WORKING PAPERS

N° 1219

May 2021

“Factor and factor loading augmented estimators for panel regression”

Jad Beyhum and Eric Gautier

Toulouse
School of
Economics

Factor and factor loading augmented estimators for panel regression

Jad Beyhum[*] Eric Gautier[†]

Abstract

This paper considers linear panel data models where the dependence of the regressors and the unobservables is modelled through a factor structure. The asymptotic setting is such that the number of time periods and the sample size both go to infinity. Non-strong factors are allowed and the number of factors can grow to infinity with the sample size. We study a class of two-step estimators of the regression coefficients. In the first step, factors and factor loadings are estimated. Then, the second step corresponds to the panel regression of the outcome on the regressors and the estimates of the factors and the factor loadings from the first step. Different methods can be used in the first step while the second step is unique. We derive sufficient conditions on the first-step estimator and the data generating process under which the two-step estimator is asymptotically normal. Assumptions under which using an approach based on principal components analysis in the first step yields an asymptotically normal estimator are also given. The two-step procedure exhibits good finite sample properties in simulations.

KEYWORDS: panel data, interactive fixed effects, factor models, flexible unobserved heterogeneity, principal components analysis.

[*] ORSTAT, KU Leuven, jad.beyhum@kuleuven.be. This work was undertaken at the Toulouse School of Economics, Université Toulouse Capitole.

[†] Toulouse School of Economics, Université Toulouse Capitole, eric.gautier@tse-fr.eu.

Financial support from the European Research Council (ERC POEMH grant agreement No. 337665) and the French National Research Agency (grant ANR-17-EURE-0010) is gratefully acknowledged. The authors thank Domenico Giannone, Jihyun Kim, Pascal Lavergne, Thierry Magnac and Nour Meddahi for helpful comments and ideas.

1 Introduction

This paper considers inference on $\beta \in \mathbb{R}^K$ in the following model:

$$Y_{it} = \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_j + E_{it}, , \tag{1.1}$$

where the data consists of the outcome Y_{it} and the regressors X_{kit} for all $k = 1, \dots, K$, $i = 1, \dots, N$ and $t = 1 \dots, T$. The random vectors λ_i and f_t in $\mathbb{R}^{r_N}$ are factor loadings and factors, δ is a nonrandom vector in $\mathbb{R}^{r_N}$, r_N is the number of factors.

This is a panel data model with interactive fixed effects (see **?**). It allows for flexible cross-section and serial correlation thanks to the factor structure in the regression error. Several techniques have been developed to estimate this model. **?** proposes to estimate jointly the regression coefficient and the factors and factor loadings. **?** and **?** study a nuclear-norm penalized estimator. In contrast, the CCE estimator of **?** and the factor-augmented regression estimator studied in **?** and **?** model the dependence between the regressors and the unobservables $\sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_j + E_{it}$. They assume that, for $k \in \{1, \dots, K\}$, there exists $\lambda_{k1}, \dots, \lambda_{kN}$ which are random vectors in $\mathbb{R}^{r_N}$ and mean-zero errors $E_1, \dots, E_K$ which are $N \times T$ random matrices such that $X_{kit} = \sum_{r=1}^{r_N} \lambda_{kir} f_{tr} + E_{kit}$ for $k \in \{1, \dots, K\}$. This means that the regressors have a factor structure with the same factors as the error term but possibly different factor loadings.

In the papers of **?**, **?** and **?**, a strong factor assumption is imposed. It means that the ratio of the singular values of Γ and $\sqrt{NT}$ has a finite deterministic limit as $N, T \to \infty$, where $\Gamma_{it} = \sum_{r=1}^{r_N} \lambda_{ij} f_{tj}$. It holds if $(1/N) \sum_{i=1}^{N} \lambda_i \lambda_i^\top \ (1/T) \sum_{t=1}^{T} f_t f_t^\top$ has a finite deterministic limit in probability. The number of factors is also assumed to be fixed with the sample size. It is worth noting that some papers have sought to relax these assumptions in the context of the CCE (**?**) and factor augmented (**?**) estimators.

This paper proposes instead to model the dependence of the regressors with both the factors and the factor loadings by assuming that there exists $\delta_k \in \mathbb{R}^{r_N}$ for $k \in \{1, \dots, K\}$ and errors $E_1, \dots, E_K$ which are $N \times T$ random matrices such that

$$X_{kit} = \sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_{kj} + E_{kit}, \ k \in \{1, \dots, K\}. \tag{1.2}$$

The role of the vectors $\delta, ..., \delta_K$ is to model the dependence between the regressors and the unobservables $\sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_j + E_{it}$. The structure that we impose can be seen as the generalisation to dimension 3 (the third dimension being the one of variables) of the usual factor models for matrices as in **?**. Such a modelling was already introduced in the psychometrics literature in **?** and **?**. The mathematical foundations behind this approach lie in the tensor decomposition literature, see **?** for a survey.

We study a class of two-step estimators of the proposed model ((1.1) and (1.2)). In the first step , the factors and the factor loadings are estimated. Then, in the second step the outcome is regressed on the covariates augmented by estimates of the factors and the factor loadings. We provide sufficient conditions on the first-step estimator under which the two-step estimator is asymptotically normal. We present assumptions under which a first-step estimator based on principal components analysis (henceforth PCA) satisfies these conditions. All the results are developed under an asymptotic regime where the sample size N goes to infinity and T is a function of N going to infinity with N. Moreover, the number of factors is unknown and allowed to grow (possibly to infinity) with the sample size. Factors are not assumed to be strong. The proposed principal components augmented estimator exhibits better finite sample properties than alternatives in Monte-Carlo simulations.

When a strong factor assumption is imposed and the number of factors is assumed to be fixed, the proposed two-step estimator is found to be asymptotically normal under weaker conditions on N and T than for the factor-augmented estimator in **?**. This suggests that augmenting the panel regression with estimates of the factor loadings leads to improved estimation properties. The estimator of Section 4.7.1 of **?** is a special case of the two-step procedure of this paper. In this other article, a first-step estimator based on hard-thresholding of a nuclear-norm penalized estimator is used. The procedure is pivotal in the sense that it does not require knowledge of the variance of the error terms and that the thresholding level is data-driven. The first-step estimator uses a penalty which level depend on the distribution of the operator norms of the errors while the approach with PCA that we develop here does not. It also relies on the fact that a compatibility constant is bounded away from 0 with probability approaching 1. Such an assumption is absent in the present paper.

This paper is organized as follows. The two-step estimator is introduced in Section 2. Sufficient conditions for asymptotic normality are derived in Section 3. Section 4 is devoted to the analysis of the two-step procedure when PCA is used in the first step. Section 5 describes our simulations. All the proofs are deferred to the Appendix.

Preliminaries. The transpose of a $N \times T$ matrix A is written $A^\top$ and its trace is $\mathrm{tr}(A)$. Its k^{th} singular value is $\sigma_k(A)$ and $\mathrm{rank}(A)$ is its rank. $A = \sum_{k=1}^{\mathrm{rank}(A)} \sigma_k(A) u_k(A) v_k(A)^\top$ is the singular value decomposition of A, where $\{u_k\left(A\right)\}_{k=1}^{\mathrm{rank}(A)}$ is a family of orthonormal vectors of $\mathbb{R}^N$ and $\{v_k\left(A\right)\}_{k=1}^{\mathrm{rank}(A)}$ is a family of orthonormal vectors of $\mathbb{R}^T$. The scalar product in the space of $N \times T$ matrices is $\langle A, B\rangle = \mathrm{tr}(A^\top B)$. The nuclear norm is $|A|_* = \sum_{k=1}^{\mathrm{rank}(A)} \sigma_k(A)$, and the operator norm is $|A|_{\mathrm{op}} = \sigma_1(A) = \max_{h\in\mathbb{R}^T \text{ s.t. } |h|_2=1} |Ah|_2$. For two integers, N and T,

$N \vee T$ is the maximum of N and T , $N \wedge T$ is the minimum of N and T and $\lfloor N \rfloor$ is the integer part of N. For $N \in \mathbb{N}$, I_N is the identity matrix of size N.

We consider sequences of data generating processes indexed by N. T is a function of N that goes to infinity with N. This paper studies an asymptotic where N goes to infinity. For a probabilistic event $\mathcal{A}$, its complement is denoted $\mathcal{A}^c$ and we write that $\mathcal{A}$ happens with probability approaching 1 or w.p.a 1 if $\mathbb{P}(\mathcal{A}) \to 1$.

2 The estimator

The model can be rewritten in matrix form as $Y = \Pi_0 + E_0$, $X_k = \Pi_k + E_k$ for $k \in \{1, \dots, K\}$, where $\Pi_{kit} = \sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_{kj}$ for $i \in \{1, \dots, N\}$ and $t \in \{1, \dots, T\}$, $\Pi_0 = \sum_{k=1}^{K} \beta_k \Pi_k + \Gamma$, $\Gamma_{it} = \sum_{r=1}^{r_N} \lambda_{ir} f_{tr} \delta_r$ and $E_0 = \sum_{k=1}^{K} \beta_k E_k + E$. Notice that E_0 and E are different. E is the remainder term in (1.1), while E_0 is the remainder term in the expression of Y as the sum of a term with a statistical factor structure and a remainder. Remark also that we do not assume that the error terms $E, E_0, \dots, E_K$ have mean zero, hence they can be the sum of an error term with mean zero and a small remainder as in **?**.

