Factor and Factor Loading Augmented Estimators for Panel Regression With Possibly Nonstrong Factors

Jad Beyhum and Eric Gautier

ABSTRACT
This article considers linear panel data models where the dependence of the regressors and the unobservables is modeled through a factor structure. The number of time periods and the sample size both go to infinity. Unlike in most existing methods for the estimation of this type of models, nonstrong 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. The estimators enjoy double robustness. 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. The approach is illustrated by an empirical application on fiscal policy.

1. Introduction
This article considers inference on $\beta \in \mathbb{R}^K$ in the following model:

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

(1.1)

where the data consist of the outcome $Y_{it}$ and the regressors $X_{kit}$ for all $k = 1, \ldots, K$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The random vectors $\lambda_{ij}$ and $f_j$ in $\mathbb{R}^{r_N}$ are factor loadings and factors, $\delta$ 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 Pesaran 2015). It allows for flexible cross-section and serial correlation thanks to the factor structure in the regression error. This article proposes 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 $k \in \{1, \ldots, K\}$ and errors $E_1, \ldots, E_K$ which are $N \times T$ random matrices such that

$$X_{kit} = \sum_{j=1}^{r_N} \lambda_{ij} f_j \delta_j + E_{kit}, \quad k \in \{1, \ldots, K\}.$$  

(1.2)

The role of the vectors $\delta_1, \ldots, \delta_K$ is to model the dependence between the regressors and the unobservables $\sum_{j=1}^{r_N} \lambda_{ij} f_j \delta_j + E_i$. They ensure that the model is symmetric with respect to the factors and the factor loadings, which is more natural in the context of this article. The classical model of the factor augmented regression literature (as in Greenaway-McGrevy et al. 2012) can be rewritten as (1.1)–(1.2). Indeed, consider the simple model

$$Y_{it} = \beta_1 X_{1it} + \tilde{\lambda}_{1it} \tilde{f}_i + E_{it};$$

$$X_{1it} = \tilde{\gamma}_{1it} + E_{1it},$$

where $\tilde{\lambda}_{1i}, \tilde{\gamma}_{1i}, i = 1, \ldots, N, t = 1, \ldots, T$ are scalar random variables. This is the model of the factor augmented regression literature with one factor and one regressor. Then, it can be rewritten as our model (1.1) and (1.2) with $r_N = 2, \lambda_{11} = \tilde{\lambda}_{1i}, \lambda_{12} = \tilde{\gamma}_{1i}, f_{11} = \tilde{f}_i, \delta = (1, 0)^{\top}, \delta_1 = (0, 1)^{\top}$. This example is generalized to the case of several regressors and factors in Section 4.2. Because some coordinates of $\delta_1, \ldots, \delta_K$ can be equal to 0, our model does not imply that the regressors share the same factors. An economic interpretation of Equations (1.1) and (1.2) can be given as follows. There is a set of $r_N$ unobserved time-varying factors $f_j, j = 1, \ldots, r_N$, which affect both the regressors and the outcome. These unobservables can create omitted variable bias. The effect of the $j$th factor on the level of variable $k$ for subject $i$ is modulated by the quantities $\delta_{kj}$ and $\lambda_{ij}$. Given that the components of the latter vectors are allowed to be 0, some factors may not affect some individuals or regressors. In our framework, $T$ goes to infinity as a function of $N$ and we allow the number of factors $r_N$ to go to infinity with $N$. Such an asymptotic regime models situations where new factors can enter the model when additional subjects and dates are added. This interpretation of the model is further discussed in the context our empirical application in Section 6. The structure that we impose can be seen as the generalization to...
dimension 3 (the third dimension being the one of variables) of the usual factor models for matrices as in Bai and Wang (2016). Such a modeling was already introduced in the psychometrics literature in Carroll and Chang (1970) and Harshman (1970). The mathematical foundations behind this approach lie in the tensor decomposition literature, see Kolda and Bader (2009) for a survey.

