*Econometric Theory*, **0**, 2026, 1–24.
doi:10.1017/S0266466625100303

# CAN PRINCIPAL COMPONENT ANALYSIS PRESERVE THE SPARSITY IN FACTOR LOADINGS?

JIE WEI
*Huazhong University of Science and Technology*

YONGHUI ZHANG
*Renmin University of China*

This article studies the principal component analysis (PCA) estimation of weak factor models with sparse loadings. We uncover an intrinsic *near-sparsity preservation* property for the PCA estimators of loadings, which comes from the approximately (block) *upper triangular* structure of the rotation matrix. It suggests an asymmetric relationship among factors: the sparsity of the rotated loadings for a stronger factor can be contaminated by the loadings from weaker ones, but the sparsity of the rotated loadings of a weaker factor is almost unaffected by the loadings of stronger ones. Then, we propose a simple alternative to the existing penalized approaches to sparsify the loading estimators by screening out the small PCA loading estimators directly, and construct consistent estimators for factor strengths. The proposed estimators perform well in finite samples, as shown by a set of Monte Carlo simulations.

## 1. INTRODUCTION

Factor models have been widely used in economics and finance. In a factor model,

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it}, i = 1, \ldots, N, t = 1, \ldots, T, \tag{1.1}$$

where $X_{it}$ is the observed data for the $i$th individual at time $t$, $F_t^0$ is an $r \times 1$ vector of latent factors, $\lambda_i^0$ is the corresponding $r \times 1$ vector of factor loadings, and $e_{it}$ is the idiosyncratic error with possible weak dependence across $i$ or/and over $t$. Until very recently, a large body of literature on factor models has built on the

We are deeply indebted to the co-editor and three anonymous referees for their many constructive comments and suggestions, which improve the article substantially. We are grateful to Simon Freyaldenhoven and Takashi Yamagata for helpful discussions as well as seminar and conference participants at Peking University, Renmin University, 2023 IAAE, and 2023 IPDC. We also thank Simon Freyaldenhoven, Cheng Yong Tang, and Yoshimasa Uematsu for sharing their codes. J.W. gratefully acknowledges the financial support from the Huazhong University of Science and Technology Double First-Class Funds for Humanities and Social Sciences and the Fundamental Research Funds for the Central Universities (HUST: 2022WKYXZX014). Y.Z. gratefully acknowledges the financial support from the National Natural Science Foundation of China (Projects No. 71973141 and No. 71873033). All errors are the authors' sole responsibil ties. MATLAB codes are available upon request. Address correspondence to Yonghui Zhang, School of Economics, Renmin University of China, Beijing, China, e-mail: yonghui.zhang@hotmail.com.

© The Author(s), 2026. Published by Cambridge University Press. 

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

assumption that all factors are strong in the sense that both $T^{-1}\sum_{t=1}^{T} F_t^0 F_t^{0\prime}$ and $N^{-1}\sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime}$ converge to some positive definite (p.d.) matrices. However, the assumption of strong factors may fail due to the sparsity in loadings or many nonzero but small loadings. In such cases, we generally obtain a weak factor model, where $N^{-1}\sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\prime}$ tends to be singular while $T^{-1}\sum_{t=1}^{T} F_t^0 F_t^{0\prime}$ converges to a p.d. matrix. In this article, we focus on weak factor models with sparse loadings, i.e., sparsity-induced weak factor models.

Many empirical studies support the widespread existence of sparse factor loadings. For instance, both Stock and Watson (2002) and Ludvigson and Ng (2009) find that the extracted PC factors from a large set of macroeconomic variables can fit the time-series observations of only a few variables well, suggesting a sparsity structure in loadings (see also Kristensen, 2017; Freyaldenhoven, 2022; Uematsu and Yamagata, 2023a, 2023b, UY hereafter). Moreover, hierarchical or group factor models, where factors have nonzero loadings only for some specific cross-sectional (CS) units or group, also echo the sparsity in loadings. For example, an oil supply shock only affects industrial production sectors but not others; size-sorted portfolios are influenced by the size factor, but momentum-sorted portfolios may not be. Additionally, weak factors may emerge as unsystematic risk due to market incompleteness (Dello-Preite et al., 2024). For more examples of sparsity in loadings, see Ando and Bai (2017) and Choi, Lin, and Shin (2023).

When loadings are sufficiently sparse, what we encounter turns to be a weak factor model. A *sparse weak* factor only affects a small subset of individuals and the model with sparse weak factors is called "*sparse weak factor model*".[1] The sparsity structure in loadings has attracted a lot of research interest. For example, Pelger and Xiong (2022) consider the sparse approximation to factor models, and Freyaldenhoven (2023) and Despois and Doz (2023) study the identification of factor models based on sparsity. In the presence of sparse weak factors, Giglio, Xiu, and Zhang (2023) show that the prediction based on factor-augmented regression (FAR), a popular method for macroeconomic forecasting, is inconsistent (see also Chao and Swanson, 2022; Chao, Liu, and Swanson, 2022 for inconsistent prediction in factor-augmented vector autoregression [FAVAR]). Giglio, Xiu, and Zhang (2025) show that in the 3-step estimation of risk premium in Giglio and Xiu (2021), sparse loadings raise a severe attenuation bias. For extremely weak factor models, Onatski (2012) shows that the PC estimator is inconsistent.[2] For general weak factors, Bai and Ng (2023) re-investigate the asymptotics for principal component analysis (PCA) based on the singular value decomposition. In a linear regression framework with interactive fixed effects, such as in Bai (2009) and Su and Chen (2013), Armstrong, Weidner, and Zeleneev (2025) show that when some factors are weak, previously developed estimators and confidence intervals (CIs) can be heavily biased. They subsequently propose improved estimators and

[1] We formally define the sparse weak factor model in Section 2.

[2] In Onatski (2012), factors are extremely weak in the sense that eigenvalues corresponding to the factors are of the same order of magnitude as those of the idiosyncratic component.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

bias-aware CIs that are uniformly valid, regardless of whether the factors are strong or weak.

To avoid the problems caused by sparse weak factors, two main solutions have been proposed. The first is to recover the sparsity in loadings using penalized regressions. UY (2023a) propose a sparse orthogonal factor regression (SOFAR) estimator using $\ell_1$-regularization, and UY (2023b) further provide an inferential procedure to determine whether each loading component is zero while controlling the false discovery rate (FDR). The second solution is to pursue a strong factor model via screening out irrelevant individuals or variables. Giglio, Xiu, and Zhang (2023, 2025) propose a supervised PCA method for two distinct purposes: selecting test assets to evaluate new factors and choosing predictors to forecast a target macroeconomic variable. By eliminating assets that lack exposure to the factor under evaluation or dropping predictors that are uncorrelated to the target variable, each factor within the model becomes strong. Similar ideas are also employed in Chao and Swanson (2022) and Chao, Liu, and Swanson (2022) in FAVAR.

Regularization-based approaches are often relatively complicated and carry heavy computational burdens due to the search of tuning parameters or the use of complex iterative algorithms. In contrast, we directly investigate the properties of the PCA method and propose recovering the sparsity in loadings via screening the PC estimator. Our article makes several contributions to the existing literature. *First*, we show the PC estimators for loadings are almost *sparsity-preserving*. To the best of our knowledge, ours is the first paper to reveal this intrinsic sparsity preservation feature of the PCA estimator for factor models. This finding contrasts sharply with the prevailing understanding that the sparsity of loadings cannot be preserved due to a matrix rotation (see Bailey, Kapetanios, and Pesaran, 2021, BKP hereafter; Freyaldenhoven, 2023). *Second*, given the near-invariance of sparsity and the consistency of PC estimators, we can easily recover the sparsity of loadings with proper screening to the PC estimator for loadings. The revelation of sparsity is important since it can lead to a better interpretation and understanding of, e.g., which factors are relatively important to which individuals. Another appealing advantage of our method is its negligible computational burden compared to regularization-based procedures. We only need to run standard PCA once and then apply our screening device directly to the estimated loadings. *Third*, based on the sparsified loading estimators, we can straightforwardly estimate distinct factor strengths and establish consistency of the strength estimators. Our approach allows for different factor strengths across factors, thereby improving BKP (2021), which only identify the strength of the strongest latent factor and require a complex sequential procedures. The estimated factor strengths can serve as valuable indicators for forecasting and policy-making. *Fourth*, for sparse weak factor models with varying sparsity degrees, we establish the large sample properties for the PC estimators of factors, loadings, and common components. Complementary to Bai and Ng (2023), we provide uniform convergence results for the PC estimates, which are useful for further analyses, such as constructing diffusion indices and factor-augmented instrumental variable regressions.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

Finally, we need to mention the works most closely related to ours: UY (2023a) and Jiang, Uematsu, and Yamagata (2024). UY (2023a) define a pseudo-true parameter as a representation of $\Lambda^0$ and $F^0$ that satisfies $T^{-1}F^{0\prime}F^0 = I_r$ and the diagonality of $\Lambda^{0\prime}\Lambda^0$, even though this representation need not correspond to the true data-generating process (DGP). Jiang, Uematsu, and Yamagata (2024) recently prove that such pseudo-true parameter always exists, thereby clarifying the estimand of the plain PC method. Our work provides a complementary perspective: while Jiang, Uematsu, and Yamagata (2024) begin with the PC estimator and ask what pseudo-true parameter it consistently estimates, we begin with a sparse representation of the factor model and analyze whether the PC estimators inherit that sparsity in the loadings. Thus, both studies address the identification of PC estimators, but from fundamentally different angles.

The rest of this article is organized as follows. Section 2 formally introduces the sparse weak factor model, i.e., a weak factor model arising from sparse loadings, and the PC estimation method. In Section 3, we study the asymptotic properties of the PC estimators for factors, loadings, and common components under the weak factor framework, establishing their consistency, asymptotic distributions, and uniform convergence rates. Building on the result from Section 3 that the rotation matrix is approximately (block) *upper triangular*, we propose a screening method to sparsify the estimated loadings and provide consistent estimators for factor strengths in Section 4. Monte Carlo simulations are reported in Section 5, and Section 6 concludes. All technical proofs for the main results in Sections 3 and 4, as well as the determination of the number of factors, additional simulation results, and an empirical application, are relegated to the Supplementary Material.

