}_m-\tilde{\delta}\right)1\{i=j\}\\
&= \frac{1}{NrG(\mathtt{t}_m)}\frac{G(\mathtt{t}_m-\tilde{\delta})}{G(\mathtt{t}_m)} \leq \log^{(1-a_1)/2}(Nr)(1+o(1/b)) = o\left(1/b\right).
\end{aligned}
$$

Next consider the case of $i\neq j$. By Lemma 6 with the inequality $1/(1-\rho^2)^{1/2}\leq 1+|\rho|/(1-\rho^2)^{1/2}$ and the same argument above, the first term of (A.6) is bounded as

$$
\begin{aligned}
&\frac{G(\mathtt{t}_m-\tilde{\delta})^2}{N^2r^2G(\mathtt{t}_m)^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\frac{\mathbb{P}\left(|\mathcal{Z}^*_{ik}|\geq \mathtt{t}_m-\tilde{\delta}, |\mathcal{Z}^*_{j\ell}|\geq \mathtt{t}_m-\tilde{\delta}\right)}{G(\mathtt{t}_m-\tilde{\delta})^2}1\{i\neq j\}\\
&\leq \frac{1+o(1/b)}{N^2r^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\frac{1}{(1-\rho_{ijk\ell}^2)^{1/2}}1\{i\neq j\}\\
&\leq 1+\frac{1}{N^2r^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\frac{|\rho_{ijk\ell}|}{(1-\rho_{ijk\ell}^2)^{1/2}}1\{i\neq j\}+o(1/b),
\end{aligned}
\tag{A.8}
$$

where $\rho_{ijk\ell}=\rho_{ij}1\{k=\ell\}$ is characterized in Assumption as

$$
|\rho_{ij}| \in \begin{cases} \left[0, c/\log^{\xi} N\right] & \text{for } (i,j)\in P_1 \text{ (weak)},\\ \left(c/\log^{\xi} N, \bar{\rho}\right] & \text{for } (i,j)\in P_2 \text{ (strong)},\end{cases}
$$

for given constants $c>0$, $\xi>1$, and $\bar{\rho}\in(1/2,1)$. Note that $P_1\cap P_2=\varnothing$ and $P_1\cup P_2=\{(i,j)\in[N]\times[N]: i\neq j\}$ with $|P_1|=N^2-N-|P_2|$ and $|P_2|=O(N^2\log^{-\xi}N)$. From

(A.8), we obtain

$$\frac{1}{N^2r^2}\sum_{(i,k)\in\mathcal{S}^c}\sum_{(j,\ell)\in\mathcal{S}^c}\frac{|\rho_{ijk\ell}|}{(1-\rho_{ijk\ell}^2)^{1/2}}1\{i\neq j\}$$
$$\leq \frac{|P_1|}{N^2r^2}\max_{(i,j)\in P_1}\frac{|\rho_{ijk\ell}|}{(1-\rho_{ijk\ell}^2)^{1/2}}+\frac{|P_2|}{N^2r^2}\max_{(i,j)\in P_2}\frac{|\rho_{ijk\ell}|}{(1-\rho_{ijk\ell}^2)^{1/2}}$$
$$=O(1)O(\log^{-\xi}N)+O(\log^{-\xi}N)O(1)=o(1/b). \tag{A.9}$$

Combining the obtained results reveals that (A.6) is $o(1/b)$. Therefore, (A.5) holds. This completes the proof. □

A.4 Proof of Theorem 4

Proof. Define

$$\mathtt{t}_*=\Phi^{-1}\left(1-\frac{qs}{2Nr}(1-x_N)\right)\quad\text{with}\quad x_N=\frac{1}{\log N}. \tag{A.10}$$

A direct use of Lemma 7 with condition $s/N=o(1/\log N)$ establishes that

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

Furthermore, Lemma 8 gives

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

Using these results yields

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

This completes the proof. □

A.5 Proof of Theorem 5

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

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

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

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

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

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

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 7, one minus the power is bounded as

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

which converges to zero under the assumed condition. 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) and (ii) simultaneously hold with probability at least $1 - O((N \vee T)^{-\nu})$:*

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

Moreover, if Assumption 4 is satisfied instead of Assumption 3, then for any matrix norm $\|\cdot\|$, the inequalities (iii) and (iv) hold with probability at least $1 - O((N \vee T)^{-\nu})$:

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

Proof. Let $L_n = N \vee T$ throughout the proof. For (*i*), we only give a sketch of the proof since it is essentially the same as that of Lemma 1 in Uematsu and Yamagata (2021). Note that

$$\left\| T^{-1} \sum_{t=1}^{T} \left(\mathbf{f}_t^0 \mathbf{f}_t^{0\prime} - \mathbf{I} \right) \right\|_{\max} = \left\| T^{-1} \sum_{t=1}^{T} \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \mathbf{\Psi}_\ell \left(\boldsymbol{\zeta}_{t-\ell} \boldsymbol{\zeta}_{t-m}^\prime - \mathbf{E} \boldsymbol{\zeta}_{t-\ell} \boldsymbol{\zeta}_{t-m}^\prime \right) \mathbf{\Psi}_m^\prime \right\|_{\max}$$
$$\lesssim \sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} \|\mathbf{\Psi}_\ell\|_2 \|\mathbf{\Psi}_m\|_2 \left\| T^{-1} \sum_{t=1}^{T} \left(\boldsymbol{\zeta}_{t-\ell} \boldsymbol{\zeta}_{t-m}^\prime - \mathbf{E} \boldsymbol{\zeta}_{t-\ell} \boldsymbol{\zeta}_{t-m}^\prime \right) \right\|_{\max}.$$

Decompose the double summation as

$$\sum_{\ell=0}^{\infty} \sum_{m=0}^{\infty} = \sum_{\ell=0}^{L_n-1} \sum_{m=0}^{L_n-1} + \sum_{\ell=L_n}^{\infty} \sum_{m=0}^{L_n-1} + \sum_{\ell=0}^{L_n-1} \sum_{m=L_n}^{\infty} + \sum_{\ell=L_n}^{\infty} \sum_{m=L_n}^{\infty}$$

and evaluate each sum. Under Assumption 1, the first term is bounded as

$$\begin{aligned}
&\sum_{\ell=0}^{L_n-1}\sum_{m=0}^{L_n-1}\|\boldsymbol{\Psi}_\ell\|_2\|\boldsymbol{\Psi}_m\|_2\left\|T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'-\mathbf{E}\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'\right)\right\|_{\max}\\
&\leq \max_{\ell=0,\dots,L_n-1}\max_{m=0,\dots,L_n-1}\left\|T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'-\mathbf{E}\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'\right)\right\|_{\max}\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}\|\boldsymbol{\Psi}_\ell\|_2\|\boldsymbol{\Psi}_m\|_2\\
&\lesssim T^{-1/2}\log^{1/2}(N\vee T)
\end{aligned}$$

with probability at least $1-O((N\vee T)^{-\nu})$, where we have used the union bound, the Bernstein inequality for a sum of i.i.d. sub-exponential random variables (Vershynin, 2018, Theorem 2.8.1), and $\sum_{\ell=0}^{\infty}\|\boldsymbol{\Psi}_\ell\|_2<\infty$. To derive the upper bound for the other terms that holds with high probability, we evaluate its expectation in view of the Markov inequality. By the sub-exponential property and the summability condition of Assumption 1, the remaining terms are bounded by $O(T^{-1/2}\log^{1/2}(N\vee T))$ with probability at least $1-O((N\vee T)^{-\nu})$. This gives the proof of (i).

The proof of (ii) is found in Uematsu and Yamagata (2021). The proof of (iv) is also obtained in the same manner.