Let $\Pi_u = (\Pi_0, \dots, \Pi_K)$, $\Pi_v = ((\Pi_0)^\top, \dots, (\Pi_K)^\top)$. For $z = u, v$, we denote by P_z the projector on the vector space spanned by the columns of Π_z and M_z the projector on the orthogonal of the vector space spanned by the columns of Π_z. Let r_z be the rank of Π_z. Note that $r_z \leq r_N$, by definition.

The proposed estimator is as follows. In a first step, one estimates M_u and M_v by estimators $\widehat{M_u}$ and $\widehat{M_v}$. From there, the estimator of β is

$$\widehat{\beta} \in \underset{b \in \mathbb{R}^K}{\operatorname{argmin}} \left| \widehat{E}_0 - \sum_{k=1}^{K} b_k \widehat{E}_k \right|_2^2, \tag{2.1}$$

where $\widehat{E}_0 = \widehat{M_u} Y \widehat{M_v}$ and $\widehat{E}_k = \widehat{M_u} X_k \widehat{M_v}$ for $k \in \{1, \dots, K\}$.

As argued in the introduction, the estimator (2.1) can be seen as the regression of the outcome on the regressors and estimated factor loadings and factors as shown in the following lemma. Let us introduce $\widehat{r}_u = \operatorname{rank}\left(I_N - \widehat{M_u}\right)$, $\widehat{r}_v = \operatorname{rank}\left(I_T - \widehat{M_v}\right)$ and $X_{it} = (X_{1it}, \dots, X_{Kit})^\top$.

Lemma 2.1 *Let $\{\widehat{\lambda}_i\}_{i=1}^N$ (resp. $\{\widehat{f}_t\}_{t=1}^T$) be a family of vectors in $\mathbb{R}^{\widehat{r}_u}$ (resp. $\mathbb{R}^{\widehat{r}_v}$) such that $\{(\widehat{\lambda}_{1j} \dots, \widehat{\lambda}_{Nj})^\top\}_{j=1}^{\widehat{r}_u}$ (resp. $\{(\widehat{f}_{1j} \dots, \widehat{f}_{Tj})^\top\}_{j=1}^{\widehat{r}_v}$) is a generating family of the orthogonal of* *the null space of $\widehat{M}_u$ (resp. $\widehat{M}_v$). Then, it holds that*

$$\widehat{\beta} \in \underset{b\in\mathbb{R}^K}{\operatorname{argmin}} \min_{\substack{\phi_1,\dots,\phi_T \in \mathbb{R}^{\widehat{r}_u},\\ l_1,\dots,l_N \in \mathbb{R}^{\widehat{r}_v}}} \sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - X_{it}^\top b - \widehat{\lambda}_i^\top \phi_t - l_i^\top \widehat{f}_t\right)^2 .$$

3 Sufficient assumptions for asymptotic normality

In this section, we present sufficient conditions for asymptotic normality of $\widehat{\beta}$ and consistent estimation of its asymptotic variance. The first assumption concerns the asymptotic behaviour of the error matrices. For a $N \times T$ matrix A, define $\widetilde{A} = M_u A M_v$.

Assumption 3.1 *The following holds:*

(i) *There exists a $K \times K$ positive definite matrix Σ such that, for $k,l \in \{1,\dots,K\}$, $\left\langle \widetilde{E}_k, \widetilde{E}_l \right\rangle /(NT) \xrightarrow{\mathbb{P}} \Sigma_{kl}$;*

(ii) *There exists $\sigma > 0$ such that $\left|\widetilde{E}\right|_2^2/(NT) \xrightarrow{\mathbb{P}} \sigma^2$ and $\left(\left\langle \widetilde{E}_k, \widetilde{E} \right\rangle\right)_{k=1}^K /\sqrt{NT} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\Sigma\right)$.*

This assumption is similar to Assumption 9 (v) and (vi) in **?**. The next lemma provides sufficient conditions for Assumption 3.1.

Lemma 3.1 *Assume that*

(i) $\mathbb{E}\left[|P_u E|_2 + |E P_v|_2\right] + \sum_{k=1}^K \mathbb{E}\left[|P_u E_k|_2 + |E_k P_v|_2\right] = o_P(\sqrt{NT})$;

(ii) *There exists a positive definite matrix Σ such that, for $k,l \in \{1,\dots,K\}$, $\langle E_k, E_l\rangle/(NT) \xrightarrow{\mathbb{P}} \Sigma_{kl}$;*

(iii) *For $k \in \{1,\dots,K\}$, $\langle E_k, P_u E\rangle / |P_u E|_2 = O_P(1)$ and $\langle E_k, M_u E P_v\rangle / |M_u E P_v|_2 = O_P(1)$;*

(iv) *There exists $\sigma > 0$ such that $|E|_2^2/(NT) \xrightarrow{\mathbb{P}} \sigma^2$ and $(\langle E_k, E\rangle)_{k=1}^K /\sqrt{NT} \xrightarrow{d} \mathcal{N}(0, \sigma^2\Sigma)$.*

Then, conditions (i) and (ii) in Assumption 3.1 hold.

The next corollary gives an example of data generating process under which Assumption 3.1 holds.

Corollary 3.1 *Let us assume that $E, E_1, \dots, E_k$ are independent, $r_u + r_v = o_P\left(\sqrt{N \wedge T}\right)$ and there exists $\sigma, \sigma_1, \dots, \sigma_k > 0$ such that $\{E_{it}\}_{it}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ and $\{E_{kit}\}_{it}$ are i.i.d. $\mathcal{N}(0, \sigma_k^2)$. If also $(E, E_1, \dots, E_k)$ is independent of (Π_u, Π_v), then Assumption 3.1 holds.*

The last set of conditions concerns the performance of the estimators of the projectors $\widehat{M}_u$ and $\widehat{M}_v$. Let $\{u_N\}_N$ and $\{v_N\}_N$ be real-valued sequences such that $\left|\widehat{M}_u - M_u\right|_2 = O_P(u_N)$ and $\left|\widehat{M}_v - M_v\right|_2 = O_P(v_N)$. Let also $\{h_N\}_N$ and $\{\rho_N\}_N$ be real-valued sequences such that $\max\limits_{k\in\{0,\dots,K\}} |\Pi_k + E_k|_2 = O_P(h_N)$ and $\max\limits_{k\in\{0,\dots,K\}} |E_k|_{\text{op}} = O_P(\rho_N)$. The estimators satisfy the following assumption.

Assumption 3.2 *The following holds:*

(i) $\widehat{M}_u$ *and* $\widehat{M}_v$ *are symmetric almost surely;*

(ii) $\mathbb{P}(\widehat{r}_u = r_u) \to 1$ *and* $\mathbb{P}(\widehat{r}_v = r_v) \to 1$*;*

(iii) $u_N \vee v_N = o(1)$ *and* $h_N^2 = O(NT)$*;*

(iv) $\sqrt{2r_N}(u_N \vee v_N)\rho_N^2 = o(NT)$*;*

(v) $u_N v_N h_N^2 = o(NT)$*;*

(vi) $\sqrt{2r_N}(u_N \vee v_N)\rho_N = o\left(\sqrt{NT}\right)$*;*

(vii) $u_N v_N \rho_N h_N = o\left(\sqrt{NT}\right)$*;*

This assumption plays a similar role as conditions (i) to (iv) in Assumption 9 in **?**. It is difficult to understand the strength of Assumption 3.2 without examples of u_N and v_N for specific first-step estimators. Hence, we discuss it in Section 4, where we derive the properties of $\widehat{M}_u$ and $\widehat{M}_v$ when they are estimated by a method relying on PCA. The next theorem constitutes the main result of this paper.

Theorem 3.1 (Asymptotic Normality) *Under assumptions 3.1 and 3.2, we have*

$$\sqrt{NT}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\Sigma^{-1}\right).$$

Also, for $k, l \in \{1, \dots, K\}$, $\widehat{\Sigma}_{kl} = \left\langle \widehat{E}_k, \widehat{E}_l \right\rangle / (NT) \xrightarrow{\mathbb{P}} \Sigma_{kl}$ *and* $\widehat{\sigma}^2 = \left|\widehat{E}_0 - \sum_{k=1}^K \widehat{\beta}_k \widehat{E}_k\right|_2^2 / (NT) \xrightarrow{\mathbb{P}} \sigma^2$.