*Notations*. For a set $S$, let $|S|$ be its cardinality. Let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$ for two real numbers $a$ and $b$. For two random sequences $\{a_n\}$ and $\{b_n\}$, $a_n \lesssim_p b_n$ denotes $a_n/b_n$is stochastically bounded and $a_n \asymp_p b_n$ if $a_n \lesssim_p b_n$ and $b_n \lesssim_p a_n$. When $\{a_n\}$ and $\{b_n\}$ are deterministic, $a_n \lesssim b_n$ denotes $a_n/b_n$is bounded and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For a square matrix $A$, $\mu_{\min}(A)$ and $\mu_{\max}(A)$ are the smallest and largest eigenvalues of $A$, respectively. Let $\|B\| = \left[\operatorname{Tr}\left(BB'\right)\right]^{1/2}$ be the Frobenius norm of matrix $B$ and $\|B\|_{sp} = \mu_{\max}^{1/2}\left(B'B\right)$ be its spectral norm. For an $n \times 1$ vector $a$, its $\ell_0$-norm is $\|a\|_0 = \sum_{j=1}^{n} \mathbf{1}\left(a_j \neq 0\right)$, where $\mathbf{1}(\cdot)$ is the usual indicator function.

## 2. THE SPARSE WEAK FACTOR MODELS

### 2.1. The Definition of Sparse Weak Factor Models

In this article, we consider the following factor model with a sparse representation:

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it} = \sum_{k=1}^{r} \lambda_{ik}^0 F_{tk}^0 + e_{it}, \tag{2.1}$$

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

where $\lambda_i^0 = \left(\lambda_{i1}^0, \ldots, \lambda_{ir}^0\right)'$ and $F_t^0 = \left(F_{t1}^0, \ldots, F_{tr}^0\right)'$.[3] In a matrix form, the model in (2.1) can be written as $X = \Lambda^0 F^{0\prime} + e$, where $X$ and $e$ are both $N \times T$ matrices with $X(i,t) = X_{it}$ and $e(i,t) = e_{it}$, respectively, $\Lambda^0 = \left(\lambda_1^0, \ldots, \lambda_N^0\right)'$ and $F^0 = \left(F_1^0, \ldots, F_T^0\right)'$. We assume that the factors are well-behaved in the sense that $T^{-1}\sum_{t=1}^T F_t^0 F_t^{0\prime} \to_p \Sigma_F$ as $T \to \infty$ for some p.d. and finite matrix $\Sigma_F$. For the $k$th factor, $k = 1, \ldots, r$, the $\ell_0$-norm of the $N \times 1$ factor loadings $\Lambda_{\cdot,k}^0 := \left(\lambda_{1k}^0, \ldots, \lambda_{Nk}^0\right)'$ is of order $N^{\alpha_k}$ for some $\alpha_k \in [0,1]$, that is,

$$\left\|\Lambda_{\cdot,k}^0\right\|_0 = \sum_{i=1}^N \mathbf{1}\left(\lambda_{ik}^0 \neq 0\right) \asymp N^{\alpha_k}. \tag{2.2}$$

According to (2.2), $\alpha_k = 1$ defines a strong factor which affects almost all CS units; $\alpha_k = 0$ leads to an extremely weak factor which at most affects a finite number of units; any $\alpha_k \in (0,1)$ gives a sparse structure on the $N$ loadings. A larger $\alpha_k$ implies a stronger factor which affects more CS units. So the strength of the $k$th factor or the sparsity degree in its loadings can be represented by $\alpha_k$. As BKP (2021), we define the parameter $\alpha_k$ as the *factor strength* of the $k$th factor. Further, we arrange the $r$ factors in (2.1) by their factor strengths such that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_{r-1} \geq \alpha_r$. A decreasing sequence $\{\alpha_k\}_{k=1}^r$ can capture the sparsity structure of the factor models.

For any bounded and invertible rotation matrix $R \in \mathbb{R}^{r\times r}$, we denote observationally equivalent factors and loadings as: $F = F^0 R$ and $\Lambda = \Lambda^0 \left(R'\right)^{-1}$. Consequently, we can define the factor strength vector for $F$, analogously to (2.2), as $\alpha(R) \equiv (\alpha_1(R), \ldots, \alpha_r(R))'$ with $\alpha_j(R) \in [0,1]$ for $j = 1, \ldots, r$, and $\alpha_j(R) \geq \alpha_k(R)$ for $j < k$. Due to the intrinsic identification indeterminacy of model (2.1), to characterize our sparse weak factor model with well-defined factor strengths, we introduce the following definition of the sparsest representation.

Definition 2.1 (The sparsest representations of a factor model). *The rotation matrix $R^* \in \mathbb{R}^{r\times r}$ gives a sparsest representation of factor loadings (up to order), if the factors after rotation $F^* = F^0 R^*$ satisfy that $T^{-1}F^{*\prime}F^*$ is p.d. and finite in probability limit, and each element of factor strength vector $\alpha(R^*)$ cannot be further reduced for any other rotation matrix, that is $\alpha_k(R^*) \leq \alpha_k(R)$ for any invertible and bounded $R \in \mathbb{R}^{r\times r}$ and $k = 1, \ldots, r$.*

**Remark 1.** The sparsest representation indeed exists and is well defined. Definition 2.1 pins down uniquely the factor strengths for the sparsest representations, regardless of common or distinct factor strengths (or mixed of both). We will show

[3] Note that while the model in (2.1) is a sparse representation of the true model or a structural model, this does not imply that the true model or the structural factor model must necessarily take the sparsest form of loadings. In reality, the true model can exhibit either a sparse or even a non-sparse (given the existence of one strong factor) structure in loadings. Given the rotation equivalence property of factor structure, we focus on the sparsest representation of the true factor model and directly make assumptions on it.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

that all factor strengths can be consistently estimated in Section 4.2. Here, we need to mention that the sparsest representation is not unique because many rotations give the same factor strengths vector $\alpha(R^*)$.

As discussed later in Section 3 with the relevant normalization in Assumptions 1 and 2 therein, it implies that the underlying $\left(F^0, \Lambda^0\right)$ has the sparsest representation with $R^* = I_r$ and factor strengths $\alpha(I_r)$, where $I_r \in \mathbb{R}^{r\times r}$ is the identity matrix. In other words, there is no other rotation matrix that can further reduce any element of the factor strength vector $\alpha(I_r)$. Given that two factors may possess the same degree of strength, we distribute $\alpha_j$'s into $G$ groups so that: $\mathcal{I}^{[1]} \equiv \left\{1, \ldots, m_1 : \alpha_1 = \cdots = \alpha_{m_1}\right\}, \ldots, \ \mathcal{I}^{[G]} \equiv \left\{m_{G-1}+1, \ldots, r : \alpha_{m_{G-1}+1} = \cdots = \alpha_r\right\}$ with $G \leq r$ and $\alpha_{m_1} > \alpha_{m_2} > \cdots > \alpha_{m_G}$, and $G = r$ when every factor's strength is unique. For ease of notation, let us also define the cardinality of $\mathcal{I}^{[g]}$ by $r_g \equiv |\mathcal{I}^{[g]}|$, for $g = 1, \ldots, G$. Clearly, $r = \sum_{g=1}^{G} r_g$. In addition, let $\alpha_{[j]}$ be the same strength shared in group $g$, i.e., $\alpha_{[g]} = \alpha_{m_g}$ for $g = 1, \ldots, G$.

As BKP (2021) point out, for a factor with an extremely weak signal with $\alpha_k \in (0, 1/2)$, it cannot be identified without *prior* restrictions and is not relevant in most financial and macroeconomic applications. Moreover, Freyaldenhoven (2022) discusses the reason why only factors affecting proportionally more than $\sqrt{N}$ of individuals are relevant in arbitrage pricing theory and aggregate fluctuations in macroeconomics. Hence, we restrict the factor strength so that $\alpha_k \in [\underline{\alpha}, 1]$ for $k = 1, \ldots, r$ with some $\underline{\alpha} > 1/2$.[4] Admittedly, the restriction on the weakest factor's strength is stronger than that in Bai and Ng (2023) and UY (2023a, 2023b).

## 2.2. Empirical Relevance of the Sparse Representations

What is the relevance of the sparse representation in empirical applications, compared with alternative (observationally equivalent) representations? Here, we provide two important motivational examples in studying such a sparse representation with $\Lambda^0$ and $F^0$.

**Example 2.2** (Empirical asset pricing). The sparse representation would help identify the strength of observed or constructed factors robustly in empirical asset pricing models. For observed factors such as the three factors in Fama and French (1993), the Fama–MacBeth two-pass regression has been developed to estimate the loadings and risk premia. Based on estimated loadings, the factor strengths can be estimated (BKP 2021), and can provide important information for identifying useful factors from the "factor zoo" (Cochrane, 2011). However, such a model is highly likely misspecified due to omitting important factors (see Giglio and Xiu, 2021; Kim, Raponi, and Zaffaroni, 2024). Instead, one could estimate the sparse latent factor model such as (2.1) and obtain the estimators for factors $\{\widehat{F}_t\}_{t=1}^T$ and sparsified loadings estimator $\widehat{\Lambda}$. Then, for a given observed factor $g_t$, using the

[4] Factors with the strength parameter larger than $1/2$ are called "semi-strong" factors in BKP (2021).

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

linear projection of $g_t$ on $\widehat{F}_t$, one could estimate the factor strength of $g_t$ and further evaluate its risk premium.

**Example 2.3** (Detect relevant factors for individuals). The sparse representation can be used to identify redundant factors for each individual. To fix ideas and simplify, let us consider model (2.1) with $r = 2$ and factor loadings $\Lambda^\diamond = \begin{pmatrix} \Lambda_{\cdot,1} & a\Lambda_{\cdot,1} \\ b\Lambda_{\cdot,2} & \Lambda_{\cdot,2} \end{pmatrix}$ with $ab \neq 0$ or 1, where each of the four blocks is an $\frac{N}{2} \times 1$ vector and $N$ is assumed to be even.[5] For a given $i$, it is sufficient to capture the time-series variation of $X_{it}$ by only one factor. The sufficiency is clearly expressed by $\Lambda^\diamond$'s sparse representation via rotation:[6] $\Lambda^0 = \begin{pmatrix} \Lambda_{\cdot,1} & 0 \\ 0 & \Lambda_{\cdot,2} \end{pmatrix}$. Clearly, the sparsity structure in (2.3) implies that there is only one relevant factor for each unit. In addition, identification of (ir)relevant factors for certain individuals may play a fundamental role in solving the weak factor problem. As proposed by Giglio, Xiu, and Zhang (2025), the knowledge of factor relevancy helps screen irrelevant individuals (test assets) for a given $F^0_{\cdot,k}$, and evaluate $F^0_{\cdot,k}$ only using the relevant ones to which $F^0_{\cdot,k}$ is pervasive.