### 1.1. Contributions

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. This is the reason why we call our estimators factor and factor-loading augmented. This approach gives some double robustness to the estimator as in Chernozhukov et al. (2018). Errors on the estimation of the factors can be offset by precise estimation of the factor loadings (see Section 2). The reverse is also true. We provide high-level sufficient conditions on the first-step estimator under which the two-step estimator is asymptotically normal. Then, we apply this theory to the case where the first-step estimator is based on principal components analysis (henceforth PCA). When the number of factors is fixed and the factors are strong, we derive rate conditions on $N$ and $T$ such that the estimator is asymptotically normal. Under this regime, our estimator is asymptotically normal under weaker conditions on $N$ and $T$ than the usual factor augmented estimator (Greenaway-McGrevy et al. 2012). This demonstrates that including estimates of the factor loadings in the second-step regression may improve estimation performance. It results from the double robustness property of the estimator. Then, we show that this estimator can satisfy our high-level conditions even when the numbers of factors is unknown and goes to infinity with the sample size, and that the factors are not strong. This is in contrast with the main alternative approaches for panel data models with interactive fixed effects. 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 $N$. Such a regime makes the notation less cumbersome, but the results would remain valid if $N$ and $T$ were to jointly go to infinity. The proposed principal components augmented estimator exhibits better finite sample properties than alternatives in Monte-Carlo simulations. Finally, we apply our procedure to real data on government resources.

### 1.2. Related Literature

Several alternative techniques have been developed to estimate panel data models with interactive fixed effects. Assuming (1.1) but not (1.2), Bai (2009) proposed to estimate jointly the regression coefficient and the factors and factor loadings. Moon and Weidner (2018) and Beyhum and Gautier (2019) study a nuclear-norm penalized estimator of a model similar to that of Bai (2009). In contrast, the CCE estimator of Pesaran (2006) and the factor augmented regression estimator studied in Kapetanios and Pesaran (2005) and Greenaway-McGrevy et al. (2012) model the dependence between the regressors and the unobservables $\sum_{j=1}^{K} \tilde{\lambda}_{k} f_{j} \delta_{j} + E_{k}$. They assume that, for $k \in \{1, \ldots, K\}$, there exists $\tilde{\lambda}_{k1}, \ldots, \tilde{\lambda}_{kN}$ which are random vectors in $\mathbb{R}^{N}$ and mean-zero errors $E_{1}, \ldots, E_{K}$ which are $N \times T$ random matrices such that $X_{kt} = \sum_{r=1}^{K} \tilde{\lambda}_{k} f_{r} f_{r}^{T} + E_{kt}$ for $k \in \{1, \ldots, K\}$. This means that the regressors have a factor structure with the same factors as the error term but possibly different factor loadings. In the articles of Pesaran (2006), Bai (2009), and Greenaway-McGrevy et al. (2012), a strong factor assumption is imposed. It means that the empirical second moments matrices of the factors and the factor loadings converge in probability to their population counterparts. The number of factors is also assumed to be fixed with the sample size. It is worth noting that some articles have sought to relax these assumptions in the context of the CCE (Chudik, Pesaran, and Tosetti 2011) and factor augmented (Reese and Westerlund 2018) estimators.

The differences with the aforementioned works are as follows. First, our estimator can be asymptotically normal even when the number of factors is not fixed and the factors are not strong. Second, the model of Greenaway-McGrevy et al. (2012) can be rewritten as ours (see Section 4.2). When the factors of the model of Greenaway-McGrevy et al. (2012) are strong, then those of our model (1.1) and (1.2) are strong too. The estimator of Greenaway-McGrevy et al. (2012) differed from ours because its second-step regression only includes estimates of the factors (but not of the factor loadings). As already mentioned, we show that, under this strong factor assumption, our estimator is asymptotically normal under weaker conditions on $N$ and $T$ than the factor augmented estimator of Greenaway-McGrevy et al. (2012). This is due to the aforementioned double robustness property. As a result, estimation performance is improved by inclusion of estimates of the factor loadings in the second-step regression. This is illustrated by our simulation experiment, where our PCA-based estimator performs better than the factor augmented estimator. Third, the theory in Bai (2009) is developed for an estimator corresponding to the global minimum of a least-square criterion. However, in practice, this least-square criterion is nonconvex and only a local minimum of the estimator can be computed. In contrast, our proposed estimation approach only relies on singular values decomposition and ordinary least square. Therefore, our estimator can be computed in practice.