Finally Prove (iii). By the same decomposition as above, we obtain

$$\begin{aligned}
&\left\|T^{-1}\sum_{t=1}^{T}\left(\mathbf{f}_t^0\mathbf{f}_t^{0\prime}-\mathbf{I}\right)\left(e_{ti}^2-\mathbb{E}\,e_{ti}^2\right)\right\|_{\max}\\
&\lesssim\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}\|\boldsymbol{\Phi}_\ell\|_2\|\boldsymbol{\Phi}_m\|_2\left\|T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'-\mathbb{E}\,\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}\right)\left(e_{ti}^2-\mathbb{E}\,e_{ti}^2\right)\right\|_{\max}.
\end{aligned}$$

We then decompose the double sum as above and only consider the term of $\sum\sum_{\ell,m=0}^{L_n-1}$ because the remaining terms can be bounded via the Markov inequality with the summability condition. Using Assumptions 1 and 4 along with the argument of Vershynin (2018), we first note that each component of $\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'-\mathbb{E}\,\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}$, and $e_{ti}^2-\mathbb{E}\,e_{ti}^2$ are sub-exponential random variables. Furthermore, by Theorem 2.1 of Vladimirova et al. (2020), 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) together with the union bound, there exist some constants $c_1,c_2>0$

such that

$$\mathbb{P}\left(\max_{\ell,m=0,\dots,L}\left\|T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}'-\mathbb{E}\,\boldsymbol{\zeta}_{t-\ell}\boldsymbol{\zeta}_{t-m}\right)\left(e_{ti}^2-\mathbb{E}\,e_{ti}^2\right)\right\|>u\right)$$
$$\leq L^2r^2\exp\left(-c_1Tu^2\right)+L^2r^2T\exp\left(-c_2T^{1/2}u^{1/2}\right).$$

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

$$1-L^2r^2\exp\left(-c_1\log(N\vee T)\right)-L^2r^2T\exp\left(-c_2T^{1/4}\log^{1/4}(N\vee T)\right)$$
$$=1-O((N\vee T)^{-\nu}).$$

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 (16).

Proof. Result (i) follows from Uematsu and Yamagata (2021). 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 (16). 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* $1 - O((N \vee T)^{-\nu})$*:*

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

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

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

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

Similarly, we obtain

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

Next, we see that

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

Similarly, we have

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

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,

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

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 4 and 5, 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

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

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

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

By Lemma 3, the second parentheses can be bounded as

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

Therefore we eventually have

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

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

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

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 we have decomposition $\mathtt{T}_{ik} = Z_{ik}/\sigma_i + W_{ik}$ such that for any $\mathtt{t} > 0$ the following results hold:

$$\begin{aligned}
&(i) \quad \max_{i,k} \mathbb{P}\left(|W_{ik}| \geq \delta\right) = O((N \vee T)^{-\nu}),\\
&(ii) \quad \mathbb{P}\left(|\mathtt{T}_{ik}| \geq \mathtt{t}\right) \geq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_i} \geq \mathtt{t} + \delta\right) + O((N \vee T)^{-\nu}),\\
&(iii) \quad \mathbb{P}\left(|\mathtt{T}_{ik}| \geq \mathtt{t}, |\mathtt{T}_{j\ell}| \geq \mathtt{t}\right) \leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_i} \geq \mathtt{t} - \delta, \frac{|Z_{j\ell}|}{\sigma_j} \geq \mathtt{t} - \delta\right) + O((N \vee T)^{-\nu}),\\
&(iv) \quad (Z_{ik}, Z_{j\ell}) = (\mathcal{Z}_{ik}, \mathcal{Z}_{j\ell}) + O(T^{-\kappa}) \quad \textit{a.s. for some constant } \kappa > 0,
\end{aligned}$$

where

$$\begin{pmatrix} \mathcal{Z}_{ik} \\ \mathcal{Z}_{j\ell} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \sigma_{ij} 1\{k = \ell\} \\ \sigma_{ij} 1\{k = \ell\} & \sigma_j^2 \end{pmatrix}\right)$$

with $\sigma_i^2 = \mathrm{Var}(e_{ti})$ and $\sigma_{ij} = \mathrm{Cov}(e_{ti}, e_{tj})$.

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

$$\mathtt{T}_{ik} = \frac{T^{1/2} \hat{b}_{ik}^d}{\hat{\sigma}_i} = \frac{Z_{ik} + R_{ik}}{\hat{\sigma}_i} = \frac{Z_{ik}}{\sigma_i} + \frac{R_{ik}}{\sigma_i} + \left(\frac{\sigma_i}{\hat{\sigma}_i} - 1\right)\left(\frac{Z_{ik} + R_{ik}}{\sigma_i}\right) =: \frac{Z_{ik}}{\sigma_i} + W_{ik}.$$

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

$$\mathbb{P}\left(|\mathtt{T}_{ik}| \geq \mathtt{t}\right) \geq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_i} - |W_{ik}| \geq \mathtt{t}\right) \geq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_i} \geq \mathtt{t} + \delta\right) - |\mathcal{S}^c| \max_{i,k} \mathbb{P}\left(|W_{ik}| > \delta\right),$$

and similarly

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

Thus the proofs of (ii) and (iii) complete if (i) is true.

Prove (i). We have

$$\mathbb{P}\left(|W_{ik}| \geq \delta\right) \leq \mathbb{P}\left(\frac{|R_{ik}|}{\sigma_i} + \left|\frac{\sigma_i}{\hat{\sigma}_i} - 1\right| \left(\frac{|Z_{ik}|}{\sigma_i} + \frac{|R_{ik}|}{\sigma_i}\right) \geq \delta\right)$$
$$\leq \mathbb{P}\left(\frac{|Z_{ik}|}{\sigma_i} \geq \frac{\delta - \delta_1}{\delta_1} - \delta_1\right) + \mathbb{P}\left(\left|\frac{\sigma_i}{\hat{\sigma}_i} - 1\right| \geq \delta_1\right) + \mathbb{P}\left(\frac{|R_{ik}|}{\sigma_i} \geq \delta_1\right). \tag{A.12}$$

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

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

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

$$\mathbb{P}\left(\left|\frac{\sigma_i}{\hat{\sigma}_i} - 1\right| > \delta_1\right) = \mathbb{P}\left(\left|\hat{\sigma}_i^2 - \sigma_i^2\right| > \delta_1 \left|\hat{\sigma}_i\right| \left|\hat{\sigma}_i + \sigma_i\right|\right)$$
$$\lesssim \mathbb{P}\left(\left|\hat{\sigma}_i^2 - \sigma_i^2\right| \gtrsim \delta_1\right) = O((N \vee T)^{-\nu}).$$

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

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

where the last equality is due to Lemma 1(iii). Combining the results gives the proof of (i).

Finally prove (iv). Under Assumptions 1 and 4, it is easily seen that $(Z_{ik}, Z_{j\ell}) = T^{-1/2}\sum_{t=1}^{T}(e_{ti}f_{tk}, e_{tj}f_{t\ell})$ is the sum of square integrable martingale difference sequence with respect to the filtration $\mathcal{F}_t = \sigma\{e_{t-s,i}, e_{t-s,j}, f_{t-s,k}, f_{t-s,\ell} : s = 0, 1, 2, \dots\}$. Then Theorem 3.2 of Kifer (2013) establishes the strong approximation that without changing its distribution the sequence $(e_{ti}f_{tk}, e_{tj}f_{t\ell})$ can be redefined on a richer probability space where there exists a bivariate standard normal random vector $(\mathcal{Z}_{ik}, \mathcal{Z}_{j\ell})$ with the covariance(correlation) ρ such that $(Z_{ik}, Z_{j\ell}) = (\mathcal{Z}_{ik}, \mathcal{Z}_{j\ell}) + O(T^{-\kappa})$ a.s. holds for some $\kappa > 0$, provided that the sufficient conditions (3.4) and (3.5) in the paper are true. Therefore, the proof of (iv) completes if these two conditions are verified.