## 2.3. PC Estimation

To begin with, we assume that the true number of factors $r$ is known, and leave the determination of $r$ in Section D of the Supplementary Material. The estimation of factors and their loadings is via the method of PC in minimizing

$$\min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i' F_t\right)^2,$$

subject to the usual identification restriction that $F'F/T = I_r$ and $\Lambda'\Lambda$ being diagonal. Since it is known that the PC estimator $\widetilde{F}$ is identified up to a full rank rotation matrix, the identification restriction $F'F/T = I_r$ is only employed to pin down $\widetilde{F}$ for the purpose of estimation. The estimated factors, denoted by $\widetilde{F}$, are $\sqrt{T}$ times the eigenvectors corresponding to the $r$ largest eigenvalues of the $T \times T$ matrix $\frac{X'X}{NT}$ in decreasing order. Then, $\widetilde{\Lambda} = X\widetilde{F}/T$ and $\widetilde{e} = X - \widetilde{\Lambda}\widetilde{F}'$. Also, let $\widetilde{V} = \text{diag}\left(\widetilde{V}_1, \ldots, \widetilde{V}_r\right)$ be the $r \times r$ diagonal matrix consisting of the $r$ largest eigenvalues of $\frac{X'X}{NT}$ in decreasing order. We also define the common component estimator $\widetilde{C} = \widetilde{\Lambda}\widetilde{F}'$ as the estimator for $C^0 = \Lambda^0 F^{0\prime}$.

[5] The example used here just illustrates the idea since the two factors involved are, strictly speaking, not weak by our previous definition.

[6] It can be easily verified that $\Lambda^0 = \Lambda^\diamond H$ with $H = \begin{pmatrix} 1 & -a \\ -b & 1 \end{pmatrix}$.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

## 3. LARGE SAMPLE PROPERTIES FOR PC ESTIMATORS

### 3.1. Main Assumptions

We first define a scale matrix $A$ as follows, $A \equiv \text{diag}\,(N^{\alpha_1}, \ldots, N^{\alpha_r})$, which will be frequently used in this article, and define the non-null set with regard to $\lambda_{ik}^0$'s for $k = 1, \ldots, r$, as $\mathcal{L}_k^0 \equiv \{i : \lambda_{ik}^0 \neq 0, i = 1, \ldots, N\}$.

**Assumption 1.** $E\|F_t^0\|^4 < \infty$ for $t = 1, \ldots, T$ and $T^{-1}F^{0\prime}F^0 \to_p \Sigma_F$ as $T \to \infty$ for some p.d. matrix $\Sigma_F$.

**Assumption 2.** For the factor loadings,

(i) $0 < \underline{\lambda} \leq \left|\lambda_{ik}^0\right| \leq \overline{\lambda} < \infty$, $\forall i \in \mathcal{L}_k^0$, for $k = 1, \ldots, r$, where $\underline{\lambda}$ and $\overline{\lambda}$ are two constants;

(ii) $A^{-1/2}\Lambda^{0\prime}\Lambda^0 A^{-1/2} \to \Sigma_\Lambda^*$ as $N \to \infty$ for some p.d. matrix $\Sigma_\Lambda^*$, and for the $k$th factor, the number of nonzero factor loading is $n_k \equiv \left\|\Lambda_{\cdot,k}^0\right\|_0 \asymp N^{\alpha_k}$ for $\alpha_k \in [\underline{\alpha}, 1]$ with $\underline{\alpha} > 1/2$ for $k = 1, \ldots, r$.

**Assumption 3.** There exists a positive constant $M \leq \infty$, such that for all $N$ and $T$,

(i) $E(e_{it}) = 0$ and $E(e_{it}^2) < M$ for all $i$ and $t$;

(ii) $E(e_s'e_t/N) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ for all $s$, and $T^{-1}\sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s,t)| \leq M$;

(iii) $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $\left|\tau_{ij,t}\right| \leq \tau_{ij}$ for all $t$ with some $\tau_{ij} > 0$. In addition, for $\forall \mathcal{S} \subset \{1, \ldots, N\}$, $\sum_{j=1}^N \tau_{ij}\mathbf{1}(j \in \mathcal{S}) \leq M$ and $|\mathcal{S}|^{-1/2}T^{-1/2}\sum_{j=1}^N \sum_{t=1}^T \left(e_{it}e_{jt} - \tau_{ij,t}\right)\mathbf{1}(j \in \mathcal{S}) = O_p(1)$;

(iv) for every $(t,s)$, $E\left|N^{-1/2}\sum_{i=1}^N [e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$;

(v) $(n_kT)^{-1/2}\sum_{i=1}^N \sum_{t=1}^T F_t^0\lambda_{ik}^0 e_{it} = O_p(1)$ for $k = 1, \ldots, r$;

(vi) $\max_s \sum_{t=1}^T \left\|\gamma_{N,F}(s,t)\right\|^2 \leq M$ and $\max_s E\left\|\varpi(s)\right\|^2 \leq M$, where $\gamma_{N,F}(s,t) = N^{-1}E(F_t^0e_t'e_s)$ and $\varpi(s) = (NT)^{-1/2}\sum_{i=1}^N \sum_{t=1}^T \left[F_t^0e_{it}e_{is} - E(F_t^0e_{it}e_{is})\right]$;

(vii) $E\left(N^{-1}\sum_{i=1}^N \left\|T^{-1/2}\sum_{t=1}^T F_t^0e_{it}\right\|^2\right) \leq M$;

(viii) the eigenvalues of $\Sigma_\Lambda^*\Sigma_F$ are distinct.

**Assumption 4.** $\|e\|_{sp}^2 = O_p(\max\{N,T\})$.

Assumption 1 imposes a moment condition on the factors and requires the existence of a p.d. probability limit of $T^{-1}F^{0\prime}F^0$. Assumption 2 imposes boundedness conditions on nonzero loadings and specifies the sparsest structure in loadings. The deterministic loadings can be relaxed to be stochastic with some additional moment conditions. It is worth mentioning that Assumptions 1 and 2 together imply that the structure $\left(F^0, \Lambda^0\right)$ satisfies the *sparsest* representation, as the column-wise nonzero loadings in $\Lambda^0$ matches the eigenvalues of $\Sigma_C \equiv \Lambda^0\Sigma_F\Lambda^{0\prime}$ which is rotation invariant. Assumption 3(i)–(iii) impose moment conditions on errors and allow for weak CS/serial dependence as Bai (2003). Note that 3(iii) is weaker than Bai (2003) and thus generalizes the counterpart under a strong factor

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

model. 3(vi) is standard in panel factor models, and it is not redundant since we do not assume $F_t^0$ and $e_s$ are independent. Assumption 4 is also adopted by Bai and Ng (2002, 2023). It surely holds for independently identically distributed (iid) data with uniformly bounded 4th moments, and may also hold for weakly dependent data across $i$ and $t$.

Our article focuses on the sparse loadings, where elements are either zero or bounded away from zero, as specified in Assumption 2(i). Our framework is not as general as that of Bai and Ng (2023), which accommodate two scenarios—sparse loadings and shrinking loadings. The former involves many zero entries with the remainder being non-degenerate, while the latter are dense, with all loadings shrinking to zero. In practice, a mixture of these two cases may certainly exist. Nevertheless, sparse factor loadings are more critical for estimating key parameters (e.g., Giglio, Xiu, and Zhang, 2025) and are also commonly used to interpret latent factors (Ludvigson and Ng, 2009).

### 3.2. Consistency and Limiting Distributions

We first present one of the key interesting results for PC estimation of weak factor models. Recall that $\widetilde{V} = \text{diag}\left(\widetilde{V}_1, \ldots, \widetilde{V}_r\right)$ is the diagonal matrix consisting of the $r$ largest eigenvalues of $\frac{X'X}{NT}$ in decreasing order. We show that the eigenvalue matrix $\widetilde{V}$ preserves the magnitude of factor strength, as stated in Proposition 3.1 below.

PROPOSITION 3.1. *Under Assumptions 1–4,* $\widetilde{V}_k \asymp_p N^{\alpha_k - 1}$ *for* $k = 1, \ldots, r$.

**Remark 2.** Unlike in a strong factor model, the diagonal elements of matrix $\widetilde{V}$ vanish at various rates determined by their corresponding factor strengths. It raises more challenges to our asymptotic theory later for terms involving $\widetilde{V}^{-1}$ frequently, whereas in a strong factor model, all these terms are $O_p(1)$.

Next, we turn to the consistency and asymptotic distributions for PC estimators. We introduce several rotation matrices as below. Define

$$H = \frac{\Lambda^{0\prime}\Lambda^0}{N}\frac{F^{0\prime}\widetilde{F}}{T}\widetilde{V}^{-1},\ H_1 = \Lambda^{0\prime}\Lambda^0(\widetilde{\Lambda}'\Lambda^0)^{-1},\ H_2 = \left(F^{0\prime}F^0\right)^{-1}F^{0\prime}\widetilde{F},$$

$$H_3 = Q^{-1} \text{ with } Q = \frac{\widetilde{F}'F^0}{T}, \text{ and } H_4 = \Lambda^{0\prime}\widetilde{\Lambda}(\widetilde{\Lambda}'\widetilde{\Lambda})^{-1},$$

where we have suppressed the dependence on sample sizes for these matrices to ease the burden of notations. The introduction of different rotation matrices is mainly for the convenience in the proofs of different results. We will show the asymptotic equivalence of the matrices $H$ and $H_j$ for $j = 1, 2, 3, 4$, which also generalizes Lemma 3 of Bai and Ng (2023). These rotation matrices and the equivalence results will play an indispensable role in establishing the convergence rates and the asymptotic distributions of our PC estimators. Before doing so, we would like to first introduce an interesting finding related to the matrix $Q$ defined above. One more additional condition is needed.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

**Assumption 5.** $N^{1-\alpha_r}/T^{1/2} \to 0$.

It is worth mentioning that Bai and Ng (2023) impose a weaker condition ($N^{1-\alpha_r}/T \to 0$) for consistency. They do not impose $N^{1-\alpha_r}/T^{1/2} \to 0$ until in proving distributional theory. However, we impose Assumption 5 at an early stage for showing Proposition 3.2, which plays a fundamental role in the sparsity recovery in Section 4. Besides, with Proposition 3.2, we can obtain a sharper result for the convergence rate (see Proposition 3.3(ii) later). Also note that under $N/T \to c \in (0,\infty)$ as in Freyaldenhoven (2022), Assumption 5 is trivial.