4 Estimation of the projectors using principal components analysis

4.1 Strength of the factors

In this section, we discuss the estimation of the projectors using a method based on PCA. We make assumptions regarding the asymptotic behaviour of $\sigma_j(\Pi_z)$ for $z = u, v$. The purpose of this subsection is to show that there exists data generating processes (henceforth DGP) that generate various asymptotic behaviours of the singular values of Π_z. When $\sigma_j(\Pi_z)/\sqrt{NT}$ has a finite deterministic limit in probability, then we say that the j^{th} factor is strong. The following lemma shows that there exists a wide variety of DGP under which such a strong factor assumption holds. Let $\Lambda = (\lambda_1, \dots, \lambda_N)$, $F = (f_1, \dots, f_N)$, and $\Delta = (\delta_0, \dots, \delta_K)$, where $\delta_0 = \delta + \sum_{k=1}^{K} \beta_k \delta_k$.

Lemma 4.1 *Assume that r_N is fixed and*

(i) There exists a $r_N \times r_N$ positive definite matrix Σ_Λ such that $\Lambda\Lambda^\top/N \xrightarrow{\mathbb{P}} \Sigma_\Lambda$;

(ii) There exists a $r_N \times r_N$ positive definite matrix Σ_F such that $FF^\top/T \xrightarrow{\mathbb{P}} \Sigma_F$;

(iii) $\Delta\Delta^\top$ does not depend on N.

Then, for $z = u, v$, the ratio of the singular values of Π_z and $\sqrt{NT}$ has a finite deterministic limit in probability.

If instead $\sigma_j(\Pi_z)/\alpha_{jN}$ has a finite deterministic limit for $\alpha_{jN} = o\left(\sqrt{NT}\right)$, then the j^{th} factor is not strong. For a detailed discussion of the concept of non-strong factors, see **?**. The following lemma shows how to generate non-strong factors and a growing number of factors in the case where F, Λ and Δ are nonrandom.

Lemma 4.2 *Let $\{\alpha_{jN}\}_N$ for $j \in \mathbb{N}$ be real-valued sequences with positive values. Maintain*

(i) Λ is nonrandom and $(\Lambda\Lambda^\top)_{jj} = I_{r_N}$;

(ii) F is nonrandom and $FF^\top$ is equal to the $r_N \times r_N$ diagonal matrix with coefficients $\alpha_{1r_N}^2, \dots, \alpha_{jN}^2$;

(iii) $\Delta\Delta^\top$ is a diagonal matrix such that $\alpha_{1N}^2 \left(\Delta\Delta^\top\right)_{11} \geq \dots \geq \alpha_{r_N N}^2 \left(\Delta\Delta^\top\right)_{r_N r_N}$.

Then, for $z = u, v$ and $r \in \mathbb{N}$, $\sigma_j(\Pi_z) = \alpha_{jN}\sqrt{\left(\Delta\Delta^\top\right)_{jj}}$.

Notice that the last two lemmas give sufficient conditions for Assumption 8 in **?**.

4.2 Convergence results

The econometrician can use different methods to estimate the projectors M_u and M_v. The approach in **?** relies on a nuclear-norm penalised estimator followed by hard-thresholding of the singular values. It has the advantage of being data-driven, in the sense that it does not use any knowledge of the number of factors or the variance of the errors. Another interesting and computationally advantageous procedure is the double IV estimator of **?**. In this paper, we focus the theoretical presentation on yet another method, based on the PCA. r_u and r_v are estimated via the eigenvalue ratio estimator from **?**. For $z = u, v$, let us define

$$\widehat{r}_z \in \underset{j\in\left\{1,\dots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}}{\operatorname{argmax}} \frac{\sigma_j\left(Y_z\right)}{\sigma_{j+1}\left(Y_z\right)}, \tag{4.1}$$

where $Y_u = (Y, X_1, \dots, X_K)$ and $Y_v = (Y^\top, X_1^\top, \dots, X_K^\top)$. It may be that there exists $r \in \left\{1,\dots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}$, $\sigma_{j+1}\left(Y_z\right) = 0$. To ensure that the estimators are defined, throughout this section, we use the convention that the division of a positive number by 0 is equal to ∞. The estimator in **?** is of the form $\widehat{r}_z \in \operatorname{argmax}_{j\in\{1,\dots,\lfloor d^*(N\wedge T)\rfloor\}} \sigma_j\left(Y_z\right)/\sigma_{j+1}\left(Y_z\right)$, where $d^* \in (0,1]$. Therefore, the estimators in (4.1) correspond to the one in **?** for a particular choice of d^*. Our theoretical analysis is different from the one of **?** because it allows for non-strong factors and a growing number of factors. Contrarily to the estimators in **?**, the advantage of the eigenvalue ratio estimator is that it does not require to choose a penalty level. To ensure consistency of the eigenvalue ratio estimator, we make the following assumption. Let $E_u = (E_0, \dots, E_K)$ and $E_v = (E_0^\top, \dots, E_K^\top)$.

Assumption 4.1 (Eigenvalue Ratio) *For $z = u, v$, it holds that $r_z \leq \sqrt{N\wedge T}$ almost surely, $|E_z|_{\text{op}} = o_P\left(\sigma_{r_z}\left(\Pi_z\right)\right)$ and there exists $C < 1$ such that*

$$\mathbb{P}\left(\left(\max_{j\in\{1,\dots,r_z-1\}} \frac{\sigma_j\left(\Pi_z\right)}{\sigma_{j+1}\left(\Pi_z\right)}\right) \vee \left(\max_{j\in\left\{r_z+1,\dots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}} \frac{|E_z|_{\text{op}}}{\sigma_{r_z+j}\left(E_z\right)}\right) \leq C\frac{\sigma_{r_z}\left(\Pi_z\right)}{\sigma_{2r_z+1}\left(E_z\right)}\right) \to 1.$$

Let us give sufficient conditions fo Assumption 4.1.

Lemma 4.3 *For $z = u, v$, assume that $r_z \leq \sqrt{N\wedge T}$, $|E_z|_{\text{op}} = O_P\left(\sqrt{N\vee T}\right)$, $|E_z|_2^2/(NT)$ has a finite deterministic limit in probability and there exists a sequence $\{z_N\}_N$ such that $\sigma_{r_z}(\Pi_z) = O_P(z_N)$, $\sqrt{N\vee T} = o\left(z_N\right)$ and $\max_{j\in\{1,\dots,r_z-1\}} \sigma_j\left(\Pi_z\right)/\sigma_{j+1}\left(\Pi_z\right) = o_P\left(z_N/\sqrt{N\vee T}\right)$, then Assumption 4.1 holds.*

This Lemma shows that our assumption allows for non-strong factors and a growing number of factors. The condition $\max_{j\in\{1,\dots,r_z-1\}} \sigma_j\left(\Pi_z\right)/\sigma_{j+1}\left(\Pi_z\right) = o_P\left(z_N/\sqrt{N\vee T}\right)$ implies that the singular values of Π_z cannot decrease too quickly with $j \in \{1, \ldots, r_z\}$. The assumption that $|E_z|_{\mathrm{op}} = O_P\left(\sqrt{N \vee T}\right)$ is standard in the panel data literature and holds under flexible cross-sectional and serial correlations. For a detailed discussion, see Appendix A.1 in **?**. Let us now state the main result regarding the eigenvalue ratio estimator.

Lemma 4.4 *Under Assumption 4.1, we have* $\mathbb{P}\left(\widehat{r}_u = r_u, \widehat{r}_v = r_v\right) \to 1$.

Given the estimators $\widehat{r}_u$ and $\widehat{r}_v$, we set $\widehat{M_u} = I_N - \sum_{j=1}^{\widehat{r}_u} u_j\left(Y_u\right) u_j\left(Y_u\right)^{\top}$ and $\widehat{M_v} = I_T - \sum_{j=1}^{\widehat{r}_v} u_j\left(Y_v\right) u_j\left(Y_v\right)^{\top}$. Then, we have the following theorem which states the rates of convergence of the estimators of the projectors.

Theorem 4.1 *For* $z = u, v$, *if* $\mathbb{P}\left(\widehat{r}_z = r_z\right) \to 1$, *we have* $\left|\widehat{M_z} - M_z\right|_2 = O_P\left(\sqrt{r_z}\left|E_z\right|_{\mathrm{op}} / \sigma_{r_z}\left(\Pi_z\right)\right)$.

4.3 Examples

Let us now show how Assumption 3.2 can hold under different assumptions on the singular values of Π_u, Π_v, E_u and E_v. In both examples, we assume that, for $z = u, v$, $|E_z|_{\mathrm{op}} = O_P\left(\sqrt{N \vee T}\right)$ and $|E_z|_2^2/(NT)$ has a deterministic finite limit. The assumption on the errors implies that we can choose $\rho_N = \sqrt{N \vee T}$ because $|E_k|_{\mathrm{op}} \leq |E_u|_{\mathrm{op}}$.