Check condition (3.4) of the theorem. For any (i,k) and (j,ℓ), an application of Lemma 1 yields

$$\begin{aligned}&\left|T^{-1}\sum_{t=1}^{T}\left(\mathbb{E}[e_{ti}e_{tj}f_{tk}f_{t\ell}\mid\mathcal{F}_{t-1}]-\sigma_{ij}1\{k=\ell\}\right)\right|\\&=|\sigma_{ik}|\left|T^{-1}\sum_{t=1}^{T}\sum_{a,b=0}^{\infty}\boldsymbol{\psi}_{k\cdot}^{(a)}\left(\boldsymbol{\zeta}_{t-a}\boldsymbol{\zeta}_{t-b}'-\mathbf{I}_r\right)\boldsymbol{\psi}_{\ell\cdot}^{(b)\prime}\right|\lesssim T^{-\eta}\end{aligned}$$

for any constant $\eta\in(0,1/2)$ with probability at least $1-O((N\vee T)^{-\nu})$. This verifies the first condition.

Next check condition (3.5) of the theorem. By Assumptions 1 and 3 and equivalence of norms for finite-dimensional vectors, we observe that

$$\mathbb{E}\left[\|(e_{ti}f_{tk},e_{tj}f_{t\ell})\|^4\right]\lesssim\mathbb{E}\,e_{ti}^4\,\mathbb{E}\,f_{tk}^4=O(1).$$

Therefore for any constant $\eta\in(0,1/2)$, we have

$$\begin{aligned}&\sum_{t=1}^{\infty}t^{\eta-1}\,\mathbb{E}\left[\|(e_{ti}f_{tk},e_{tj}f_{t\ell})\|^2 1\left\{\|(e_{ti}f_{tk},e_{tj}f_{t\ell})\|^2\geq 1/t^{\eta-1}\right\}\right]\\&\leq\sum_{t=1}^{\infty}t^{2\eta-2}\,\mathbb{E}\left[\|(e_{ti}f_{tk},e_{tj}f_{t\ell})\|^4\right]\lesssim\sum_{t=1}^{\infty}t^{2\eta-2}<\infty,\end{aligned}$$

which gives the second condition. This completes the proof. □

Lemma 6. *For any bivariate standard normal random vector $(\mathcal{Z}_1,\mathcal{Z}_2)$ with correlation $\rho\in(-1,1)$, it holds that*

$$\sup_{z\in[0,\infty)}\frac{\mathbb{P}(\mathcal{Z}_1\geq z,\mathcal{Z}_2\geq z)}{Q(z)^2}\leq\frac{1}{\sqrt{1-\rho^2}}.$$

Proof. Let $(x,y)\mapsto\phi(x,y;\rho)$ denote the density function of the bivariate standard normal random vector with correlation $\rho \in (-1, 1)$. For any $z \geq 0$, we have

$$\mathbb{P}\left(\mathcal{Z}_1 \geq z, \mathcal{Z}_2 \geq z\right) = \int_z^\infty \int_z^\infty \phi(x, y; \rho) dx dy$$
$$= \frac{1}{\sqrt{1-\rho^2}} \int_z^\infty \int_z^\infty \phi(x, y; 0) \exp\left\{-\frac{2-\rho^2}{2(1-\rho^2)}\left(x^2+y^2\right) + \frac{2\rho}{2(1-\rho^2)} xy\right\} dx dy$$
$$\leq \frac{1}{\sqrt{1-\rho^2}} \int_z^\infty \int_z^\infty \phi(x, y; 0) dx dy \max_{0 \leq x, y < \infty} \exp\left\{-\frac{2-\rho^2}{2(1-\rho^2)}\left(x^2+y^2\right) + \frac{2\rho}{2(1-\rho^2)} xy\right\}.$$

Note that $\int_z^\infty \int_z^\infty \phi(x, y; 0) dx dy = Q(z)^2$ and

$$\max_{0 \leq x, y < \infty} \left\{-\frac{2-\rho^2}{2(1-\rho^2)}\left(x^2+y^2\right) + \frac{2\rho}{2(1-\rho^2)} xy\right\} \leq \max_{0 \leq x, y < \infty} \frac{-2+\rho+\rho^2}{2(1-\rho^2)}(x^2+y^2) = 0$$

uniformly in $\rho \in (-1, 1)$. Combining the results gives the desired uniform bound. □

Lemma 7. *If all the conditions of Theorem 4 are satisfied, then for any $(i, k) \in \mathcal{S}$ 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

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

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

$$\mathbb{P}\left(|\mathtt{T}_{ik}| \leq \mathtt{t}_*\right) = \mathbb{P}\left(|T^{1/2} b_{ik}^0/\sigma_i + Z_{ik}^* + R_{ik}^*| \leq \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_*\right)$$
$$\leq \mathbb{P}\left(|T^{1/2} b_{ik}^0/\sigma_i| - |R_{ik}^*| - \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_* \leq |Z_{ik}^*|\right)$$
$$\leq \mathbb{P}\left(|T^{1/2} b_{ik}^0/\sigma_i| - |R_{ik}^*| - \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_* \leq |Z_{ik}^*| \;\middle|\; |T^{1/2} b_{ik}^0/\sigma_i| - |R_i^*| - \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_* > \mathtt{t}_*\right)$$
$$+ \mathbb{P}\left(|T^{1/2} b_{ik}^0/\sigma_i| - |R_{ik}^*| - \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_* \leq \mathtt{t}_*\right)$$
$$\leq \mathbb{P}\left(\mathtt{t}_* \leq |Z_{ik}^*|\right) + \mathbb{P}\left(|T^{1/2} b_{ik}^0/\sigma_i| - |R_{ik}^*| - \frac{\hat{\sigma}_i}{\sigma_i} \mathtt{t}_* \leq \mathtt{t}_*\right). \tag{A.13}$$

Using inequality (A.7) in the proof of Theorem 3, we can approximate the first term of

(A.13) by $G(\mathtt{t}_*)$:

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

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

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

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

$$\begin{aligned}
&\mathbb{P}\left(|T^{1/2}b^0_{ik}/\sigma_i| - |R^*_{ik}| - \frac{\hat{\sigma}_i}{\sigma_i}\mathtt{t}_* \le \mathtt{t}_*\right) \\
&\le \mathbb{P}\left(|T^{1/2}b^0_{ik}/\sigma_i| - |R^*_{ik}| - \frac{\hat{\sigma}_i}{\sigma_i}\mathtt{t}_* \le \mathtt{t}_* \;\middle|\; \frac{\hat{\sigma}_i}{\sigma_i} \le c_1\right) + \mathbb{P}\left(\frac{\hat{\sigma}_i}{\sigma_i} > c_1\right) \\
&\le \mathbb{P}\left(|T^{1/2}b^0_{ik}/\sigma_i| - (1+c_1)\mathtt{t}_* \le |R^*_{ik}|\right) + \mathbb{P}\left(|\hat{\sigma}_i^2 - \sigma_i^2| > \sigma_i^2(c_1^2-1)\right). \qquad (A.15)
\end{aligned}$$

Note that $\mathtt{t}_* < \sqrt{2\log(Nr)}$ by the construction. Hence by Assumption 6, we obtain

$$|T^{1/2}b^0_{ik}/\sigma_i| - (1+c_1)\mathtt{t}_* > \min_{(i,k)\in\mathcal{S}} |T^{1/2}b^0_{ik}/\sigma_i| - (1+c_1)\sqrt{2\log(Nr)} \gtrsim \sqrt{2\log(Nr)}.$$