PROPOSITION 3.2. *Under Assumptions 1–5, the $r \times r$ matrix $Q = T^{-1}\widetilde{F}'F^0$ is a full rank matrix with probability approaching 1 and*

$$Q_{lk} \equiv Q(l,k) \begin{cases} \asymp_p O_p(N^{\alpha_l-\alpha_k}), \text{ for } 1 \le k \le l \le r, \\ = O_p(1), \text{ for } 1 \le l < k \le r. \end{cases}$$

**Remark 3.** Proposition 3.2 implies a very important property of $Q$ serving as the rotation matrix: $Q$ is (block) *upper triangular* (with different factor strengths) asymptotically.[7] The property is particularly useful for analyzing the sparsity preservation of the PC method in sparse weak factor models. For more discussion, see the next subsection.

The following proposition states the convergence rates of $\widetilde{F}$, $\widetilde{\Lambda}$, and $\widetilde{C}$.

PROPOSITION 3.3. *Under Assumptions 1–5, we have*
(i) $\frac{1}{T}\left\|\widetilde{F} - F^0H\right\|^2 = O_p\left(N^{\alpha_1-2\alpha_r} + N^{2(1-\alpha_r)}T^{-2}\right) = o_p(1)$;
(ii) $\frac{1}{N}\left\|\widetilde{\Lambda} - \Lambda^0Q'\right\|^2 = O_p\left(N^{-\alpha_r-1} + T^{-1}\right) = o_p(1)$;
(iii) $\frac{1}{NT}\left\|\widetilde{C} - C^0\right\|^2 = O_p\left(N^{2(\alpha_1-\alpha_r)-1} + T^{-1}\right) = o_p(1)$.

**Remark 4.** (i) The associated convergence rate for $\widetilde{F}$ in Proposition 6.i of Bai and Ng (2023) is $O_p\left(N^{-\alpha_r} + N^{2(1-\alpha_r)}T^{-2}\right) = o_p(1)$, which is a bit better than ours. The reason is that they use a slightly different rotation matrix from our $H$, as indicated by the comment below their Proposition 6. We employ the rotation matrix $H$ because it is more closely related to $Q$, which will play a significant role in our subsequent analysis. (ii) Note that our rate for $\widetilde{\Lambda}$ is better than what is stated in Proposition 6.ii of Bai and Ng (2023). The reason is that we have utilized the upper triangularity of rotation $Q$ in our proof, which helps sharpen the bound for estimation errors. (iii) It is noted that the convergence rate of $\widetilde{C}$ in general depends on weak factor strengths, or precisely, the discrepancy of strengths between the strongest and weakest factors. It is only when all factor strengths are the same that the convergence rate is maximal and coincides with that under strong factor

[7] In the special case of all factors with the *same* strength, $Q$ is no longer a (block) upper triangular matrix. However, this would not alter the main conclusions in this article.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

models. Also note that this result agrees with Proposition 3 of Bai and Ng (2023) when all the factors have the same factor strength.

**Remark 5.** In the proof of Lemma A.6 in Section C of the Supplementary Material, we show that $H_j = H + O_p(\gamma_{NT})$ for $j = 1, \ldots, 4$, where $\gamma_{NT} = \max\left(N^{\frac{\alpha_1}{2}-\alpha_r}T^{-1/2}, N^{1-\alpha_r}T^{-1}, N^{\alpha_1-2\alpha_r}\right)$. Recall that $H_3 = Q^{-1}$. Lemma A.6 in the Supplementary Material thus implies that $Q^{-1} = H + O_p(\gamma_{NT})$. This equivalence is used in showing the consistency of the estimator $\widetilde{C}$ for the common component. In addition, Lemma A.6 in the Supplementary Material generalizes Lemma 1 of Bai and Ng (2019) for strong factor models, and Lemma 3 of Bai and Ng (2023) for weak factor models with a common factor strength.

For reference later, let $\overline{Q} \equiv A^{-1/2}QA^{1/2}$, and further denote its probability limit as $Q^0$. Also, let $V^0$ denote the probability limit of $NA^{-1}\widetilde{V}$. The following assumption is used to establish the limiting distribution of PC estimators for weak factor models.

**Assumption 6.** The following hold for each $i$ and $t$ as $(N,T) \to \infty$:

(i) $A^{-1/2}\Lambda^{0\prime}e_t \xrightarrow{d} N(0, \Gamma_t)$, where $\Gamma_t = \lim_{N\to\infty} A^{-1/2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i^0\lambda_j^{0\prime} E\left(e_{it}e_{jt}\right)A^{-1/2}$.

(ii) $T^{-1/2}\sum_{t=1}^{T}F_t^0 e_{it} \xrightarrow{d} N(0, \Phi_i)$, where $\Phi_i = \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} E\left(F_t^0 F_s^{0\prime} e_{it}e_{is}\right)$.

One more regularity condition is imposed on sample size $(N,T)$ and the weakest factor strength $\alpha_r$ to guarantee the distributional theory, which is also employed by Assumption C'(iv) in Bai and Ng (2023).

**Assumption 7.** $N^{\frac{3}{2}-\alpha_r}T^{-1} \to 0$.

**Assumption 8.** $\sqrt{T}N^{\alpha_1-2\alpha_r} \to 0$.

Assumption 8 parallels Assumption C(ii) of Bai and Ng (2023), and a smaller gap between $\alpha_1$ and $\alpha_r$ makes the assumption more likely to hold.

We next present the limiting distribution for $\widetilde{F}$, $\widetilde{\Lambda}$, and $\widetilde{C}$. To this end, recall $r_G$ from Section 2.3, and define the $r \times r$ matrix $S^\dagger = \text{diag}(\underbrace{0, \ldots, 0}_{r-r_G}, \underbrace{1 \ldots, 1}_{r_G})$.

THEOREM 3.4. *Under Assumptions 1–8, we have (i)* $A^{1/2}\left(\widetilde{F}_t - H_4'F_t^0\right) \xrightarrow{d} N\left(0, \Psi^0\Gamma_t\Psi^{0\prime}\right)$, *where* $\Psi^0 = V^0Q^0$; *(ii)* $\sqrt{T}(H')^{-1}\left(\widetilde{\lambda}_i - Q\lambda_i^0\right) \xrightarrow{d} N(0, \Phi_i)$; *and (iii)* $\frac{\widetilde{C}_{it}-C_{it}^0}{\sqrt{N^{-\alpha_r}V_{it}+T^{-1}U_{it}}} \xrightarrow{d} N(0,1)$, *where* $V_{it} = \lambda_i^{0\prime}S^\dagger\Sigma_\Lambda^{*-1}\Gamma_t\Sigma_\Lambda^{*-1}S^\dagger\lambda_i^0$ *and* $U_{it} = F_t^{0\prime}\Sigma_F^{-1}\Phi_i\Sigma_F^{-1}F_t^0$.

**Remark 6.** Theorem 3.4 (i) and (ii) are also in Bai and Ng (2023, Prop. 7). Note that Bai and Ng (2023) establish that the matrix $Q$ has a well-defined and -behaved limit, so do $Q^0$ and $V^0$ (appearing in $\Psi^0$) as the limit of $\overline{Q}$.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

### 3.3. Uniform Convergence Rates

In this section, we establish uniform convergence rate results for $\widetilde{F}_t$, $\widetilde{\lambda}_i$, and $\widetilde{C}_{it}$ over $i$ or (and) $t$. These results can be exploited in recovering model sparsity in Section 4, in the factor-augmented forecast regression, and are perhaps also of independent interest.

Given that dependence is allowed across both $i$ and $t$, we next define strong mixing conditions over time and across CS units, respectively. First, let $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{k}^{\infty}$ denote the $\sigma$-algebras generated by $\left\{\left(F_t^0, e_t\right) : -\infty \le t \le 0\right\}$ and $\left\{\left(F_t^0, e_t\right) : k \le t \le \infty\right\}$, respectively. Then, define the mixing coefficient $\alpha_G(k) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{k}^{\infty}} |P(A)P(B) - P(AB)|$. Second, following Ma, Linton, and Gao (2021), we assume that the CS units have a latent ordering indexed by $i^*$ such that the dependence between any two units $i^*$ and $j^*$ decays with the distance $|i^* - j^*|$. Then, we define a mixing rate applied for the random field $\{e_{i^*} : 1 \le i^* \le N\}$,

$$\alpha_e(k) = \sup\left\{|P(A)P(B) - P(AB)| : A \in \sigma(e_{i^*}), B \in \sigma\left(e_{j^*}\right), \text{ and } \left|i^* - j^*\right| = k\right\},$$

where $\sigma(\cdot)$ denotes a sigma-field. For the purpose of estimation, we do not need to know the true labeling $\{i^*\}$. Ma, Linton, and Gao (2021) show that their inference is valid as long as the number of mis-assigned indices is $o\left(N^{1/2}\right)$. In conducting inference, our approach is effective with the true labeling $\{i^*\}$ being completely unknown, and thus further relaxes the assumption by Ma, Linton, and Gao (2021).

Finally, for $k = 1, 2, \ldots$, define $\alpha(k) = \max\{\alpha_G(k), \alpha_e(k)\}$. We now present the additional assumption for establishing the uniform convergence rates.

**Assumption 9.** (i) $\{e_t\}_{t\ge1}$ and $\left\{F_t^0\right\}_{t\ge1}$ are both stationary and ergodic;
(ii) there exist $s_2 > 0$ and $K > 0$ such that $\forall t \in \mathbb{Z}^+, \alpha(t) \le \exp(-Kt^{s_2})$;
(iii) there exist $s_1 > 0, b_1 > 0, s_3 > 0, b_2 > 0$, satisfying $3s_1^{-1} + s_2^{-1} > 1$ and $3s_3^{-1} + s_2^{-1} > 1$, such that $\forall w > 0, P(|e_{it}| > w) \le \exp(-(w/b_1)^{s_1})$ and $P\left(\left|F_{tk}^0\right| > w\right) \le \exp(-(w/b_2)^{s_3})$ for $k = 1, \ldots, r$.

Assumption 9 imposes restrictions on the mixing rate decay and the tail behavior of the factors and errors, as well as a mild growth rate condition on the sample size $(N, T)$, a requirement also employed in Fan, Liao, and Mincheva (2011).