Example 1. In this first example, we assume that r_N is fixed and that the strong factor assumption holds, that is, for $z = u, v$ and $j \in \{1, \ldots, r\}$, $\sigma_j(\Pi_z)/\sqrt{NT}$ has a finite deterministic limit. This implies that we can choose $h_N = \sqrt{NT}$ because $|\Pi_u|_2 = O_P\left(\sqrt{NT}\right)$ and that Assumption 4.1 holds by Lemma 4.3. Theorem 4.1 yields $u_N = v_N = 1/\sqrt{N \wedge T}$. All conditions in Assumption 3.2 except (vii) are satisfied whatever the value of N and T. Condition (vii) holds if $\sqrt{N \vee T}/(N \wedge T) = o(1)$. The latter correponds to the condition for asymptotic normality of the debiased estimator in **?** and is weaker than the conditions for asymptotic normality in **?**.

Example 2. In this case, we assume that $r_u = r_v = r_N$ can grow with the sample size, and that, for $z = u, v$ and $j \in \{1, \ldots, r_N\}$, $\sqrt{r_N}\sigma_j(\Pi_z)/\sqrt{NT}$ has a finite deterministic limit. This is a case with non-strong factors and a growing number of factors. This implies that we can choose $h_N = \sqrt{NT}$ because $|\Pi_u|_2 = O_P\left(\sqrt{NT}\right)$ and that Assumption 4.1 holds by Lemma 4.3. From Theorem 4.1, we obtain $u_N = v_N = r_N/\sqrt{N \wedge T}$. Conditions (i)-(iii) hold for any value of N, T and r_N. For (iv)- (vi) to hold, it is enough that $r_N^{\frac{3}{2}}/(N \wedge T) = o(1)$. Finally, condition (vii) is satisfied if $r_N^2\sqrt{N \vee T}/(N \wedge T) = o(1)$.

5 Simulations

We consider a data generating process with a single regressor and two factors:

$$
\begin{aligned}
Y_{it} &= X_{1it} + \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + E_{it}, \\
X_{1it} &= \frac{1}{2}\lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + E_{1it},
\end{aligned}
$$

where f_{tl}, λ_{il}, E_{1it}, and E_{it} for all indices are mutually independent, $f_{tl} \sim \mathcal{N}(1/2, 1)$, $\lambda_{il} \sim \mathcal{N}(1, 1)$ and $E_{1it}, \ldots, E_{Kit}$ and E_{it} are standard normals. The matrix X_1 has a statistical factor structure with a low-rank component of rank 2. Recall that $\widehat{\beta}^{LS} \in \operatorname{argmin} b \in \mathbb{R} \, |Y - bX_1|_2^2$ is the least-squares estimator of the linear regression of the outcome on the regressors. $\widehat{\beta}^{FA} \in \operatorname{argmin} b \in \mathbb{R} \left| Y\widehat{M_v} - bX_1\widehat{M_v} \right|_2^2$ is the factor augmented regression estimator where $\widehat{M_v}$ is computed as in Section 4. $\widetilde{\beta}^{(1)}$ and $\widetilde{\beta}^{(2)}$ are the two-stage estimators of Section 4.7.1 in **?**. They are computed as in the simulations of that paper, without using within transforms. $\widetilde{\beta}^{(2)}$ uses Bai’s estimator as a second stage while $\widetilde{\beta}_2$ uses the approach of this paper with two projectors and a first-step based on hard-thresholding of a nuclear-norm penalized estimator. Finally, $\widehat{\beta}^{PCA}$ is the estimator (2.1), using the procedure of Section 4 as the first-stage.

Tables 1 and 2 compare the performance of the estimators in terms of mean squared error (henceforth MSE), bias, standard error (henceforth std) and coverage of 95% confidence intervals, for different sample sizes. The coverage is not reported for $\widehat{\beta}_{LS}$ because the latter is not asymptotically normal for the DGP that we consider. We use 7300 Monte-Carlo replications which allows for an accuracy of ±0.005 with 95% for the coverage probabilities of 95% confidence intervals. In this simulation exercise, our estimator exhibits better finite samples properties than the studied alternatives.

Table 1: $N = T = 50$

	$\widehat{\beta}^{LS}$	$\widehat{\beta}^{FA}$	$\widetilde{\beta}^{(1)}$	$\widetilde{\beta}^{(2)}$	$\widehat{\beta}^{PCA}$
MSE	0.884	0.004	0.14	0.13	0.004
bias	0.939	-0.011	0.321	0.275	0.012
std	0.055	0.191	0.023	0.234	0.063
coverage		0.75	0.22	0.37	0.90

Table 2: $N = T = 150$

	$\widehat{\beta}^{LS}$	$\widehat{\beta}^{FA}$	$\widetilde{\beta}^{(1)}$	$\widetilde{\beta}^{(2)}$	$\widehat{\beta}^{PCA}$
MSE	0.887	10^{-4}	$4\ 10^{-5}$	$4\ 10^{-5}$	$4\ 10^{-5}$
bias	0.9414	-0.007	$-5\ 10^{-5}$	$-8\ 10^{-6}$	$-5\ 10^{-5}$
std	0.031	0.007	0.007	0.007	0.007
coverage		0.79	0.95	0.95	0.95

Appendix

Proof of Lemma 2.1.

Let $b \in \mathbb{R}^k$, $Y_i = (Y_{i1}, \dots, Y_{iT})^\top$ and $X_i = (X_{i1}, \dots, X_{iT})^\top$ for $i \in \{1, \dots, N\}$, $Y_t = (Y_{1t}, \dots, Y_{Nt})^\top$ and $X_t = (X_{1t}, \dots, X_{Nt})^\top$ for $t \in \{1, \dots, T\}$ and

$$\varphi(b) = \min_{\substack{\phi_1, \dots, \phi_T \in \mathbb{R}^{\widehat{r}_u}, \\ l_1, \dots, l_N \in \mathbb{R}^{\widehat{r}_v}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - X_{it}^\top b - \widehat{\lambda}_i^\top \phi_t - l_i^\top \widehat{f}_t \right)^2 .$$

By algebra, we have

$$\varphi(b) = \min_{\substack{\phi_1, \dots, \phi_T \in \mathbb{R}^{\widehat{r}_u}, \\ l_1, \dots, l_N \in \mathbb{R}^{\widehat{r}_v}}} \sum_{i=1}^{N} \left| Y_i - X_i b - (\phi_1, \dots, \phi_T)^\top \widehat{\lambda}_i - \left(\widehat{f}_1, \dots, \widehat{f}_T \right)^\top l_i \right|_2^2 .$$

Then, by definition of $\widehat{M}_v$, it holds

$$\varphi(b) = \min_{\phi_1, \dots, \phi_T \in \mathbb{R}^{\widehat{r}_u}} \sum_{i=1}^{N} \left| \widehat{M}_v \left(Y_i - X_i b - (\phi_1, \dots, \phi_T)^\top \widehat{\lambda}_i \right) \right|_2^2 .$$

Because $\widehat{M}_v$ is symmetric, this implies

$$\varphi(b) = \min_{\phi_1, \dots, \phi_T \in \mathbb{R}^{\widehat{r}_u}} \left| \left(Y - \sum_{k=1}^{K} b_k X_k - \left(\widehat{\lambda}_1, \dots, \widehat{\lambda}_N \right)^\top (\phi_1, \dots, \phi_T) \right) \widehat{M}_v \right|_2^2 . \tag{5.1}$$

Next, by definition of $\widehat{M}_u$, we obtain

$$\varphi(b) \geq \left| \widehat{M}_u \left(Y - \sum_{k=1}^{K} b_k X_k \right) \widehat{M}_v \right|_2^2 .$$

Hence, because the value of

$$\min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{\widehat{r}_u}} \left| \left(Y - \sum_{k=1}^{K} b_k X_k - \left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N \right)^\top (\phi_1, \ldots, \phi_T) \right) \widehat{M}_v \right|_2^2$$

is $\left| \widehat{M}_u \left(Y - \sum_{k=1}^{K} b_k X_k \right) \widehat{M}_v \right|_2^2$ when $\left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_N \right)^\top \phi_t = \left(I_N - \widehat{M}_u \right) (Y_t - X_t b)$, we get

$$\varphi(b) = \left| \widehat{M}_u \left(Y - \sum_{k=1}^{K} b_k X_k \right) \widehat{M}_v \right|_2^2 .$$

Proof of Lemma 3.1.

Proof that Assumption 3.1 (i) holds. For $k \in \{1, \ldots, K\}$, we have

$$E_k - M_u E_k M_v = P_u E_k + M_u E_k P_v.$$

By Markov's inequality and the fact that M_u is a projector, we have

$$|P_u E_k|_2 + |M_u E_k P_v|_2 \leq |P_u E_k|_2 + |E_k P_v|_2 = O_P \left(\mathbb{E} \left[|P_u E_k|_2 + |E_k P_v|_2 \right] \right).$$

By condition (i) in Lemma 3.1, this yields $|P_u E_k|_2 + |M_u E_k P_v|_2 = o_P \left(\sqrt{NT} \right)$. This implies $|M_u E_k M_v - E_k|_2 = o_P \left(\sqrt{NT} \right)$. By the Cauchy-Schwarz inequality, we obtain

$$\langle M_u E_k M_v, E_l \rangle - \langle E_k, E_l \rangle = \langle M_u E_k M_v - E_k, E_l \rangle = o_P (NT)$$

because $|E_k|_2 = O_P \left(\sqrt{NT} \right)$. We get

$$\frac{1}{NT} \left\langle \widetilde{E}_k, \widetilde{E}_l \right\rangle = \frac{1}{NT} \langle M_u E_k M_v, E_l \rangle = \frac{1}{NT} \langle E_k, E_l \rangle + o_P(1) \xrightarrow{\mathbb{P}} \Sigma_{kl},$$

by condition (ii) in Lemma 3.1.