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

$$\max_{(i,k)\in\mathcal{S}} \mathbb{P}\left(|T^{1/2}b^0_{ik}/\sigma_i| - |R^*_{ik}| - \frac{\hat{\sigma}_i}{\sigma_i}\mathtt{t}_* \le \mathtt{t}_*\right) = O((N\vee T)^{-\nu}). \tag{A.16}$$

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 8. *If all the conditions of Theorem 4 are satisfied, then* $\mathtt{t}_0 \leq \mathtt{t}_*$ *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

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

The Markov inequality gives

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

where the last equality holds by Lemma 7. Hence we have $R(\mathtt{t}_*) \geq s(1-x_N)$ with high probability. This lower bound with (A.17) entails that

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

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. □

C Additional Experimental Results

C.1 Empirical size of the t-test

We investigate distribution of the standardized debiased SOFAR and PC estimators for a loading, $b^0_{i,k}$, $\mathtt{T}_{ik}$ (i.e., a t-statistic). In particular, we focus on the tail of the distribution, by investigating empirical size of the t-test using $\mathtt{T}_{ik}$, referring to a standard normal distribution. Suppressing the subscript (i, k) for ease of notation, we consider two t-statistics, $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$; see the last paragraph before Section 5.1 for their definitions. $\mathtt{T}_{NW}$ is robust against error serial correlations but $\mathtt{T}_{iid}$ is not.

We consider two experimental designs. For the first design the factors and errors are serially independent ($\rho_{fk} = 0$ and $\rho_e = 0$ for all k), but for the second design they are serially correlated ($\rho_{fk} = 1/4$ and $\rho_e = 1/4$ for all k). We conduct a two-sided test at the five per cent significance level, by rejecting the null $H_0 : b^0_{ik} = 0$ when the absolute value of the t-statistic is greater than 1.96. We fix the combination (i, k) so that $b^0_{ik} = 0$. We investigate all the combinations of $N = 100$ and $T = 100, 200, 500, 1000$. The results are based on 2000 replications.

Table C.3 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 summarizes the size of the tests for serially correlated errors. Let us look at Panel A. With i.i.d. errors, the size behavior of $\mathtt{T}_{iid}$ is expected to be more efficient than that of $\mathtt{T}_{NW}$. Even when $T = 100$, the t-test based on the debiased SOFAR estimator has satisfactory size across the models, exhibiting only minor size distortions. In contrast, the size of the t-test based on the PC estimator is distorted, and the size distortion becomes severer as the model becomes weaker. Furthermore, the size distortion does not seem to disappear when T rises for the models with weaker factors.

Now let us turn our attention to the serial correlation robust test, $\mathtt{T}_{NW}$. As can be seen in Panel A, for $T = 100$, the size of $\mathtt{T}_{NW}$ based on the debiased SOFAR estimate exhibits a moderate size distortion. In contrast, the observed size distortion pattern of $\mathtt{T}_{iid}$ based on the PC estimate is exaggerated for the serial correlation robust test. Increasing T to 200 does not seem to make the distortion sufficiently reduced.

Let us turn our attention to Panel B. The first thing to note is that the size distortion of

$\mathtt{T}_{iid}$ does not ease as T rises, which would be expected since the variance estimator used for $\mathtt{T}_{iid}$ is not consistent. An interesting feature to note is that the size distortion of $\mathtt{T}_{iid}$ based on the PC estimate is much larger than that based on the debiased SOFAR estimate, and the distortion of the PC tests becomes sizable as the model becomes weaker. For example, the averages of the sizes over $T = 100, ..., 1000$ of $\mathtt{T}_{iid}$ based on the debiased SOFAR estimate 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 those of $\mathtt{T}_{iid}$ based on the PC estimate are 7.9%, 8.4% and 10.6%, respectively. For sufficiently large $T(\geq 200)$, the size of the $\mathtt{T}_{NW}$ test based on the debiased SOFAR estimate is satisfactory close to the nominal level, whilst that based on the PC estimate exhibits serious distortion. For example, the averages of the sizes over $T = 200, 500, 1000$ of $\mathtt{T}_{NW}$ based on the debiased SOFAR estimate 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 those of $\mathtt{T}_{NW}$ based on the PC estimate are 6.7%, 7.2% and 9.2%, respectively.

To conclude, the t-statistic based on the debiased SOFAR estimate is preferred to that of the PC estimate, unless the model is almost strong. When the idiosyncratic errors might be serially correlated, we recommend using the robust t-statistic, $\mathtt{T}_{NW}$ based on the debiased SOFAR estimator. The $\mathtt{T}_{NW}$ test based on the PC estimate can suffer from serious size distortions.

C.2 The global inference for the loadings

In Section 5.2, the FDR and the power, based on $\mathtt{T}_{iid}$ for the DGP with serially correlated factors and i.i.d. errors ($\rho_{fk} = 1/4$ and $\rho_e = 0$) are examined. In this section we report two sets of additional experimental results: (i) the same DGP considered in Section 5.2 but based on $\mathtt{T}_0$ (t-ratio based on the true variance of $\hat{b}_{ik}$) and $\mathtt{T}_{NW}$ (serial correlation robust t-ratio), and; (ii) the DGP where both factors and errors are serially correlated ($\rho_{fk} = 1/4$ and $\rho_e = 1/4$).

Table C.1 reports the FDR and the power of the proposed procedure based on $\mathtt{T}_0$ and $\mathtt{T}_{NW}$, along with $\mathtt{T}_{iid}$ for the DGP with serially correlated factors and i.i.d. errors. Interestingly, the FDR based on the debiased SOFAR estimate with $\mathtt{T}_0$ does not vary for different values of T, which implies that the normal approximation for each t-ratio is very accurate with $T = 100$ for all the models considered, which is not the case for the FDR based on the

PC estimate. As one might expect, the $\mathtt{T}_{NW}$ version controls the FDR less accurately than the $\mathtt{T}_{iid}$ version. The FDR based on the PC estimate tends to overshoot the level $q = 0.1$ when the model becomes weaker, whilst the FDR based on the debiased SOFAR estimate satisfactorily controls the FDR for sufficiently large T. The power properties are very similar for the $\mathtt{T}_0$, $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$ versions.

The FDR and the power of the proposed procedure for the DGP with serially correlated factors and idiosyncratic errors are summarized in Table C.2. As one might expect, the $\mathtt{T}_{NW}$ version controls the FDR more accurately than the $\mathtt{T}_{iid}$ version. The FDR based on the PC estimate grossly overshoots the level $q = 0.1$ when the model is very weak, whilst the FDR based on the debiased SOFAR estimate satisfactorily controls the FDR for sufficiently large T. The power properties are very similar for the $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$ versions.

C.3 Re-sparsified SOFAR and sparsified PC estimators with serially correlated errors

Adding to the discussion in Section 5.3, here we report the norm loss of the re-sparsified SOFAR and sparsified PC estimators, with the same DGP employed in Section 5.3 except that the idiosyncratic errors are serially correlated. We denote the re-sparsified SOFAR and sparsified PC estimators based on the serial correlation robust t-ratio $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$ and $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$, respectively. The results are summarised in Table C.4. It clearly shows that $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$ (resp. $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$) outperforms $\widehat{\mathbf{B}}^{\mathsf{r}}$ (resp. $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$), and $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$ dominates $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$ in terms of the norm loss, as expected.