From the previous section, we see that the PC estimators $\widetilde{F}_t$ and $\widetilde{\lambda}_i$ are both consistent up to certain rotation matrices. So to better state the uniform convergence result, we define the rotated factor and factor loading by $F_t^* = H_4' F_t^0$ and $\lambda_i^* \equiv Q\lambda_i^0$, respectively.

THEOREM 3.5. *Under Assumptions 1–9,*
(i) $\sup_i \left|\widetilde{\lambda}_{ik} - \lambda_{ik}^*\right| = O_p\left(\sqrt{(\ln N)/T}\right)$ *for* $k = 1, \ldots, r$;
(ii) $\sup_t \left|\widetilde{F}_{tk} - F_{tk}^*\right| = O_p\left(N^{-\alpha_k/2}\sqrt{\ln T}\right)$ *for* $k = 1, \ldots, r$;
(iii) $\sup_{i,t} \left|\widetilde{C}_{it} - C_{it}^0\right| = O_p((\ln T)^{1/s_3}\sqrt{(\ln N)/T} + N^{-\alpha_r/2}\sqrt{\ln T})$.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

**Remark 7.** Theorem 3.5 (i) implies that the estimation errors for factor loadings are uniformly dominated by $O_p\left(\sqrt{(\ln N)/T}\right)$ across all $i$. This result is very useful for exploring loading sparsity and factor strength in the next section. For each factor, result (ii) provides the uniform convergence rate that depends on the strength of that specific factor. Finally, result (iii) establishes the uniform convergence rate for the common component estimators; the rate is determined by the smallest factor strength, $\alpha_r$, and a parameter $s_3$, which controls the probability tail bound of the factors.

## 4. REVELATION OF SPARSITY

It is well known that latent factor models are subject to rotational indeterminacy. This presents a challenge for revealing the true sparsity structure of the loadings. Specifically, the large-sample properties of the principal component (PC) estimators in Section 3 indicate that loadings are identified only up to a rotation matrix. It is generally believed that such rotation obscures the underlying sparsity. Consequently, interpreting the factors becomes challenging. Interpretation typically relies on associating a factor with the CS units that have nonzero estimated loadings. For example, Ludvigson and Ng (2009) write,

"*Moreover, we caution that any labeling of the factors is imperfect, because each is influenced to some degree by all the variables in our large dataset and the orthogonalization means that no one of them will correspond exactly to a precise economic concept like output or unemployment, which are naturally correlated.*"

However, we show that the sparsity in factor loadings can be preserved even in the presence of the rotation induced by PC estimation. This finding justifies the common, albeit ad hoc, empirical practice of interpreting a factor by associating it with only a small set of observed variables that are correlated with the PC factor (see Pelger and Xiong, 2022).

### 4.1. Preservation of Sparsity Degree by the PC Rotation

The success of sparsity preservation by PC largely hinges on the property of the rotation matrix $Q$. Proposition 3.2 implies that $Q$ is a non-strictly (block) *upper triangular* matrix with full rank in probability.[8] The particular form of matrix $Q$ gives rise to sparsity recovery as $N$ grows. To illustrate, consider a simple two-factor model with factor strengths $\alpha_1 = 0.9$ and $\alpha_2 = 0.7$. Now, $Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = O_p\begin{pmatrix} 1 & 1 \\ N^{-0.2} & 1 \end{pmatrix}$. According to Proposition 3.3(ii), the PC estimator of loading matrix converges to a rotation of $\Lambda^0$: $\Lambda^* = \Lambda^0 Q'$, which

[8] A (lower, upper) triangular matrix is strictly (lower, upper) triangular if its diagonal elements are zero (see Abadir and Magnus, 2005, p. 17).

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

can be written as

$$\Lambda^* = \left(\Lambda^*_{\cdot 1}, \Lambda^*_{\cdot 2}\right) = \left(\Lambda^0_{\cdot 1}, \Lambda^0_{\cdot 2}\right) \begin{pmatrix} Q_{11} & Q_{21} \\ Q_{12} & Q_{22} \end{pmatrix}$$
$$= \left(Q_{11}\Lambda^0_{\cdot 1} + Q_{12}\Lambda^0_{\cdot 2}, Q_{21}\Lambda^0_{\cdot 1} + Q_{22}\Lambda^0_{\cdot 2}\right).$$

Given the approximately upper triangular structure of $Q$, we have

$$\left\|\Lambda^*_{\cdot 1}\right\|_0 = \left\|Q_{11}\Lambda^0_{\cdot 1} + Q_{12}\Lambda^0_{\cdot 2}\right\|_0 \asymp_p N^{0.9} + N^{0.7} \asymp_p N^{0.9}. \tag{4.1}$$

So the rotated loadings for factor 1 preserve its sparsity degree. Given that $\Lambda^*_{\cdot 2} = Q_{21}\Lambda^0_{\cdot 1} + Q_{22}\Lambda^0_{\cdot 2}$, the sparsity degree of factor 2 is contaminated by $Q_{21}\Lambda^0_{\cdot 1}$. Nevertheless, the contamination is vanishing due to diminishing $Q_{21}$. Note that each nonzero element in $Q_{12}\Lambda^0_{\cdot 2}$ is of order $O_p(1)$ and the total number of nonzero elements is of order $N^{0.7}$; the total number of nonzero elements in $Q_{21}\Lambda^0_{\cdot 1}$ is of order $N^{0.9}$, but each nonzero element is of order $O_p\left(N^{-0.2}\right)$ and converges to zero. So the first component $Q_{21}\Lambda^0_{\cdot 1}$ goes to a zero vector asymptotically. Alternatively, we define an *adaptive measure for sparsity degree* for a generic $n \times 1$ vector $a$ as $\|a\|_{\kappa_N,0} = \sum_{j=1}^n \mathbf{1}\left(|a_j| \geq \kappa_N\right)$, where $\kappa_N = N^{-\varrho}$ for some $\varrho \in (0, 0.2)$. Then,

$$\left\|\Lambda^*_{\cdot 2}\right\|_{\kappa_N,0} = \left\|Q_{21}\Lambda^0_{\cdot 1} + Q_{22}\Lambda^0_{\cdot 2}\right\|_{\kappa_N,0} \asymp_p N^{0.7}. \tag{4.2}$$

The statement in (4.1) also certainly holds if the $\ell_0$ -norm is replaced with $\|\cdot\|_{\kappa_N,0}$. This result suggests that the PC rotation can preserve the sparsity degree, up to negligible terms, so that the factor strengths remain unchanged.

To have a better understanding of how the PC rotation acts upon the sparse loading $\Lambda^0$, we illustrate with the two-factor model as above via a simulation. The DGP is similar to that in Section 5.[9] We set $(N, T) = (200, 200)$ and simulate $r = 2$ positively correlated factors.

Panel (a) of Figure 1 shows underlying $\Lambda^0$ in which $\mathcal{L}^0_1 = \{i : i = 1, \ldots, 117\}$ and $\mathcal{L}^0_2 = \{i : i = 97, \ldots, 136\}$. The rotation matrix is calculated as $Q = \begin{pmatrix} -1.07 & 1.46 \\ -0.19 & -0.67 \end{pmatrix}$. Panel (b) illustrates a common concern that $\Lambda^0$ rotated by an arbitrary nonsingular matrix, say $Q^{(b)}$, would become less sparse. To see it clearly, we randomly generate each element of $Q^{(b)}$ as $\mathcal{N}(0,1)$ so that $Q^{(b)} = \begin{pmatrix} -1.44 & -0.98 \\ -1.16 & -0.75 \end{pmatrix}$. It is obvious in Panel (b) that $\mathcal{L}^0_2$ is "inflated" considerably (by the shaded area) such that the two factors turn out to have the same strength. Panel (c) displays the rotation effect by a stylized exactly upper triangular matrix $Q^{(c)}$. $Q^{(c)}$ here is set the same as $Q$ except $Q^{(c)}_{21} = 0$, i.e., $Q^{(c)} = \begin{pmatrix} -1.07 & 1.46 \\ 0 & -0.67 \end{pmatrix}$. Clearly, the upper triangular $Q^{(c)}$ resolves inflation to $\mathcal{L}^0_2$. On the other hand, we see in panel (c) that $\mathcal{L}^0_1$ is somewhat inflated due to

[9] The only exception is that $\mathcal{L}^0_k$ is randomly determined across $i$ in Section 5, while it is fixed as prescribed here for illustration.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

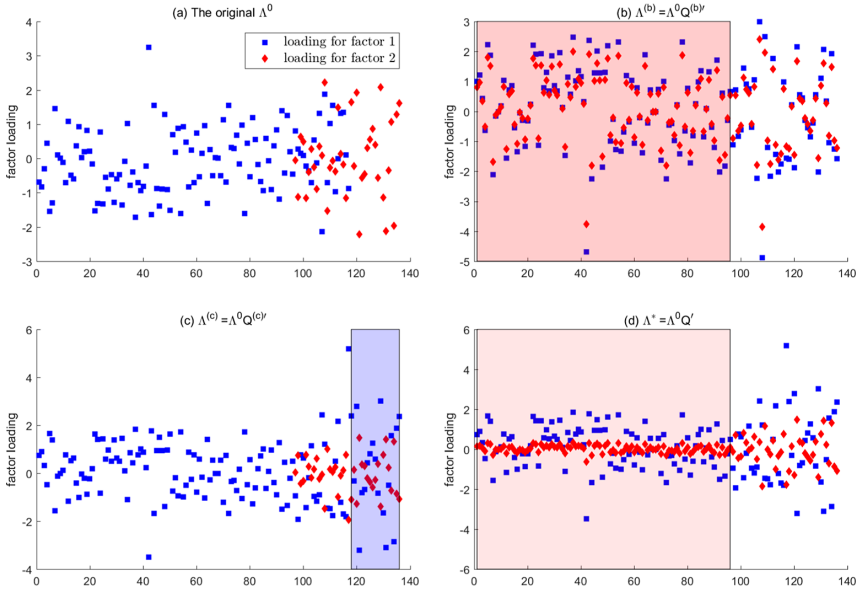



**FIGURE 1.** Illustration of sparsity degree preservation by PC rotation.

the loading on $\{i : i = 118, \ldots, 136\}$ (by the shaded area) being mixed with factor 2's nonzero loadings. Nevertheless, the total number of zero loadings in factor 1 distorted by factor 2 is negligible relative to factor 1's pervasiveness, since factor 2 is relatively weaker. Finally, panel (d) displays $\Lambda^*$ rotated by the real $Q$ in the simulation. The absolute value of $Q_{21}$ is indeed very small as expected in a finite sample ($Q_{21} \asymp_p 200^{-0.2}$), though it is not exactly 0. By virtue of this feature, the inflated part of $\Lambda^*_{\cdot,2}$ (by the shaded area) is uniformly small and vanishing at a certain rate, which can be screened off together with estimation errors to reveal sparsity. We formalize the idea in the next subsection.