Proof that Assumption 3.1 (ii) holds. The proof of $\left| \widetilde{E} \right|_2^2 / (NT) \xrightarrow{\mathbb{P}} \sigma^2$ is similar to the proof that Assumption 3.1 (i) holds. By conditions (i) and (iii) in Lemma 3.1, we have

$$\langle P_u E_k, E\rangle = O_P(1)\left|P_u E_k\right|_2 = O_P(1)o_P\left(\sqrt{NT}\right) = o_P\left(\sqrt{NT}\right)$$

and similarly $\langle M_u E_k P_v, E\rangle = o_P\left(\sqrt{NT}\right)$. Next, this yields

$$\langle E_k, E\rangle - \langle M_u E_k M_v, E\rangle \ = \langle P_u E_k, E\rangle + \langle M_u E_k P_v, E\rangle = o_P\left(\sqrt{NT}\right).$$

We obtain that

$$\frac{1}{\sqrt{NT}}\left(\left\langle \widetilde{E}_k, \widetilde{E}\right\rangle\right)_{k=1}^K = \frac{1}{\sqrt{NT}}\left(\langle M_u E_k M_v, E\rangle\right)_{k=1}^K = \frac{1}{\sqrt{NT}}\left(\langle E_k, E\rangle\right)_{k=1}^K + o_P(1) \xrightarrow{d} \mathcal{N}\left(0,\sigma^2\Sigma\right),$$

by condition (iv) in Lemma 3.1.

Proof of Corollary 3.1.

Let us prove that the assumptions of Proposition 3.1 are satisfied. (ii) and (iv) in Proposition 3.1 are direct consequences of the weak law of large numbers and the central limit theorem. Concerning (i) in Proposition 3.1, for $k \in \{1,\dots,K\}$, by Lemma A.3 in **?** and the fact that $E, E_1, \dots, E_K$ are independent of Π_u and Π_v, we have

$$\mathbb{E}\left[\left|P_u E_k\right|_2^2\right] = \sum_{t=1}^T \mathbb{E}\left[\sum_{i=1}^N (P_u E_k)_{it}^2\right] = \sum_{t=1}^T \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N (P_u E_k)_{it}^2 \middle| P_u\right]\right] = Tr_u \sigma_k^2 = o\left(\sqrt{NT}\right)$$

Similarly, one can show that $\mathbb{E}\left[\left|E_k P_v\right|_2^2\right] = o\left(\sqrt{NT}\right)$. In the same manner, we obtain that $\mathbb{E}\left[\left|P_u E\right|_2 + \left|E P_v\right|_2\right] = o(\sqrt{NT})$. To prove (iii), just notice that conditionally on $P_u E$, we have $\langle P_u E, E_k,\rangle / \left|P_u E\right|_2 \sim \mathcal{N}(0,\sigma_k^2)$, hence, for any $M \geq 0$, it holds that

$$\mathbb{P}\left(\frac{\left|\langle P_u E_k, E\rangle\right|}{\left|P_u E\right|_2 \sigma_k} > M \middle| P_u E\right) \leq 2(1-\Phi^{-1}(M)),$$

where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ distribution. This implies

$$\mathbb{P}\left(\frac{\left|\langle P_u E_k, E\rangle\right|}{\left|P_u E\right|_2 \sigma_k} > M\right) \leq 2(1-\Phi^{-1}(M)).$$

Therefore, we obtain that $\langle P_u E_k, E\rangle / \left|P_u E\right|_2 = O_P(1)$. The proof that $\langle E_k, M_u E P_v\rangle / \left|M_u E P_v\right|_2 = O_P(1)$ is the same.

Proof of Theorem 3.1.

Proof of asymptotic normality.

Because $\widehat{M}_u$ and $\widehat{M}_v$ are symmetric, a solution to (2.1) satisfies, for $l = 1, \ldots, K$,

$$\left\langle \widehat{M}_u X_l \widehat{M}_v, Y - \sum_{k=1}^{K} \widehat{\beta}_k X_k \right\rangle = 0,$$

hence

$$\begin{aligned}
&\left\langle M_u X_l M_v, +E + \sum_{k=1}^{K} \left(\beta_k - \widehat{\beta}_k\right) X_k \right\rangle \\
&= \left\langle \left(M_u - \widehat{M}_u\right) X_l M_v, E + \sum_{k=1}^{K} \left(\beta_k - \widehat{\beta}_k\right) X_k \right\rangle \\
&\quad + \left\langle M_u X_l \left(M_v - \widehat{M}_v\right), E + \sum_{k=1}^{K} \left(\beta_k - \widehat{\beta}_k\right) X_k \right\rangle \\
&\quad - \left\langle \left(M_u - \widehat{M}_u\right) X_l \left(M_v - \widehat{M}_v\right), \Gamma + E + \sum_{k=1}^{K} \left(\beta_k - \widehat{\beta}_k\right) X_k \right\rangle,
\end{aligned}$$

so

$$\begin{aligned}
&\sum_{k=1}^{K} \left(\beta_k - \widehat{\beta}_k\right) \Bigg(\langle M_u X_l M_v, X_k \rangle - \left\langle \left(M_u - \widehat{M}_u\right) X_l M_v, X_k \right\rangle \\
&\quad - \left\langle M_u X_l \left(M_v - \widehat{M}_v\right), X_k \right\rangle \\
&\quad + \left\langle \left(M_u - \widehat{M}_u\right) X_l \left(M_v - \widehat{M}_v\right), X_k \right\rangle \Bigg) + \left\langle \left(M_u - \widehat{M}_u\right) X_l M_v, +E \right\rangle \\
&\quad + \left\langle M_u X_l \left(M_v - \widehat{M}_v\right), +E \right\rangle \\
&\quad - \left\langle \left(M_u - \widehat{M}_u\right) X_l \left(M_v - \widehat{M}_v\right), \Gamma + E \right\rangle. \qquad (5.2)
\end{aligned}$$

Let us show that $\langle M_u X_l M_v, X_k \rangle$, which by Assumption 3.1 (i) diverges like NT, is the high-order term multiplying $\left(\beta_k - \widehat{\beta}_k\right)$ in (5.2). This also yields the consistency of the estimator of the covariance matrix. For a matrix M and $r \in \mathbb{N}$, let us define $|M|_{2,r}^2 = \sum_{k=1}^{r} \sigma_k(M)^2$. By symmetry of the projectors, Theorem C.5 in **?**, and Assumption 3.2 (ii)

$\left(\text{which implies rank}\left(M_u - \widehat{M_u}\right) \leq 2r_N \text{ w.p.a. } 1\right)$, we have

$$\begin{aligned}
&\left|\left\langle \left(M_u - \widehat{M_u}\right) X_l M_v, X_k \right\rangle\right| \\
&\leq \left|M_u - \widehat{M_u}\right|_2 \left|X_l M_v X_k^\top\right|_{2,2r_N} \\
&\leq \left(\sqrt{2r_N} + o_P(1)\right) \left|M_u - \widehat{M_u}\right|_2 \left|X_l M_v\right|_{\mathrm{op}} \left|X_k M_v\right|_{\mathrm{op}} \\
&= O_P\left(\sqrt{2r_N} u_N \rho_N^2\right) = o_P(NT) \quad \text{(by Assumption 3.2 (iv)).}
\end{aligned}$$

We bound similarly $\left|\left\langle M_u X_l \left(M_v - \widehat{M_v}\right), X_k \right\rangle\right|$, and, for the fourth term, use that

$$\begin{aligned}
&\left|\left\langle \left(M_u - \widehat{M_u}\right) X_l \left(M_v - \widehat{M_v}\right), X_k \right\rangle\right| \\
&\leq \left|\left(M_u - \widehat{M_u}\right) X_l \left(M_v - \widehat{M_v}\right)\right|_2 \left|X_k\right|_2 \\
&= O_P\left(u_N v_N h_N^2\right) = o_P(NT) \quad \text{(by Assumption 3.2 (v)).}
\end{aligned} \tag{5.3}$$

Let us consider now the quantities on the right-hand side in (5.2). Notice that because $E = E_0 - \sum_{k=1}^K \beta_k E_k$, it holds that $|E|_{\mathrm{op}} = O_P(\rho_N)$. Proceeding like above, we have

$$\begin{aligned}
&\left|\left\langle \left(M_u - \widehat{M_u}\right) X_l M_v, E \right\rangle\right| \\
&\leq \left|M_u - \widehat{M_u}\right|_2 \left|X_l M_v E^\top\right|_{2,2r_N} \\
&\leq \left(\sqrt{2r_N} + o_P(1)\right) u_N \left|E_l\right|_{\mathrm{op}} \left|E\right|_{\mathrm{op}} \\
&= O_P\left(\sqrt{2r_N} u_N \rho_N^2\right) = o_P(\sqrt{NT}) \quad \text{(by Assumption 3.2 (vi)).}
\end{aligned}$$

and treat similarly $\left\langle M_u X_l \left(M_v - \widehat{M_v}\right), E \right\rangle$. With the same arguments as in (5.3), the absolute value of the last term of (5.2) is smaller than $u_N v_N |X_l|_2 |\Gamma + E|_2$, which is an $O_P(u_N v_N \rho_N h_N) = o_P\left(\sqrt{NT}\right)$ because $\Gamma + E = Y - \sum_{k=1}^K \beta_k X_k$.