C.4 Robustness of the t-statistic against non-normal errors and time-series heteroskedasticity

On top of the discussion in Section 5.1, here we investigate the finite sample distribution of the t-statistic, $\mathtt{T}_{iid}$, for two different serially uncorrelated error processes ($\rho_e = 0$): (i) standardise chi-squared random variable with six degrees of freedom, ε_{ti} ~i.i.d.$[\chi^2(6) - 6]/\sqrt{12}$; (ii) GARCH(1,1) errors, $\varepsilon_{ti} = \sqrt{h_{ti}}\xi_{ti}$, $h_{ti} = \delta_0 + \delta_1\varepsilon^2_{t-1,i} + \delta_2 h_{t-1,i}$, ξ_{ti} ~i.i.d.$N(0,1)$ with $\delta_0 = 1/2$, $\delta_1 = 1/4$, and $\delta_2 = 1/4$ for $t = -50, ..., T$, setting $\xi_{-49,i} = h_{-49,i} = 0$ and discarding the first 50 time-series observations.

Figures C.1–C.6 report the Q-Q plots of $\mathtt{T}_{iid}$ against $N(0,1)$ for the model $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$ with $N = T = 100$. 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. The first, second, and the third rows show the results for standard normal, chi-squared with six degrees of freedom, and GARCH(1,1) errors, respectively. As can be seen, the quality of the normal approximation of $\mathtt{T}_{iid}$ based on the debiased SOFAR estimator (or the PC estimator) is very similar with the three different types of errors. Also, the distribution of $\mathtt{T}_{iid}$ based on the debiased SOFAR estimator is closer to the standard normal distribution than that based on the PC estimator. When the sample size is increased to $N = T = 200$, the results with which are reported in Figures C.7–C.12, the quality of the approximation improves, but the comparative performance properties are very similar to those for $N = T = 100$. This evidence supports the claim that $\mathtt{T}_{iid}$ is robust against non-normal errors and time-series heteroskedasticity.

Table C.1: FDR and power with $q = 0.1$ using $\mathtt{T}_0$, $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$, for serially correlated factors and i.i.d. errors

$\{\alpha_1, \alpha_2\}$	$\{0.9, 0.8\}$						$\{0.7, 0.6\}$						$\{0.5, 0.4\}$					
	FDR			POWER			FDR			POWER			FDR			POWER		
T, N	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500
Debiased SOFAR																		
$\mathtt{T}_0$																		
100	3.1	5.2	5.5	67.6	87.9	87.6	7.4	7.6	8.4	84.1	86.6	87.0	8.8	9.4	9.1	78.5	78.8	83.1
200	3.1	5.2	5.6	70.9	93.0	92.8	7.4	7.8	8.1	89.1	90.7	91.5	8.5	8.9	9.1	83.2	83.6	87.0
500	3.4	5.0	5.5	74.4	98.6	98.6	7.5	7.8	8.3	97.3	97.5	97.7	8.3	8.9	9.2	94.8	94.2	95.5
$\mathtt{T}_{iid}$																		
100	4.0	6.6	7.2	68.4	88.3	88.0	11.3	11.6	13.2	84.7	87.2	87.5	14.8	16.6	17.9	79.3	79.6	83.8
200	3.6	5.9	6.5	71.3	93.2	93.0	9.5	10.0	10.5	89.4	91.0	91.7	12.0	12.5	13.1	83.7	84.2	87.4
500	3.7	5.4	5.8	74.6	98.7	98.6	8.5	8.6	9.3	97.4	97.6	97.8	9.9	10.6	10.9	94.8	94.3	95.6
$\mathtt{T}_{NW}$																		
100	4.5	7.8	8.5	68.8	88.7	88.5	13.8	14.6	16.8	85.2	87.6	88.0	19.6	22.4	25.1	79.8	80.1	84.2
200	3.9	6.4	7.1	71.5	93.3	93.1	10.6	11.3	12.3	89.6	91.2	91.9	14.3	15.0	16.4	84.0	84.5	87.7
500	3.8	5.7	6.1	74.6	98.7	98.6	8.9	9.4	10.1	97.4	97.6	97.8	11.0	11.7	12.4	94.8	94.3	95.6
PC																		
$\mathtt{T}_0$																		
100	5.5	10.9	8.6	70.6	88.0	87.6	15.5	11.7	11.2	84.5	86.7	87.1	22.1	29.6	15.4	79.2	79.2	83.3
200	4.3	8.2	7.4	72.8	93.1	92.8	14.0	10.5	9.7	89.5	90.8	91.5	20.3	30.7	14.1	84.8	84.5	87.3
500	3.7	6.8	6.3	76.3	98.7	98.6	11.7	9.7	9.4	97.5	97.6	97.8	17.7	28.0	12.9	95.7	95.0	95.9
$\mathtt{T}_{iid}$																		
100	6.1	12.0	10.0	71.3	88.5	88.1	19.1	15.8	15.9	85.0	87.3	87.6	29.3	36.5	25.0	80.2	80.1	84.0
200	4.7	8.8	8.0	73.2	93.2	93.0	16.1	12.7	12.1	89.8	91.2	91.7	25.5	34.4	18.9	85.5	85.0	87.7
500	3.9	7.1	6.5	76.5	98.7	98.6	13.0	10.8	10.4	97.6	97.6	97.8	21.1	30.3	15.3	95.8	95.0	95.9
$\mathtt{T}_{NW}$																		
100	6.5	12.9	11.1	71.6	88.8	88.5	21.3	18.6	19.4	85.4	87.7	88.0	33.5	41.3	31.9	80.7	80.6	84.4
200	5.0	9.2	8.5	73.4	93.4	93.1	17.2	14.1	13.8	90.0	91.4	91.9	27.5	36.2	22.3	85.7	85.4	88.0
500	4.1	7.3	6.8	76.5	98.7	98.6	13.6	11.5	11.2	97.6	97.7	97.9	21.9	31.3	16.9	95.9	95.1	95.9

Table C.2: FDR and power with $q = 0.1$ using $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$, for serially correlated factors and errors

$\{\alpha_1, \alpha_2\}$	$\{0.9, 0.8\}$						$\{0.7, 0.6\}$						$\{0.5, 0.4\}$					
	FDR			POWER			FDR			POWER			FDR			POWER		
T,N	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500	100	200	500
Debiased SOFAR																		
$\mathtt{T}_{iid}$																		
100	4.9	8.5	9.5	69.0	88.6	88.2	15.2	16.4	18.2	85.2	87.4	87.8	22.2	24.2	27.3	79.4	80.0	84.3
200	4.7	7.8	8.6	71.7	93.2	92.9	13.4	14.3	15.6	89.6	90.9	91.9	17.9	19.2	21.7	84.3	84.6	87.7
500	4.7	7.0	7.8	74.9	98.5	98.4	12.0	12.9	13.7	97.2	97.4	97.6	15.6	16.8	17.8	94.6	94.1	95.4
$\mathtt{T}_{NW}$																		
100	4.8	8.3	9.3	68.8	88.3	87.9	15.0	16.4	18.4	85.0	87.2	87.5	23.2	25.1	28.8	79.3	79.7	84.0
200	4.2	7.0	7.8	71.2	92.7	92.5	12.0	12.9	13.9	89.1	90.5	91.4	16.0	17.1	19.2	83.5	83.9	87.2
500	4.1	6.0	6.6	74.4	98.2	98.1	9.8	10.6	11.1	96.7	97.0	97.2	12.3	13.2	13.7	93.7	93.2	94.6
PC																		
$\mathtt{T}_{iid}$																		
100	7.0	13.4	12.4	71.8	88.6	88.2	23.5	20.4	21.1	85.5	87.4	87.8	37.1	43.4	36.5	80.6	80.7	84.5
200	5.9	10.6	10.0	74.0	93.2	92.9	19.3	17.3	17.6	90.0	91.1	92.0	31.5	39.8	28.9	85.9	85.4	88.1
500	5.0	9.0	8.7	76.8	98.5	98.4	17.5	15.4	15.2	97.3	97.5	97.6	27.5	36.5	23.5	95.7	94.9	95.8
$\mathtt{T}_{NW}$																		
100	6.8	13.1	12.1	71.4	88.3	87.9	23.1	20.1	21.1	85.3	87.3	87.5	37.0	43.2	36.8	80.3	80.3	84.2
200	5.4	9.7	9.1	73.4	92.8	92.5	17.8	15.7	15.7	89.5	90.6	91.5	29.0	37.8	25.3	85.0	84.7	87.4
500	4.4	7.8	7.4	76.3	98.3	98.1	15.1	12.8	12.5	96.9	97.1	97.3	23.8	32.6	18.5	94.9	94.1	94.9