## 4.2. Sparsity Recovery with PC Estimators

Our previous analysis in Section 4.1 provides a positive identification result for $\Lambda^*$ rotated by $Q$. In practice, $\Lambda^*$ is not observable, and what is available is the PC estimator $\widetilde{\Lambda}$. In this section, we will present how the recovery of sparsity, not just its degree, is achieved by working with $\widetilde{\Lambda}$. Undoubtedly, we need to take into account the *estimation error* consisting of $\widetilde{\lambda}_i - \lambda_i^* = T^{-1}H'F^{0\prime}e_i + T^{-1}(\widetilde{F} - F^0H)'e_i$. Theorem 3.5(i) lends us a hand stating that the estimation errors are also uniformly vanishing, and thus justifies the sparsity recovery via screening the PC estimator $\widetilde{\lambda}_i$.

We shall show that the set $\mathcal{L}_k^0$ characterizing sparsity for factor $k$ can be recovered approximately well with regulated PC estimators. When the absolute value of $\widetilde{\lambda}_{ik}$ is

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

small, we can set the factor loading as 0. So we choose the nonzero factor loadings with a threshold value $c_{\lambda,NT}$:

$$\widehat{\lambda}_{ik} = \widetilde{\lambda}_{ik}\mathbf{1}\left\{\left|\widetilde{\lambda}_{ik}\right| > c_{\lambda,NT}\right\} \text{ where } c_{\lambda,NT} = \frac{1}{\sqrt{\ln(NT)}}. \tag{4.3}$$

Let us define the estimated support for the $k$th factor loading by $\widehat{\mathcal{L}}_k = \{i : \widehat{\lambda}_{ik} \neq 0\}$. To evaluate the sparsity recovery accuracy of $\widehat{\mathcal{L}}_k$, we denote the *symmetric difference* between the true nonzero set $\mathcal{L}_k^0$ and its estimator $\widehat{\mathcal{L}}_k$ by $\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k = \left(\mathcal{L}_k^0 \backslash \widehat{\mathcal{L}}_k\right) \cup \left(\widehat{\mathcal{L}}_k \backslash \mathcal{L}_k^0\right)$, where $\mathcal{L}_k^0 \backslash \widehat{\mathcal{L}}_k$ is the false negative set and $\widehat{\mathcal{L}}_k \backslash \mathcal{L}_k^0$ is the false positive set. Clearly, $\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k$ summarizes two types of errors.

To accommodate factors with identical strength, we define a set

$$\omega(k) = \{l : \alpha_l = \alpha_k, 1 \leq l \leq r\}. \tag{4.4}$$

For any $l \in \omega(k)$, $\alpha_l = \alpha_k$.

PROPOSITION 4.1. *Under Assumptions 1–5, 8, and 9, for $k = 1, \ldots, r$, we have:*
(i) *if $\alpha_k$ is unique, then $\frac{|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k|}{N^{\alpha_k}} = o_p(1)$;*
(ii) *if $\alpha_k$ is not unique such that $|\omega(k)| \geq 2$, then $\frac{|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k|}{N^{\alpha_k}} = o_p(1)$ if $\frac{\left|\cup_{k' \in \omega(k)} \mathcal{L}_{k'}^0 \backslash \mathcal{L}_k^0\right|}{N^{\alpha_k}} = o(1)$; otherwise, $\frac{|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k|}{N^{\alpha_k}} = O_p(1)$.*

**Remark 8.** In general, we cannot recover the set $\mathcal{L}_k^0$ sharply in the sense that $|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k| = o_p(1)$. We have $|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k| = O_p(N^{\alpha_k})$ or $o_p(N^{\alpha_k})$ depending on whether $\alpha_k$ is unique or the overlap degree of loading supports from the same strength group. In general, the recovered sparsity based on PC estimates is perhaps less sparse than the true underlying sparsity for factor $k$. These results seem weak but they are general enough to identify the factor strengths. For the special case of nested sparsity, i.e., $\mathcal{L}_k^0 \supseteq \left(\cup_{l \geq k} \mathcal{L}_l^0\right) \cup \left(\cup_{k' \in \omega(k)} \mathcal{L}_{k'}^0\right)$, we can show that the recovery is almost sharp in the sense that $|\mathcal{L}_k^0 \triangle \widehat{\mathcal{L}}_k| = O_p(1)$.

**Remark 9.** (i) In the sparsity recovery, $c_{\lambda,NT}$ plays the key role of screening off noises due to rotation and estimation errors. We show that the noises are $O_p\left(N^{-\zeta} + [(\ln N)/T]^{1/2}\right)$ uniformly, where $\zeta$ is the minimum discrepancy between distinct factor strengths, i.e., $\zeta = \min_{1 \leq k \leq G-1}\left(\alpha_{[k]} - \alpha_{[k+1]}\right)$. So we set $c_{\lambda,NT} = [\ln(NT)]^{-1/2}$ to dominate the noise. (ii) Alternatively, one can use threshold $\widetilde{c}_{\lambda,NT} = c\,[\ln(NT)]^{-1/2}$ with a tuning parameter $c$. We have tried $c$ over the range $[0.5, 1.5]$, and found the results robust to the tuning parameter; moreover, $c$ can be determined by some adaptive methods such as cross validation. Given that the data are all standardized as in Bai and Ng (2002), we focus on the simple threshold $c_{\lambda,NT}$. In both simulations and the empirical application, our tuning parameter-free threshold value $c_{\lambda,NT}$ works reasonably well. Note that a simple threshold value is also adopted in Fan, Liao, and Yao (2015).

**Remark 10.** Note that the uniform order of noises depends on the minimum discrepancy between different factor strengths (see the previous remark). So the

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

recovery of loading sparsity works better when the gaps between factor strengths are bigger. When the factor strengths of, say $\alpha_j$ and $\alpha_k$, become very close, our method is likely to render the same sparsity for both factor loadings, and may not perform well in distinguishing these two. Nevertheless, perhaps it does not hurt much since the true loading sparsity for factors $j$ and $k$ here is possibly very similar in the first place. Also notice that in this case it will not cause a problem for working with factor $l$ with $\alpha_l$ far away from $\alpha_j$ and $\alpha_k$.

**Remark 11.** (i) Given sparsified loading estimator $\widehat{\lambda}_{ik}$, it is natural to consider further updating the estimator of factors. However, it is easy to show that the updated estimators of factors have the same convergence rates and asymptotical distributions as simple PC estimators. It may be interesting to compare their finite sample performance and further investigate their higher-order asymptotics. We leave it for further research. (ii) As the loadings would be of interest in many empirical applications, one may consider further refining the PC estimate to reduce its contamination, to achieve more precise estimation and sparsity recovery. Such consideration is reasonable especially when the strengths of some factors are close, as the loading vectors of these two groups are more intermingled. This may call for a more complicated procedure, e.g., in further combining SOFAR under its rotation-specific assumptions, and raise considerable theoretical challenges. We also leave it for further research.

We would like to mention that UY's (2023a) SOFAR estimator recovers sparsity alternatively by $\ell_1$ regularization. Interestingly, if we contrast the estimate of factor loadings by PC (their Figures 2 and 7) with that by SOFAR (their Figures 9 and 11) in their empirical results, we immediately realize that PC and SOFAR estimates are almost identical, except for many small noises introduced by PC. This observation suggests the validity of combining PC estimate with proper screening to recover sparsity, which in fact underlines the method used in our article.

As a byproduct of sparsity recovery, the factor strength estimates are readily available. Specifically, for $k = 1, \ldots, r$, define $\widehat{D}_k = \sum_{i=1}^{N} \mathbf{1}\left\{\left|\widetilde{\lambda}_{ik}\right| > c_{\lambda,NT}\right\}$, where $c_{\lambda,NT} = [\ln(NT)]^{-1/2}$. Our estimator for $\alpha_k$ is simply given by

$$\widehat{\alpha}_k = \frac{\ln \widehat{D}_k}{\ln N}. \tag{4.5}$$

Note that the above estimator $\widehat{\alpha}_k$ in (4.5) is also employed by UY (2023a) for latent factors and by BKP (2021) primarily for observed factors.

THEOREM 4.2. *Under Assumptions 1–5, 8, and 9, for* $k = 1, \ldots, r, \widehat{\alpha}_k \xrightarrow{p} \alpha_k$.

**Remark 12.** In practice, one can rely on the estimated factor strength to determine a latent factor is strong or weak. As argued by BKP (2021), the precise estimation of factor strength relies on a large cross-section sample size $N$. So we follow BKP (2021) and suggest a conservative way to treat a factor with (estimated)

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

strength above 0.95 as strong enough, while a factor with strength below 0.90 as weak enough.

## 5. MONTE CARLO SIMULATIONS

In this section, we study the finite sample performance of our proposed estimators for sparsity-induced weak factor models. Our study includes a comparison of simple PC estimators with various regularized estimators for estimating factors, loadings, common components, and factor strengths. We also propose the use of singular value thresholding (SVT) to estimate the number of factors, as it is simpler to apply and compute. The details of this approach and its theoretical justification are provided in Section D of the Supplementary Material.

### 5.1. The Data-Generating Process

We consider the following DGP:

$$X_{it} = \lambda_i^{0\prime} F_t^0 + e_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T,$$
$$F_{t1}^0 = 0.5 F_{t-1,1}^0 + u_{t1},\ \text{and}\ F_{tk}^0 = (-0.8)^k F_{t1}^0 + u_{tk},\ k = 2, \ldots, r.$$

The simulated factors are correlated with each other with various degrees. We let $u_{tk}$ be mutually uncorrelated $N(0,1)$ for $k = 1, \ldots, r$. To specify the factor loadings $\lambda_{ik}^0$, for each $k = 1, \ldots, r$, we first *randomly* select $\lfloor N^{\alpha_k} \rfloor$ of $\{\lambda_{ik}^0\}_{i=1}^N$ and specify them as iid $N(0,1)$, and then set the rest of $\{\lambda_{ik}^0\}_{i=1}^N$ as zero.