Let us now look at the first terms on the left-hand side and on the right-hand side of (5.2). By Assumption 3.2 (vi), for all $k, l \in \{1, \ldots, K\}$, we have $\langle M_u X_l M_v, X_k \rangle = \langle M_u E_l M_v, E_k \rangle + o_P(NT)$. Hence because of Assumption 3.1 (i), $\langle M_u X_l M_v, X_k \rangle$ are the high-order terms on the left-hand side of (5.2). Similarly, by Assumption 3.1 (ii), the high-order terms on the right-hand side of (5.2) are $\langle M_u E_l M_v, E \rangle$. As a result, $\widehat{\beta}$ is asymptotically equivalent to the ideal estimator $\overline{\beta}$

$$\overline{\beta} \in \underset{\beta \in \mathbb{R}^K}{\operatorname{argmin}} \left|M_u \left(Y - \sum_{k=1}^K \beta_k X_k\right) M_v\right|_2^2. \tag{5.4}$$

Hence, we obtain by usual arguments that $\sqrt{NT}(\widehat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, \sigma\Sigma^{-1})$.

Proof of the consistency of $\widehat{\sigma}$. We use

$$\begin{aligned}
NT\widehat{\sigma}^2 &= \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, \widehat{M}_u \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) \widehat{M}_v \right\rangle \\
&= \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, \left(\widehat{M}_u - M_u\right) \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) \left(\widehat{M}_v - M_v\right) \right\rangle \\
&\quad + \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, \left(\widehat{M}_u - M_u\right) \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) M_v \right\rangle \\
&\quad + \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, M_u \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) \left(\widehat{M}_v - M_v\right) \right\rangle \\
&\quad + \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, M_u \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) M_v \right\rangle .
\end{aligned}$$

Now, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left| \left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, \left(\widehat{M}_u - M_u\right) \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) \left(\widehat{M}_v - M_v\right) \right\rangle \right| \\
&\leq \left| Y - \sum_{k=1}^K \widehat{\beta}_k X_k \right|_2 \left| \left(\widehat{M}_u - M_u\right) \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) \left(\widehat{M}_v - M_v\right) \right|_2 \\
&\leq \left| Y - \sum_{k=1}^K \widehat{\beta}_k X_k \right|_2^2 \left| \widehat{M}_u - M_u \right|_2 \left| \widehat{M}_v - M_v \right|_2 \\
&= O_P(h_N^2 u_N v_N) \quad (\text{by the fact that } \widehat{\beta} - \beta = o_P(1)) \\
&= o_P(NT) \quad (\text{by Assumption 3.2 (iii)}).
\end{aligned}$$

Similarly, we can show that

$$\left\langle Y - \sum_{k=1}^K \widehat{\beta}_k X_k, \left(\widehat{M}_u - M_u\right) \left(Y - \sum_{k=1}^K \widehat{\beta}_k X_k\right) M_v \right\rangle = o_P(NT)$$

and

$$\left\langle Y-\sum_{k=1}^{K}\widehat{\beta}_k X_k, M_u\left(Y-\sum_{k=1}^{K}\widehat{\beta}_k X_k\right)\left(\widehat{M_v}-M_v\right)\right\rangle = o_P(NT).$$

Hence, we have

$$\begin{aligned}
&NT\widehat{\sigma}^2\\
&=\left\langle Y-\sum_{k=1}^{K}\widehat{\beta}_k X_k, M_u\left(Y-\sum_{k=1}^{K}\widehat{\beta}_k X_k\right)M_v\right\rangle + o_P(NT)\\
&=\left\langle Y-\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k-\sum_{k=1}^{K}\beta_k X_k, M_u\left(Y-\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k-\sum_{k=1}^{K+1}\beta_k X_k\right)M_v\right\rangle + o_P(NT)\\
&=\left\langle \sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k, M_u\left(\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k\right)M_v\right\rangle + \left\langle E_0, M_u\left(\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k\right)M_v\right\rangle\\
&\quad+\left\langle \sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k, M_u E_K M_v\right\rangle + \left|\widetilde{E}\right|_2^2 + o_P(NT).
\end{aligned}$$

Now, by the Cauchy-Schwarz inequality, Assumption 3.2 and the fact that $\widehat{\beta}-\beta = o_P(1)$, one can show that

$$\begin{aligned}
\left\langle \sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k, M_u\left(\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k\right)M_v\right\rangle &= o_P(NT)\\
\left\langle E, M_u\left(\sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k\right)M_v\right\rangle &= o_P(NT);\\
\left\langle \sum_{k=1}^{K}\left(\widehat{\beta}_k-\beta_k\right)X_k, M_u E M_v\right\rangle &= o_P(NT).
\end{aligned}$$

We conclude the proof using Assumption 3.1.

Proof of Lemma 4.1.

Let $\Lambda = (\lambda_1, \dots \lambda_N)$. For $t \in \{1,\dots,T\}$ and $k \in \{0,\dots,K\}$, we use the notation $\psi_{tk} = \left(f_{t1}\delta_{k1},\dots,f_{tr_N}\delta_{kr_N}\right)^\top$. We also introduce $\Psi = (\psi_{10},\dots,\psi_{T0},\dots,\psi_{1K},\dots,\psi_{TK})$. It holds that, for $j,j' \in \{1,\dots,r_N\}$, $\left(\Psi\Psi^\top\right)_{jj'}/T = (\Delta\Delta^\top)_{jj'}\left(FF^\top\right)_{jj'}/T$. Therefore, $\Lambda\Lambda^\top\Psi\Psi^\top/(NT)$ converges in probability to $\Sigma_\Lambda\Sigma_{\Delta F}$, where, for $j,j' \in \{1,\dots,r_N\}$, $(\Sigma_{\Delta F})_{jj'} = (\Delta\Delta^\top)_{jj'}(\Sigma_F)_{jj'}$. Next, let $U = (u_1(\Pi_u)),\dots,u_{r_N}(\Pi_u))$, $V = (v_1(\Pi_u),\dots,v_{r_N}(\Pi_u))$ and D be the $r_N \times r_N$ diagonal matrix for which $D_{jj} = \sigma_j(\Pi_u)$. We have $UDV^\top = \Lambda^\top\Psi$, which implies $UD^2U^\top =$

$\Lambda^\top \Psi \Psi^\top \Lambda$. This yields $\Lambda U D^2 = \Lambda\Lambda^\top \Psi\Psi^\top \Lambda U$. On the event $\mathcal{E} = \{\mathrm{rank}(\Lambda U) = r_N\}$, we obtain $\Lambda\Lambda^\top \Psi\Psi^\top = \Lambda U D^2 (\Lambda U)^{-1}$. Therefore, the diagonal elements of $D^2/(NT)$ are the eigenvalues of $\Lambda\Lambda^\top\Psi\Psi^\top/(NT)$ on $\mathcal{E}$. Because $\Lambda\Lambda^\top/N$ converges in probability to a positive definite matrix, the set of full rank matrices is an open set and the determinant is a continuous mapping, we have $\mathbb{P}(\mathrm{rank}(\Lambda) = r_N) \to 1$, which implies $\mathbb{P}(\mathcal{E}) \to 1$. For $j \in \{1, \dots, r_N\}$ and $\xi > 0$, we get

$$\begin{aligned}
&\mathbb{P}\left(\left|\sigma_j\left(\frac{D^2}{NT}\right) - \sigma_j\left(\Sigma_\Lambda \Sigma_{\Delta F}\right)\right| \le \xi\right)\\
&\ge \mathbb{P}\left(\left\{\left|\sigma_j\left(\frac{D^2}{NT}\right) - \sigma_j\left(\Sigma_\Lambda \Sigma_{\Delta F}\right)\right| \le \xi\right\} \cap \mathcal{E}\right)\\
&= \mathbb{P}\left(\left\{\left|\sigma_j\left(\Lambda\Lambda^\top\Psi\Psi^\top/(NT)\right) - \sigma_j\left(\Sigma_\Lambda \Sigma_{\Delta F}\right)\right| \le \xi\right\} \cap \mathcal{E}\right) \to 1,
\end{aligned}$$

where the last statement holds because $\Lambda\Lambda^\top\Psi\Psi^\top/(NT) \xrightarrow{\mathbb{P}} \Sigma_\Lambda\Sigma_{\Delta F}$, $A \in \mathbb{R}^{r_N \times r_N} \mapsto (\sigma_1(A), \dots, \sigma_{r_N}(A))$. is a continuous mapping and $\mathbb{P}(\mathcal{E}) \to 1$.