Table C.3: Size of the t-test for $H_0 : b_{ik}^0 = 0$

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*	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$
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*	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$	$\mathtt{T}_{iid}$	$\mathtt{T}_{NW}$
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 $x_{ti} = \sum_{k=1}^{r} b_{ik}^0 f_{tk}^0 + \sqrt{\theta} e_{ti}$, $t = 1, .., T, i = 1, .., N$. The factor loadings b_{ik}^0 and factors f_{tk}^0 are formed such that $N^{-1} \sum_{i=1}^{N} b_{ik}^0 b_{i\ell}^0 = 1\{k = \ell\}$ and $T^{-1} \sum_{t=1}^{T} f_{tk}^0 f_{t\ell}^0 = 1\{k = \ell\}$, 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}^0 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 the debiased SOFAR and the PC methods. $\mathtt{T}_{iid}$ and $\mathtt{T}_{NW}$ are the t-statistics for $H_0 : b_{ik}^0 = 0$ for a specific (i, k) with $b_{ik}^0 = 0$, 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.

Table C.4: 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 ($\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$) 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
$T = 100$										
	$\widehat{\mathbf{B}}$	176.4	174.7	181.6	218.8	230.6	244.2	223.0	237.9	261.9
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	162.5	167.3	175.1	272.7	302.7	341.5	432.8	515.8	648.2
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	172.9	178.7	179.7	294.8	309.5	346.9	476.5	588.9	674.5
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	154.4	150.7	149.2	176.3	176.8	176.4	196.1	201.8	212.7
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$	154.3	150.6	149.0	175.4	175.9	175.8	196.3	201.5	213.4
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	167.5	165.3	156.6	205.0	186.5	185.2	255.2	319.4	247.5
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$	167.3	165.0	156.2	203.8	185.1	184.4	252.7	316.8	245.4
$T = 200$										
	$\widehat{\mathbf{B}}$	128.1	128.3	131.4	156.1	165.5	171.5	158.7	167.4	181.1
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	116.4	118.7	122.7	194.6	216.0	244.0	309.6	366.2	460.2
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	120.0	121.7	123.7	203.9	219.8	246.5	337.2	440.6	474.7
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	110.4	107.1	104.9	125.9	126.7	124.2	142.7	143.2	147.9
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$	110.2	106.9	104.6	124.5	125.0	122.1	140.1	140.0	143.5
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	115.3	111.7	107.1	137.6	131.4	127.7	177.7	255.8	164.3
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$	115.1	111.4	106.6	135.9	129.5	125.2	173.4	252.0	157.6
$T = 500$										
	$\widehat{\mathbf{B}}$	79.0	80.3	86.2	96.8	105.1	110.4	100.3	109.8	113.0
	$\widehat{\mathbf{B}}^{\mathsf{d}}$	74.2	76.5	78.9	124.9	137.6	154.8	198.8	233.5	292.0
	$\widehat{\mathbf{B}}^{\mathsf{PC}}$	75.4	77.4	79.1	130.6	140.1	156.3	215.0	283.8	299.9
	$\widehat{\mathbf{B}}^{\mathsf{r}}$	69.1	66.4	65.7	78.8	78.8	76.7	92.3	90.6	90.3
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$	68.8	66.1	65.3	77.6	77.2	74.7	90.2	87.9	86.7
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC}}$	70.9	68.0	66.4	86.1	81.6	78.6	112.0	165.9	98.3
	$\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$	70.6	67.6	65.9	84.5	79.7	76.4	108.6	162.1	93.4

Notes: The DGP is the same as that for Table 1 except $\rho_f = 1/4$ and $\rho_e = 1/4$, i.e., factors and idiosyncratic errors are serially correlated. For the re-sparsified estimator, the target FDR level is set $q = 0.1$. $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{NW}}$ and $\widehat{\mathbf{B}}^{\mathsf{r}}_{\mathsf{PC,NW}}$ are re-sparsified and sparsified estimators based on the robust t-ratio, $\mathtt{T}_{NW}$.

Figures C.1–C.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 $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$ with standardized normal, $\chi^2(6)$, and GARCH(1,1) errors, $N = T = 100$.

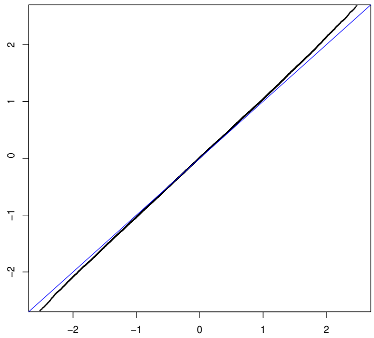

Figure C.1: debiased SOFAR, $N(0,1)$ error

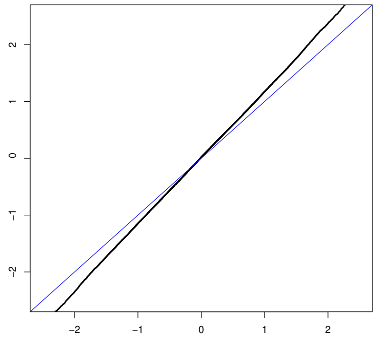

Figure C.2: PC, $N(0,1)$ error

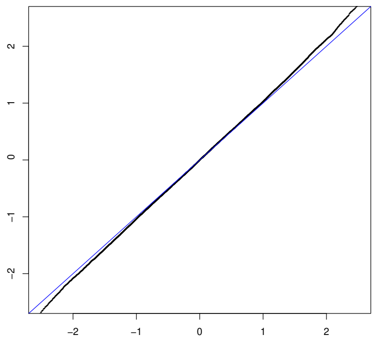

Figure C.3: debiased SOFAR, $\chi^2(6)$ error

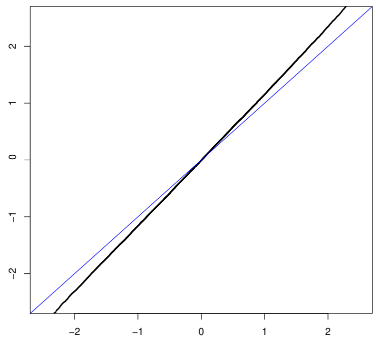

Figure C.4: PC, $\chi^2(6)$ error

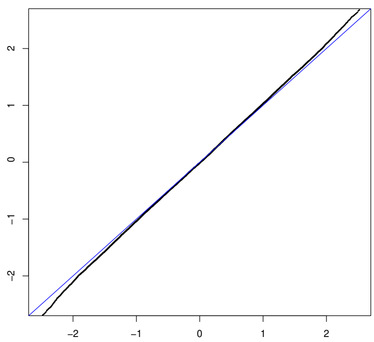

Figure C.5: debiased SOFAR, GARCH(1,1) error

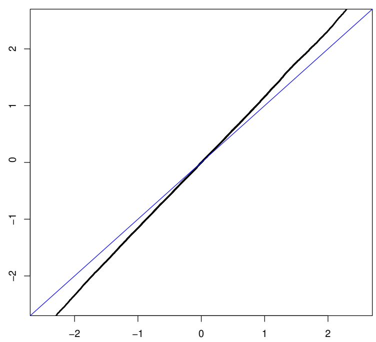

Figure C.6: PC, GARCH(1,1) error

Figures C.7–C.12 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 $\{\alpha_1, \alpha_2\} = \{0.5, 0.4\}$ with standardized normal, $\chi^2(6)$, and GARCH(1,1) errors, $N = T = 200$.