For the $N \times 1$ vector $e_t$, we specify the (marginal) distribution of $e_{it}$ as the student-$t$ (5) to allow for heavy tails. The CS dependence across $e_{it}$ is admitted through the $N \times N$ covariance matrix $\Sigma_e$ as follows. $\Sigma_e = \text{diag}\{\Sigma_1, \ldots, \Sigma_{N/4}\}$ as a block-diagonal matrix with $4 \times 4$ blocks located along the main diagonal. Each $\Sigma_i$ is assumed to be $I_4$ initially. We then randomly choose $\lfloor N^{0.3} \rfloor$ blocks among them and make them non-diagonal by setting $\Sigma_i(m,n) = 0.5^{|m-n|}$. The design of CS dependence follows Fan, Liao, and Yao (2015) except that the dependence is stronger here.

We have tried simulations with the number of factors $r = 3$ and 5. For $r = 3$, we set $\alpha = (0.9, 0.75, 0.6)$; while for $r = 5$, we set $\alpha = (1, 0.9, 0.8, 0.7, 0.6)$. To preserve space, we only include results under $r = 3$ in the main text and delegate the results under $r = 5$ to Section F.1 of the Supplementary Material. The replication number of simulations is set as 2,000.

### 5.2. Simulation Results

We first compare different methods to determine the number of factors with ours (WZ). The alternative selection rules range from the $IC_{p1}$ by Bai and Ng (2002, BN), Guo, Chen, and Tang (2023, GCT), Freyaldenhoven (2022, FR), ED by

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

**TABLE 1.** Estimating the number of factors when $r = 3$

| | | RMSE | | | | | | Bias | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $T$ | WZ | BN | GCT | FR | ED | AH | WZ | BN | GCT | FR | ED | AH |
| 100 | 100 | 0.291 | 0.676 | 0.939 | 1.962 | 0.415 | 1.998 | 0.078 | −0.446 | −0.882 | −1.944 | 0.034 | −1.998 |
| | 200 | 0.213 | 0.425 | 0.867 | 1.967 | 0.412 | 2.000 | −0.028 | −0.174 | −0.752 | −1.950 | 0.083 | −2.000 |
| | 400 | 0.269 | 0.311 | 0.782 | 1.972 | 0.952 | 1.999 | −0.070 | −0.094 | −0.608 | −1.959 | 0.322 | −1.998 |
| 200 | 100 | 0.183 | 0.594 | 0.912 | 1.968 | 0.282 | 2.000 | 0.028 | −0.347 | −0.832 | −1.944 | 0.058 | −2.000 |
| | 200 | 0.143 | 0.224 | 0.722 | 1.949 | 0.289 | 2.000 | −0.015 | −0.046 | −0.519 | −1.904 | 0.063 | −2.000 |
| | 400 | 0.180 | 0.092 | 0.454 | 1.961 | 0.397 | 2.000 | −0.033 | −0.008 | −0.199 | −1.927 | 0.075 | −2.000 |
| 400 | 100 | 0.140 | 0.676 | 0.928 | 1.950 | 0.225 | 2.000 | 0.018 | −0.455 | −0.862 | −1.902 | 0.048 | −2.000 |
| | 200 | 0.077 | 0.217 | 0.632 | 1.795 | 0.288 | 2.000 | −0.004 | −0.045 | −0.397 | −1.609 | 0.064 | −2.000 |
| | 400 | 0.102 | 0.032 | 0.225 | 1.583 | 0.261 | 1.999 | −0.010 | 0.000 | −0.012 | −1.253 | 0.055 | −1.998 |

Onatski (2010), and Ahn and Horenstein (2013, AH). $r_{\max}$ is set to be 8 if needed.[10] The root mean square error (RMSE) and bias of the estimated number of factors by each method are reported in Table 1. It is obvious that GCT, FR, and AH are all subject to underestimation of $r$ in the presence of weak factors. ED and BN are not very bad but they are outperformed by our proposed method, since they tend to over- and under-estimate $r$, respectively. The ED estimator is not as effective as found previously in weak factor models, e.g., by UY (2023a), implying that the ED performance may be sensitive to the choice of the wedge parameter $\delta$, as remarked in Section D of the Supplementary Material. Our proposed estimator of factor numbers is outstanding against all alternatives at almost all sample sizes.

To study the finite sample performance of PC estimators under weak factor models, we compare PC regression with the SOFAR proposed by UY (2023a, 2023b), assuming the true number of factors to be known. Based on Lasso penalization on factor loadings for sparsity, UY (2023a) develop the Adaptive (Ada) SOFAR estimator. They later make an inference based on the SOFAR estimator by introducing the Debiased (Deb) SOFAR estimator to recover its asymptotic normality. Furthermore, they address the multiple testing problem of loading sparsity and construct the Resparsified (Res) SOFAR estimator to fulfill FDR control. Note that the various SOFAR estimators in UY (2023b) come from penalized regression which targets sparsity explicitly. Thus, it seems natural to suggest that one should be better off using their approaches to deal with weak factor models than using PC. However, as we have mentioned before, the simple PC estimators enjoy a nice property of automatic sparsity recognition, so that PC could even outperform the complicated SOFAR.

Given our emphasis on sparsity recovery in this article, we are also curious about how well sparsity can be recovered based on PCA. To address this question, we look into the FDR and power performance for recover sparsity of the loading matrix, following UY (2023b). To define the two terms, let $\mathcal{S}$ denote an index set of nonzero elements (e.g., $\mathcal{L}_k^0$), and $\widehat{\mathcal{S}}$ be a set discovered by some procedure (e.g.,

[10]The results with different $r_{\max}$ are robust, provided in Section F.3 of the Supplementary Material.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

**TABLE 2.** FDR and power under $r = 3$

| $N$ | $T$ | $\text{FDR}_1$ | $\text{FDR}_2$ | $\text{FDR}_3$ | $\overline{\text{FDR}}$ | $\text{Power}_1$ | $\text{Power}_2$ | $\text{Power}_3$ | $\overline{\text{Power}}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | Panel A: PC+Screening | | | | | | | |
| 100 | 100 | 0.208 | 0.412 | 0.448 | 0.295 | 0.806 | 0.434 | 0.492 | 0.657 |
| | 200 | 0.212 | 0.403 | 0.431 | 0.292 | 0.845 | 0.445 | 0.532 | 0.688 |
| | 400 | 0.201 | 0.396 | 0.416 | 0.281 | 0.887 | 0.462 | 0.556 | 0.721 |
| 200 | 100 | 0.215 | 0.430 | 0.438 | 0.297 | 0.838 | 0.450 | 0.535 | 0.694 |
| | 200 | 0.213 | 0.399 | 0.369 | 0.280 | 0.872 | 0.474 | 0.579 | 0.727 |
| | 400 | 0.206 | 0.390 | 0.358 | 0.271 | 0.909 | 0.497 | 0.612 | 0.760 |
| 400 | 100 | 0.209 | 0.419 | 0.355 | 0.277 | 0.852 | 0.492 | 0.623 | 0.734 |
| | 200 | 0.205 | 0.404 | 0.255 | 0.257 | 0.893 | 0.507 | 0.677 | 0.770 |
| | 400 | 0.201 | 0.405 | 0.245 | 0.254 | 0.923 | 0.517 | 0.687 | 0.793 |
| | | Panel B: SOFAR_Adaptive | | | | | | | |
| 100 | 100 | 0.018 | 0.085 | 0.252 | 0.080 | 0.737 | 0.729 | 0.723 | 0.733 |
| | 200 | 0.011 | 0.084 | 0.267 | 0.079 | 0.801 | 0.790 | 0.776 | 0.794 |
| | 400 | 0.203 | 0.483 | 0.687 | 0.403 | 0.908 | 0.736 | 0.675 | 0.827 |
| 200 | 100 | 0.017 | 0.064 | 0.188 | 0.056 | 0.730 | 0.709 | 0.717 | 0.723 |
| | 200 | 0.010 | 0.054 | 0.170 | 0.046 | 0.803 | 0.786 | 0.781 | 0.795 |
| | 400 | 0.208 | 0.511 | 0.635 | 0.396 | 0.905 | 0.755 | 0.771 | 0.848 |
| 400 | 100 | 0.202 | 0.361 | 0.348 | 0.249 | 0.785 | 0.408 | 0.543 | 0.662 |
| | 200 | 0.202 | 0.456 | 0.470 | 0.307 | 0.856 | 0.623 | 0.660 | 0.775 |
| | 400 | 0.204 | 0.543 | 0.578 | 0.374 | 0.897 | 0.757 | 0.786 | 0.849 |
| | | Panel C: SOFAR_Resparsified | | | | | | | |
| 100 | 100 | 0.046 | 0.177 | 0.338 | 0.142 | 0.785 | 0.774 | 0.781 | 0.781 |
| | 200 | 0.036 | 0.129 | 0.317 | 0.115 | 0.851 | 0.843 | 0.826 | 0.845 |
| | 400 | 0.212 | 0.532 | 0.712 | 0.435 | 0.930 | 0.787 | 0.717 | 0.860 |
| 200 | 100 | 0.063 | 0.211 | 0.346 | 0.157 | 0.764 | 0.727 | 0.756 | 0.753 |
| | 200 | 0.035 | 0.126 | 0.273 | 0.099 | 0.849 | 0.843 | 0.843 | 0.847 |
| | 400 | 0.218 | 0.558 | 0.665 | 0.435 | 0.925 | 0.792 | 0.826 | 0.876 |
| 400 | 100 | 0.221 | 0.588 | 0.690 | 0.416 | 0.838 | 0.493 | 0.613 | 0.725 |
| | 200 | 0.216 | 0.605 | 0.721 | 0.440 | 0.886 | 0.509 | 0.648 | 0.764 |
| | 400 | 0.215 | 0.610 | 0.739 | 0.463 | 0.920 | 0.613 | 0.760 | 0.824 |

$\widehat{\mathcal{L}}_k$). Then, $\text{FDR} = \mathbb{E}[\text{FDP}]$ with $\text{FDP} = \frac{|\mathcal{S}^c \cap \widehat{\mathcal{S}}|}{|\widehat{\mathcal{S}}| \vee 1}$ and $\text{Power} = \mathbb{E}\left[\frac{|\mathcal{S} \cap \widehat{\mathcal{S}}|}{|\mathcal{S}| \vee 1}\right]$. Specifically, $\text{FDR}_k$ and $\text{Power}_k$ denote the results for sparsity for factor $k$'s loadings, while $\overline{\text{FDR}}$ and $\overline{\text{Power}}$ are for the overall factors' loadings.[11] Table 2