Proof of Lemma 4.2.

We only prove the result for Π_u, the proof for Π_v being similar. We use the same notations as in the proof of Lemma 4.1. We have, for $j, j' \in \{1, \dots, r_N\}$, $\left(\Psi\Psi^\top\right)_{jj'} = (FF^\top)_{jj'}(\Delta\Delta^\top)_{jj'} = 0$ if $j \neq j'$ and $\left(\Psi\Psi^\top\right)_{jj} = (FF^\top)_{jj}(\Delta\Delta^\top)_{jj} = \alpha_{jN}^2(\Delta\Delta^\top)_{jj}$ if $j = j'$. Therefore, $\Lambda\Lambda^\top\Psi\Psi^\top$ is the diagonal matrix with diagonal coefficients $\alpha_{1N}^2(\Delta\Delta^\top)_{11}, \dots, \alpha_{r_N N}^2(\Delta\Delta^\top)_{r_N r_N}$. Because Λ has full rank, $\Lambda\Lambda^\top\Psi\Psi^\top = \Lambda U D^2(\Lambda U)^{-1}$ and, therefore, the diagonal coefficients of D^2 are $\alpha_{1N}^2(\Delta\Delta^\top)_{11}, \dots, \alpha_{r_N N}^2(\Delta\Delta^\top)_{r_N r_N}$.

Results on PCA

Let us consider a $N \times T$ random matrix A. We do not observe A but $\widetilde{A} = A + Z$, where Z is an $N \times T$ random matrix. Let r be the rank of A. $A = \sum_{j=1}^r \sigma_j u_j v_j^\top$ is the singular value decomposition of A, where $\sigma_1 \ge \dots \ge \sigma_r \ge 0$ and $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_r\}$ are orthonormal families of $\mathbb{R}^N$ and $\mathbb{R}^T$, respectively. With similar notations, $\widetilde{A} = \sum_{j=1}^{\widetilde{r}} \widetilde{\sigma}_j \widetilde{u}_j \widetilde{v}_j^\top$ is the singular value decomposition of $\widetilde{A}$ and $\widetilde{r}$ is the rank of $\widetilde{A}$. $Z = \sum_{j=1}^{N \wedge T} \sigma_j(Z) u_j(Z) v_j(Z)^\top$ is a singular value decomposition of Z. $T = T(N)$ is a function of N going to ∞ when $N \to \infty$ and and the asymptotic setting is such that $N \to \infty$. For $s \in \{1, \dots, N \wedge T\}$, we consider the following estimators of A and P, $\widehat{A}_s = \sum_{j=1}^s \widetilde{\sigma}_j \widetilde{u}_j \widetilde{v}_j^\top$ and $\widehat{P}_s = \sum_{j=1}^s \widetilde{u}_j \widetilde{u}_j^\top$. Let also $\widehat{M}_s = I_N - \widehat{P}_s$.

Lemma 5.1 $\left|\widehat{A}_r - A\right|_{\mathrm{op}} \le 2\left|Z\right|_{\mathrm{op}}$.

Proof. We have $\left|\widehat{A}_r - A\right|_{\mathrm{op}} = \left|\widehat{A}_r - \widetilde{A} + \widetilde{A} - A\right|_{\mathrm{op}} \leq \left|\sum_{j=r+1}^{N\wedge T} \widetilde{\sigma}_j \widetilde{u}_j \widetilde{v}_j^\top\right|_{\mathrm{op}} + |Z|_{\mathrm{op}} = \widetilde{\sigma}_{r+1} + |Z|_{\mathrm{op}}$. Now, by Weyl's inequality (Theorem C.6 in **?**), it holds that $\widetilde{\sigma}_{r+1} \leq \left|\widetilde{A} - A\right|_{\mathrm{op}} = |Z|_{\mathrm{op}}$.
□

Lemma 5.2 *We have* $\left|\widehat{P}_r - P\right|_2 \leq 4\sqrt{2r}\frac{|Z|_{\mathrm{op}}}{\sigma_r}$ *almost surely.*

Proof. Following the proof of Proposition 10 in **?**, we obtain $\left|\widehat{P}_r - P\right|_2^2 \leq 2\left|\widehat{M}_r A\right|_2^2/\sigma_r^2$. We conclude using

$$\left|\widehat{M}_r A\right|_2 = \left|\widehat{M}_r\left(\widehat{A}_r - A\right)\right|_2 \leq \left|\widehat{A}_r - A\right|_2 \leq \sqrt{2r}\left|\widehat{A}_r - A\right|_{\mathrm{op}} \leq \sqrt{2r}2\,|Z|_{\mathrm{op}},$$

by Lemma 5.1 and the fact that $\widehat{M}_r$ is a projector. □

Lemma 5.3 *The following holds:*

(*i*) *For* $j \in \{1, \dots, r\}$, $\sigma_j - \left|\widehat{A}_r - A\right|_{\mathrm{op}} \leq \widetilde{\sigma}_j \leq \sigma_j + \left|\widehat{A}_r - A\right|_{\mathrm{op}}$;

(*ii*) *For* $j \in \{r+1, N \wedge T - r\}$, $\sigma_{r+j}(Z) \leq \widetilde{\sigma}_j \leq |Z|_{\mathrm{op}}$.

Proof. (i) follows from the fact that $|\widetilde{\sigma}_j - \sigma_j| \leq \left|\widehat{A}_r - A\right|_{\mathrm{op}}$ by Weyl's inequality. Weyl's inequality also yields $\widetilde{\sigma}_j \leq \left|\widetilde{A} - A\right|_{\mathrm{op}} = |Z|_{\mathrm{op}}$, which implies the right-hand side of (ii). To show the left-hand side of (ii), from (7.3.13) in **?**, we obtain $\sigma_{r+j}(Z) \leq \widetilde{\sigma}_{j-1} + \sigma_{r+1} = \widetilde{\sigma}_j$. □

Lemma 5.4 *Let* Z *be a* $N \times T$ *random matrix and* $r \in \left\{1, \dots, \left\lfloor\sqrt{N \wedge T}\right\rfloor\right\}$. *Assume that* $|Z|_{\mathrm{op}} = O_P\left(\sqrt{N \vee T}\right)$ *and there exists* $v > 0$ *such that* $|Z|_2^2/(NT) \xrightarrow{\mathbb{P}} v^2$. *Then, we have* $\sigma_{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}(Z) > 0$ *w.p.a. 1 and* $\max_{j\in\left\{1,\dots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}} |Z|_{\mathrm{op}}/\sigma_{r+j}(Z) = O_P(1)$.

Proof. We have

$$\frac{|Z|_2^2}{NT} \leq \frac{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}{NT}|Z|_{\mathrm{op}}^2 + \frac{N\wedge T}{NT}\sigma_{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}(Z)^2.$$

Thus, we obtain

$$\frac{|Z|_2^2}{NT} - \frac{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}{NT}|Z|_{\mathrm{op}}^2 \leq \frac{N\wedge T}{NT}\sigma_{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}(Z)^2.$$

Using $|Z|_2^2/(NT) \xrightarrow{\mathbb{P}} v^2$ and $|Z|_{\mathrm{op}} = O_P\left(\sqrt{N \vee T}\right)$, we get

$$\frac{|Z|_2^2}{NT} - \frac{2\left\lfloor\sqrt{N\wedge T}\right\rfloor}{NT}|Z|_{\mathrm{op}}^2 \xrightarrow{\mathbb{P}} v^2$$

and, therefore,

$$\mathbb{P}\left(\frac{\sigma_{2\lfloor\sqrt{N\wedge T}\rfloor}(Z)}{\sqrt{N\vee T}} \geq \frac{v}{2}\right) \to 1.$$

Hence, we have

$$\mathbb{P}\left(\frac{|Z|_{\text{op}}}{\sigma_{2\lfloor\sqrt{N\wedge T}\rfloor}(Z)} \leq \frac{2\,|Z|_{\text{op}}}{\sqrt{N\vee T}v}\right) \to 1.$$

Therefore, we obtain

$$\frac{|Z|_{\text{op}}}{\sigma_{2\lfloor\sqrt{N\wedge T}\rfloor}(Z)} = O_P\left(\frac{|Z|_{\text{op}}}{\sqrt{N\vee T}}\right) = O_P(1).$$

This leads to

$$\max_{j\in\{1,\ldots,\lfloor\sqrt{N\wedge T}\rfloor\}} \frac{|Z|_{\text{op}}}{\sigma_{r+j}(Z)} \leq \frac{|Z|_{\text{op}}}{\sigma_{2\lfloor\sqrt{N\wedge T}\rfloor}(Z)} = O_P(1).$$

□

Proof of Lemma 4.3.