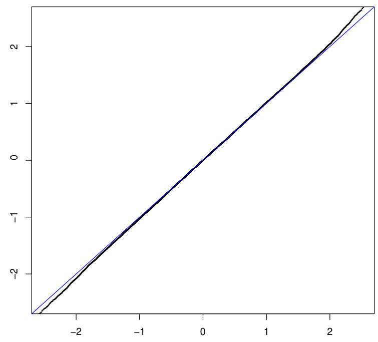

Figure C.7: debiased SOFAR, $N(0,1)$ error

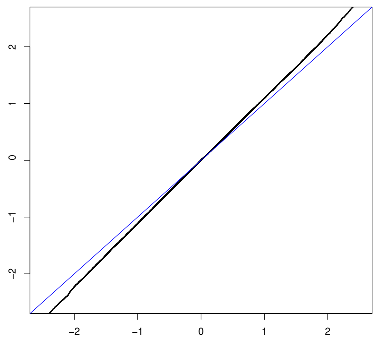

Figure C.8: PC, $N(0,1)$ error

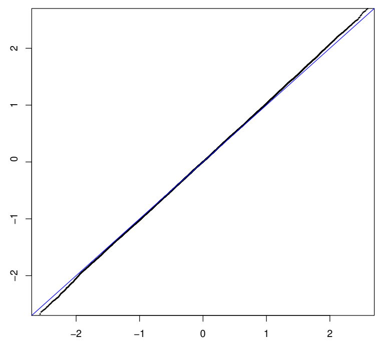

Figure C.9: debiased SOFAR, $\chi^2(6)$ error

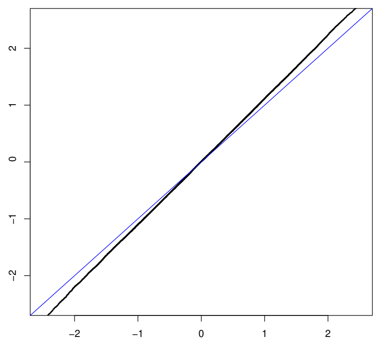

Figure C.10: PC, $\chi^2(6)$ error

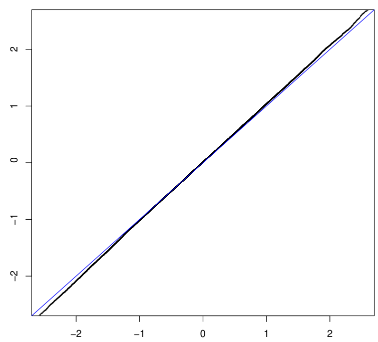

Figure C.11: debiased SOFAR, GARCH(1,1) error

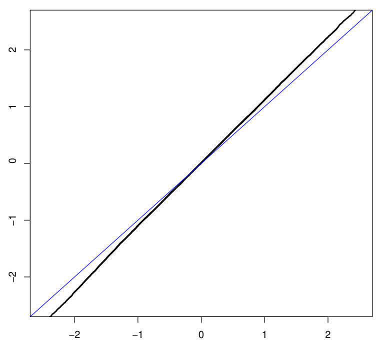

Figure C.12: PC, GARCH(1,1) error

D Additional Empirical Results

In this section we provide the PC estimation versions of the results discussed in Section 6. In a nutshell, the global inference based on the SOFAR method provides clearer information about the characteristics of the estimated factors than that based on the PC method.

D.1 Macroeconomic and financial variables

Figure D.1 below reports the sparsified PC loadings estimates for the FRED-MD data discussed in Section 6.1, which corresponds to the re-sparsified SOFAR estimates in Figure 13 in Section 6.1. The employed specifications for the global inference based on the PC estimates are very similar to those based on the SOFAR estimates: $q = 0.1$ and the t-ratio for $b^0_{i,k}$ is the serial-correlation robust version. The procedure has chosen the value of the FDR controlling threshold, $\mathtt{t}_0 = 1.89$. As a whole, the PC result is similar to the SOFAR result, but the sparsified PC loading is noisier and also has a couple of non-zero clusters, which are not observed in the re-sparsified SOFAR estimates. For example, in Figure D.1 loadings on the second factor are significant for industrial groups 1-5, whilst in Figure 13 corresponding loadings are not significant. Similar comments apply to the loadings for G6 on factor 1, the loadings for G1 on factor 3, the loadings for G7 on factor 4 and the loadings for G1 on factor 5. These differences might be attributed to the higher efficacy of the SOFAR estimates, thus, the global inference based on the debiased SOFAR estimates may be preferred for this application.

D.2 S&P500 firm security excess returns

Figure D.2 below reports the sparsified PC loadings estimates for the S&P500 firm security excess return data discussed in Section 6.2, which corresponds to the re-sparsified SOFAR estimates in Figure 14 in Section 6.2. The employed specifications for the global inference based on the PC estimates are very similar to those based on the SOFAR estimates: $q = 0.1$ and the t-ratio for $b^0_{i,k}$ is the i.i.d. version. The procedure has chosen the value of the FDR controlling threshold, $\mathtt{t}_0 = 1.82$. Comparing these two figures, the PC result is very similar to the SOFAR result, but the sparsified PC loading is slight noisier. Therefore, the global inference based on the debiased SOFAR estimates may be marginally preferred for this application.

D.3 Residuals of Fama-French five-factor regressions

In this subsection we applied the PC based global inference to the residuals of Fama-French five-factor regressions. The employed specifications for the global inference based on the PC estimates are very similar to those based on the SOFAR estimates: $q = 0.1$ and the t-ratio for $b^0_{i,k}$ is the i.i.d. version. The PC result is very different from the SOFAR result in Section 6.3. Specifically, based on the debiased SOFAR loading estimates, there are only two significant loadings (the exponent is 0.13), both of which belong to G9, whilst the PC version has 45 significant loadings (the exponent is 0.72). The numbers of significant loadings for G0-G9 are 1,2,12,5,2,1,0,11,4,7, respectively. We conclude that the SOFAR results in Section 6.3 seem more reliable than the PC result presented in this section, because it is unlikely that the five factor model is short of a semi-strong factor.