[11] For our specific purpose, $\text{FDR}_k$ and $\text{Power}_k$ are perhaps more relevant than the "maximum cosine similarity" in Freyaldenhoven (2023), which measures the correlation between the estimated loading and the true one for each factor.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

**TABLE 3.** Estimation of factor strength when $r = 3$ with $\alpha = (0.9, 0.75, 0.6)$

| | | RMSE | | | Bias | | |
|---|---|---|---|---|---|---|---|
| $N$ | $T$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\alpha}_3$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\alpha}_3$ |
| | | Panel A: PC+Screening | | | | | |
| 100 | 100 | 0.014 | 0.047 | 0.138 | 0.000 | 0.006 | 0.100 |
| | 200 | 0.014 | 0.048 | 0.169 | 0.002 | 0.003 | 0.073 |
| | 400 | 0.014 | 0.049 | 0.208 | 0.003 | 0.007 | 0.048 |
| 200 | 100 | 0.010 | 0.048 | 0.126 | 0.002 | 0.028 | 0.111 |
| | 200 | 0.009 | 0.045 | 0.138 | 0.002 | 0.023 | 0.091 |
| | 400 | 0.009 | 0.045 | 0.166 | 0.004 | 0.026 | 0.075 |
| 400 | 100 | 0.007 | 0.053 | 0.114 | 0.002 | 0.044 | 0.101 |
| | 200 | 0.006 | 0.048 | 0.103 | 0.002 | 0.040 | 0.079 |
| | 400 | 0.006 | 0.052 | 0.115 | 0.003 | 0.042 | 0.065 |
| | | Panel B: SOFAR_Debiased | | | | | |
| 100 | 100 | 0.017 | 0.094 | 0.085 | 0.004 | −0.060 | −0.019 |
| | 200 | 0.023 | 0.065 | 0.106 | 0.019 | 0.028 | 0.067 |
| | 400 | 0.030 | 0.103 | 0.162 | 0.028 | 0.096 | 0.146 |
| 200 | 100 | 0.012 | 0.095 | 0.059 | 0.000 | −0.071 | −0.021 |
| | 200 | 0.017 | 0.052 | 0.084 | 0.014 | 0.025 | 0.062 |
| | 400 | 0.025 | 0.095 | 0.148 | 0.024 | 0.091 | 0.138 |
| 400 | 100 | 0.010 | 0.102 | 0.050 | −0.004 | −0.085 | −0.033 |
| | 200 | 0.013 | 0.043 | 0.054 | 0.011 | 0.023 | 0.036 |
| | 400 | 0.020 | 0.088 | 0.114 | 0.019 | 0.086 | 0.105 |
| | | Panel C: SOFAR_Resparsified | | | | | |
| 100 | 100 | 0.024 | 0.067 | 0.095 | 0.020 | 0.044 | 0.064 |
| | 200 | 0.030 | 0.103 | 0.130 | 0.028 | 0.095 | 0.106 |
| | 400 | 0.036 | 0.137 | 0.179 | 0.035 | 0.135 | 0.164 |
| 200 | 100 | 0.017 | 0.062 | 0.096 | 0.015 | 0.057 | 0.085 |
| | 200 | 0.025 | 0.099 | 0.129 | 0.024 | 0.098 | 0.121 |
| | 400 | 0.031 | 0.130 | 0.162 | 0.031 | 0.129 | 0.157 |
| 400 | 100 | 0.013 | 0.057 | 0.090 | 0.012 | 0.054 | 0.081 |
| | 200 | 0.021 | 0.093 | 0.104 | 0.020 | 0.092 | 0.098 |
| | 400 | 0.027 | 0.123 | 0.140 | 0.026 | 0.123 | 0.136 |

shows that the results of FDR and power are better for stronger factors using either the PC or SOFAR-based method. Admittedly, the PC based method is inferior to SOFAR-based ones at relatively small or medium sample sizes. This is expected as the SOFAR-based methods explicitly target sparsity by additionally employing regularization or FDR control, while the simple PC estimators are inevitably subject to contamination distortion discussed in Section 4.1. Nevertheless, it is

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

a bit surprising to see that FDR is abnormally high when either $N$ or $T$ is large for SOFAR, while it is decreasing with sample sizes for PC.[12]

Moreover, in additional simulation, we try to work with very small gaps of factor strengths and present the FDR and power results for sparsity recovery in Table 11 in the Supplementary Material, which compares with Table 2 in the article with larger gaps of $\alpha = (0.9, 0.75, 0.6)$. The results show that under the smaller gaps of strength, for factor 1, FDR is increasing from around 0.2 to around 0.3, although its power is quite robust; for factor 2, there is only mild deterioration for its FDR and power; the weakest factor 3 suffers most from the shrinking gaps. So perhaps we need to be cautious to work with a very weak factor whose strength is close to others.

We further investigate the estimated factor strength by various methods in Table 3. All estimates become less accurate as the true strength degree decreases, as expected. It is admitted that our proposed strength estimate (PC+Screening) suffers more of overestimation with fairly weak factors, e.g., those with $\alpha_k = 0.6$. This is likely attributed to the contamination discussed in Section 4.1, and also echoes the finding in Table 2. Otherwise, our estimate is comparable to, and sometimes even better than, those from SOFAR, which is designed deliberately for sparsity recovery, especially for strength $\alpha_k \in [0.7, 0.9]$. The message here again delivers the usefulness of our factor strength estimator, which works reasonably well while avoids involving heavy computation as in SOFAR. In addition, given the discussion in Remark 9, we also replace the screening value $c_{\lambda,NT}$ with $\widetilde{c}_{\lambda,NT} = c[\ln(NT)]^{-1/2}$, and report the estimated factor strengths with $c = 0.8$ and 1.2 in Section F.2 of the Supplementary Material, which do not change much.

## 6. CONCLUSION

This article analyzes the well-known approach of PCA, and derives several novel properties of it to deal with weak latent factor models with sparse factor loadings. It unveils an interesting and fundamental fact that the PC estimators can preserve the sparsity in estimated factor loadings for sparsity-induced weak factor models. This fact facilitates the derivation of asymptotic properties of PC estimators, enables us to recover the sparsity of loadings, and estimates the strengths of each factor. In addition, the determination of the number of factors in weak factor models is also investigated. The numerical studies confirm that our proposed approach works reasonably well in finite samples.

## SUPPLEMENTARY MATERIAL

Jie Wei and Yonghui Zhang (2025): Supplement to "Can Principal Component Analysis Preserve the Sparsity in Factor Loadings?", Econometric

[12]We have also compared with the debiased SOFAR and the method proposed by Freyaldenhoven (2023). Each of the two methods results in extremely high FDR.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100303.

## REFERENCES


Abadir, K., & Magnus, J. (2005). *Matrix Algebra*. Cambridge University Press.

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

Armstrong, T. B., Weidner, M., & Zeleneev, A. (2025). Robust estimation and inference in panels with interactive fixed effects. *Journal of Political Economy*. Forthcoming.

Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519), 1182–1198.

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

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

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

Bai, J., & Ng, S. (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics*, 212(1), 78–96.

Bai, J., & Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2), 1893–1916.

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2021). Measurement of factor strength: Theory and practice. *Journal of Applied Econometrics*, 36(5), 587–613.

Chao, J. C., Liu, Y., & Swanson, N. R. (2022). Consistent factor estimation and forecasting in factor-augmented VAR models. Working Paper, Maryland University.

Chao, J. C., & Swanson, N. R. (2022). Selecting the relevant variables for factor estimation in FAVAR models, Working Paper, Maryland University.

Choi, I., Lin, R., & Shin, Y. (2023). Canonical correlation-based model selection for the multilevel factors. *Journal of Econometrics*, 233(1), 22–44.

Cochrane, J. (2011). Presidential address: Discount rates. *Journal of Finance*, 66(4), 1047–1108.

Dello-Preite, M., Uppal, R., Zaffaroni, P., & Zviadadze, I. (2024). Cross-sectional asset pricing with unsystematic risk. Working Paper, Imperial College London.

Despois, T., & Doz, C. (2023). Identifying and interpreting the factors in factor models via sparsity: Different approaches. *Journal of Applied Econometrics*, 38(4), 451–667.

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.

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

Fan, J., Liao, Y., & Yao, J. (2015). Power enhancement in high-dimensional cross sectional tests. *Econometrica*, 83(4), 1497–1541.

Freyaldenhoven, S. (2022). Factor models with local factors determining the number of relevant factors. *Journal of Econometrics*, 229(1), 80–102.

Freyaldenhoven, S. (2023). Identification through sparsity in factor models: The $\ell_1$-rotation criterion. Working Paper, Federal Reserve Bank of Philadelphia.

Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129(7), 1947–1990.

Giglio, S., Xiu, D., & Zhang, D. (2023). Prediction when factors are weak, Working Paper, NBER.

Giglio, S., Xiu, D., & Zhang, D. (2025). Test assets and weak factors. *The Journal of Finance*, 80(1), 259–319.

Guo, X., Chen, Y., & Tang, C. Y. (2023). Information criteria for latent factor models: A study on factor pervasiveness and adaptivity. *Journal of Econometrics*, 233(1), 237–250.


https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press

Jiang, P., Uematsu, Y., & Yamagata, T. (2024). Revisiting asymptotic theory for principal component estimators of approximate factor models, Working Paper, Hitotsubashi University.

Kim, S., Raponi, V., & Zaffaroni, P. (2024). Testing for weak factors in asset pricing, Working Paper, Imperial College Business School.

Kristensen, J. T. (2017). Diffusion indexes with sparse loadings. *Journal of Business & Economic Statistics*, 35(3), 434–451.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12), 5027–5067.

Ma, S., Linton, O., & Gao, J. (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics*, 222(1), 295–323.

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

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2), 244–258.

Pelger, M., & Xiong, R. (2022). Interpretable sparse proximate factors for large dimensions. *Journal of Business & Economic Statistics*, 40(4), 1642–1664.

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162.

Su, L., & Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29, 1079–1135.

Uematsu, Y., & Yamagata, T. (2023a). Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, 41(1), 126–139.

Uematsu, Y., & Yamagata, T. (2023b). Inference in sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, 41(1), 213–227.

https://doi.org/10.1017/S0266466625100303 Published online by Cambridge University Press