Because $\sqrt{N\vee T} = o(z_N)$, it holds that $|E_z|_{\text{op}} = O_P(\sigma_{r_z}(\Pi_z))$. Then, we have $\sigma_{2r_z+1}(E_z) \leq |E_z|_{\text{op}} = O_P\left(\sqrt{N\vee T}\right)$ which implies $z_N/\sqrt{N\vee T} = O_P\left(\sigma_{r_z}(\Pi_z)/\sigma_{2r_z+1}(E_z)\right)$. Moreover, by Lemma 5.4, we have $\max_{j\in\left\{1,\ldots,\lfloor\sqrt{N\wedge T}\rfloor\right\}} |E_z|_{\text{op}}/\sigma_{r_z+j}(E_z) = O_P(1)$. Because

$$\max_{j\in\{1,\ldots,r_z-1\}} \sigma_r\left(\Pi_z\right)/\sigma_{r+1}\left(\Pi_z\right) = o_P\left(z_N/\sqrt{N\vee T}\right),$$

we obtain

$$\mathbb{P}\left(\left(\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_r\left(\Pi_z\right)}{\sigma_{r+1}\left(\Pi_z\right)}\right) \vee \left(\max_{j\in\left\{r_z+1,\ldots,\lfloor\sqrt{N\wedge T}\rfloor\right\}} \frac{|E_z|_{\text{op}}}{\sigma_{r_z+j}\left(E_z\right)}\right) \leq C\frac{\sigma_{r_z}\left(\Pi_z\right)}{\sigma_{2r_z+1}\left(E_z\right)}\right) \to 1.$$

Proof of Lemma 4.4.

To prove Lemma 4.4, let us show that $\mathbb{P}\left(\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_r(Y_z)}{\sigma_{r+1}(Y_z)}\right) \to 1$. Take $j \in \{1,\ldots,r_z-1\}$, by Lemma 5.3 (i) and Lemma 5.1, we have $\sigma_j(\Pi_z) - 2\,|E_z|_{\text{op}} \leq \sigma_j(Y_z) \leq$

$\sigma_j(\Pi_z) + 2\left|E_z\right|_{\mathrm{op}}$. Then, on the event $\mathcal{A} = \left\{\sigma_{r_z}(\Pi_z) > 2\left|E_z\right|_{\mathrm{op}}\right\}$, we obtain

$$\frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} \leq \frac{\sigma_j(\Pi_z) + 2\left|E_z\right|_{\mathrm{op}}}{\sigma_{j+1}(\Pi_z) - 2\left|E_z\right|_{\mathrm{op}}} = \frac{\sigma_j(\Pi_z)}{\sigma_{j+1}(\Pi_z)} \frac{1 + \frac{2|E_z|_{\mathrm{op}}}{\sigma_j(\Pi_z)}}{1 - \frac{2|E_z|_{\mathrm{op}}}{\sigma_{j+1}(\Pi_z)}} \leq \frac{\sigma_j(\Pi_z)}{\sigma_{j+1}(\Pi_z)} \frac{1 + \frac{2|E_z|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}}{1 - \frac{2|E_z|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}}, \tag{5.5}$$

where the last equality is because $\sigma_j(\Pi_z) \geq \sigma_{j+1}(\Pi_z) \geq \sigma_{r_z}(\Pi_z)$. Also, by Lemma 5.3, on $\mathcal{A}$, it holds that

$$\frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)} \geq \frac{\sigma_{r_z}(\Pi_z) - 2\left|E_z\right|_{\mathrm{op}}}{\sigma_{2r_z+1}(E_z)} = \frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)} \left(1 - \frac{2\left|E_z\right|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}\right). \tag{5.6}$$

Let us call $\mathcal{B}$ the event

$$\left\{\left(1 - \frac{2\left|E_z\right|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}\right)^2 \Big/ \left(1 + \frac{2\left|E_z\right|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}\right) > C\right\},$$

where C is the constant in Assumption 4.1. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)}\right) \\
&\geq \mathbb{P}\left(\left\{\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)}\right\} \cap \mathcal{A} \cap \mathcal{B}\right) \\
&\geq \mathbb{P}\left(\left\{\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(\Pi_z)}{\sigma_{j+1}(\Pi_z)} < \frac{\left(1 - \frac{2|E_z|_{\mathrm{op}}}{\sigma_{r_z}}(\Pi_z)\right)^2}{1 + \frac{2|E_z|_{\mathrm{op}}}{\sigma_{r_z}(\Pi_z)}} \frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)}\right\} \cap \mathcal{A} \cap \mathcal{B}\right) \quad \text{(by (5.5) and (5.6))} \\
&\geq \mathbb{P}\left(\left\{\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(\Pi_z)}{\sigma_{j+1}(\Pi_z)} < C\frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)}\right\} \cap \mathcal{A} \cap \mathcal{B}\right) \to 1,
\end{aligned}$$

where the last statement holds because $\mathbb{P}(\mathcal{A}) \to 1$, $\mathbb{P}(\mathcal{B}) \to 1$ (given that $\left|E_z\right|_{\mathrm{op}} = O_P(\sigma_{r_z}(\Pi_z))$) and

$$\mathbb{P}\left(\max_{j\in\{1,\ldots,r_z-1\}} \frac{\sigma_j(\Pi_z)}{\sigma_{j+1}(\Pi_z)} < C\frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)}\right) \to 1$$

by Assumption 4.1. Next, let us show that, $\mathbb{P}\left(\max_{j\in\{r_z+1,\ldots,\lfloor\sqrt{N\wedge T}\rfloor\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)}\right) \to 1$. By Lemma 5.3 (ii), we have, for all $j > r_z$,

$$\frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} \leq \frac{\left|E_z\right|_{\mathrm{op}}}{\sigma_{r_z+j}(E_z)}. \tag{5.7}$$

Let $\mathcal{C} = \left\{1 - \frac{2|E_z|_{\text{op}}}{\sigma_{r_z}(\Pi_z)} > C\right\}$. This implies that

$$\begin{aligned}
&\mathbb{P}\left(\max_{j \in \left\{r_z+1,\ldots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)}\right) \\
&\geq \mathbb{P}\left(\left\{\max_{j \in \left\{r_z+1,\ldots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)} < \frac{\sigma_{r_z}(Y_z)}{\sigma_{r_z+1}(Y_z)}\right\} \cap \mathcal{A} \cap \mathcal{C}\right) \\
&\geq \mathbb{P}\left(\left\{\frac{|E_z|_{\text{op}}}{\sigma_{r_z+j}(E_z)} < \left(1 - \frac{2\,|E_z|_{\text{op}}}{\sigma_{r_z}(\Pi_z)}\right)\frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)}\right\} \cap \mathcal{A} \cap \mathcal{C}\right) \quad \text{(by (5.6) and (5.7))} \\
&\geq \mathbb{P}\left(\left\{\frac{|E_z|_{\text{op}}}{\sigma_{r+j}(E_z)} < C\frac{\sigma_{r_z}(\Pi_z)}{\sigma_{2r_z+1}(E_z)}\right\} \cap \mathcal{A} \cap \mathcal{C}\right) \to 1,
\end{aligned}$$

where the last statement holds because $\mathbb{P}(\mathcal{A}) \to 1$, $\mathbb{P}(\mathcal{C}) \to 1$ (given that $|E_z|_{\text{op}} = O_P(\sigma_{r_z})$)) and

$$\mathbb{P}\left(\frac{|E_z|_{\text{op}}}{\sigma_{r_z+j}(E_z)} < C\frac{\sigma_{r_z}(\Pi_z)}{\sigma_{r_z+1}(\Pi_z)}\right) \to 1$$

by Assumption 4.1. In the end, we obtain

$$\mathbb{P}\left(\left(\max_{j \in \{1,\ldots,r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)}\right) \vee \left(\max_{j \in \left\{r_z+1,\ldots,\left\lfloor\sqrt{N\wedge T}\right\rfloor\right\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)}\right) < \frac{\sigma_{r_z}(Y_z)}{\sigma_{2r_z+1}(Y_z)}\right) \to 1,$$

which concludes the proof.

Proof of Theorem 4.1.

We denote $\mathcal{A} = \{\widehat{r}_z = r_z\}$. We have

$$\mathbb{P}\left(\left|\widehat{M}_z - M_z\right|_2 \leq 4\sqrt{2r_z}\frac{|E_z|_{\text{op}}}{\sigma_{r_z}(\Pi_z)}\right) \geq \mathbb{P}\left(\left\{\left|\widehat{P}_z - P_z\right|_2 \leq 4\sqrt{2r_z}\frac{|E_z|_{\text{op}}}{\sigma_{r_z}(\Pi_z)}\right\} \cap \mathcal{A}\right) = \mathbb{P}(\mathcal{A}) \to 1,$$

by Lemma 5.2.