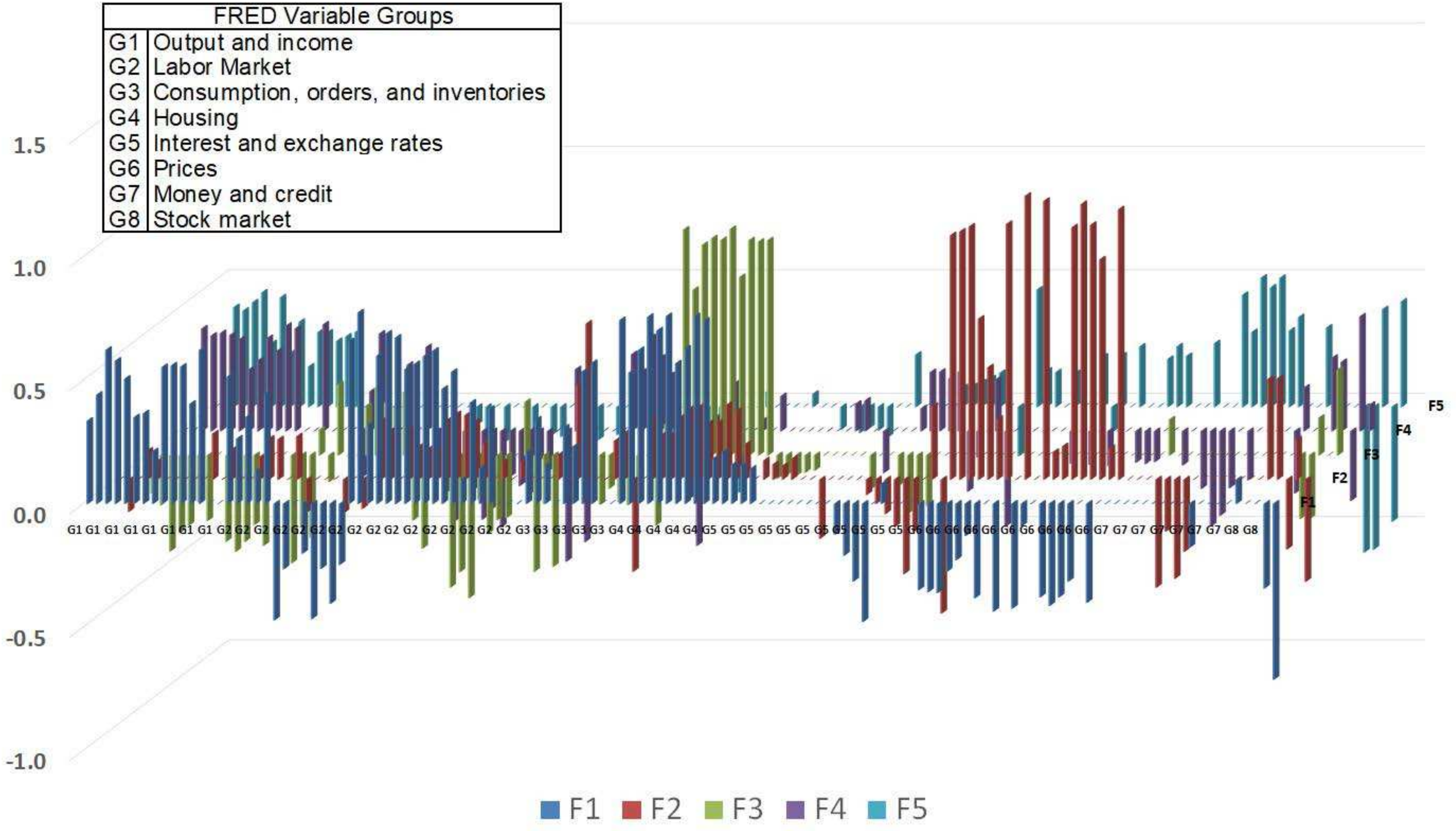

Figure D.1: Bar-chart of the sparsified PC loadings estimates for each of 128 variables with the target FDR level 0.1

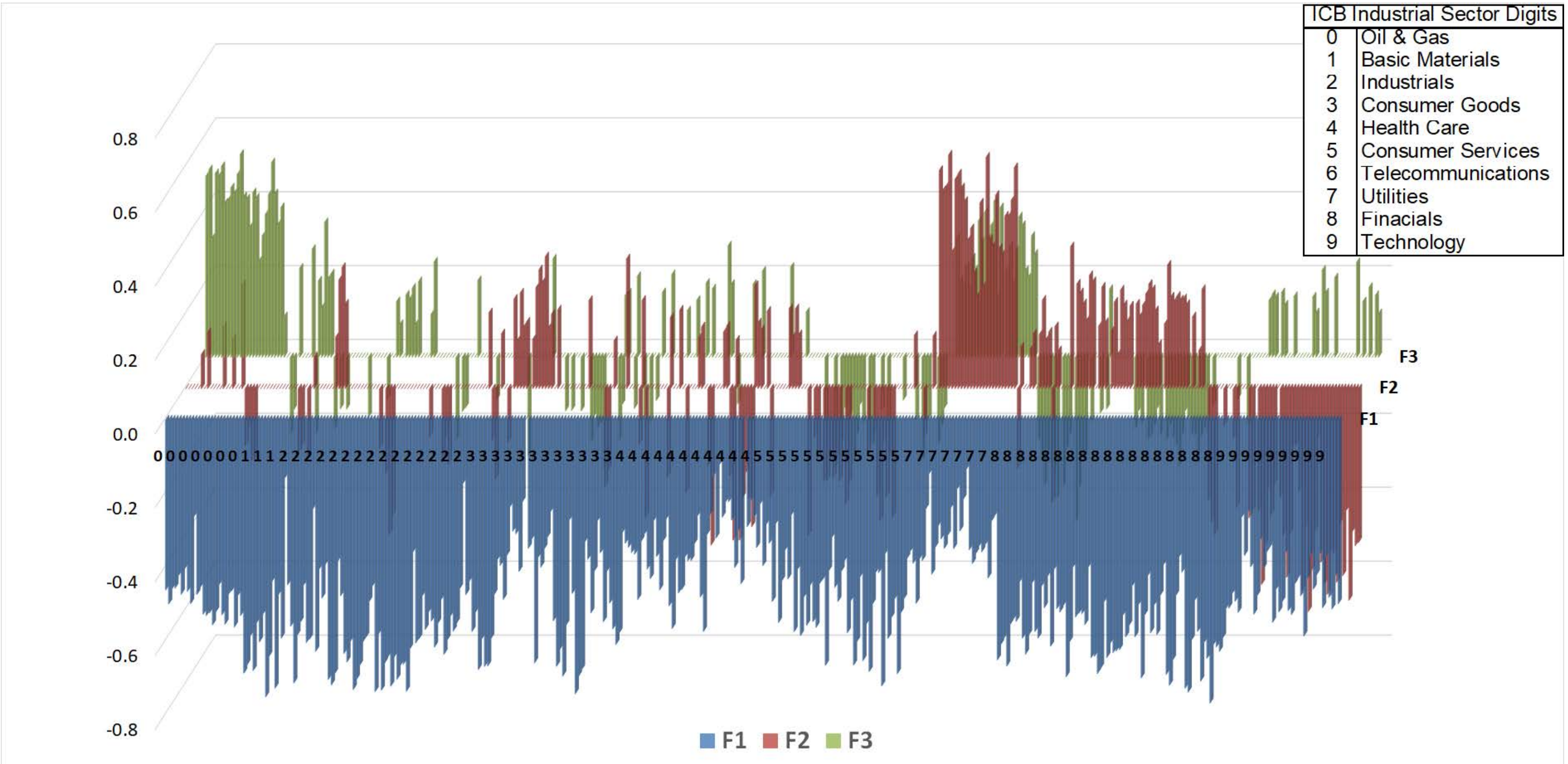

Figure D.2: Bar-chart of the sparsified PC loadings estimates for each of 376 firm security excess returns with the target FDR level 0.1

References

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications Vol. 1 (3rd ed.)*. Wiley.

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

Kifer, Y. (2013). Strong approximations for nonconventional sums and almost sure limit theorems. *Stochastic Processes and their Applications 123*, 2286–2302.

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.

Uematsu, Y. and T. Yamagata (2021). Estimation of sparsity-induced weak factor models. *SSRN: https://ssrn.com/abstract=3374750*.

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

Vladimirova, M., S. Girard, H. Nguyen, and J. Arbel (2020). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat 9*, e318.