Note also that the estimator of Beyhum and Gautier (2019, sec. 4.7.1) is a special case of the two-step procedure of this article. The differences are as follows. First, we outline a general theory of two-step estimators in our model, while the results of Beyhum and Gautier (2019) are only valid when the first-step estimator is based on hard-thresholding. Second, we propose and study a two-step estimator based on PCA. Contrarily to the approach in Beyhum and Gautier (2019), our estimator does not rely on the choice of a tuning parameter.

### 1.3. Outline

This article is organized as follows. First, the class of two-step estimators 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. Finally, we illustrate our approach by an application on fiscal policy in Section 6. Proofs are deferred to the appendix and the supplementary material.

1.4. Preliminaries

The transpose of a $N \times T$ matrix $A$ is written $A^T$ and its trace is $tr(A)$. Its $k$th singular value is $\sigma_k(A)$ and rank($A$) is its rank. $A = \sum_{k=1}^{\text{rank}(A)} \sigma_k(A)u_k(A)v_k(A)^T$ is the singular value decomposition of $A$, where $(u_k(A))_{k=1}^{\text{rank}(A)}$ is a family of orthonormal vectors of $\mathbb{R}^N$ and $(v_k(A))_{k=1}^{\text{rank}(A)}$ is a family of orthonormal vectors of $\mathbb{R}^T$. The scalar product in the space of $N \times T$ matrices is $(A,B) = tr(A^T B)$. The nuclear norm is $|A|_n = \sum_{k=1}^{\text{rank}(A)} |\sigma_k(A)|$ and the operator norm is $|A|_o = \sigma_1(A) = \max_{x \in \mathbb{R}^T, x \neq \mathbf{0}} \frac{|Ax|^2}{{|x|^2}}$. Let us also define the euclidian norm $|A|_2 = \sqrt{(A,A)}$ and the sup-norm $|A|_\infty = \max_{i \in \{1, ..., N\}, t \in \{1, ..., T\}} |A_{it}|$. The Hadamard product of matrices $A$ and $B$ is $A \odot B$. For two integers, $N$ and $T$, $N \land T$ is the maximum of $N$ and $T$, $N \lor T$ is the minimum of $N$ and $T$ and $[N]$ is the integer part of $N$. For $N \in \mathbb{N}$, $I_N$ is the identity matrix of size $N$. Convergence in probability (resp. in distribution) is denoted $P \xrightarrow{\text{d}}$ (resp. $P \xrightarrow{\text{a.s.}}$). We consider sequences of data generating processes indexed by $s$. $N$ is a function of $s$ that goes to infinity with $s$. This article studies an asymptotic where $N$ goes to infinity. For a probabilistic event $A$, its complement is denoted $\overline{A}$ and we write that $A$ happens with probability approaching 1 or w.p.a 1 if $P(\overline{A}) \to 1$.

2. The Estimator

The model can be rewritten in matrix form as $Y = \Pi_0 + E_0$, $X_k = \Pi_k + E_k$ for $k \in \{1, ..., K\}$, where $\Pi_{it} = \sum_{i=1}^{N} \lambda_i f_i \delta_{ij}$ for $i \in \{1, ..., N\}$ and $t \in \{1, ..., T\}$, $\Pi_0 = \sum_{k=1}^{K} \beta_k \Pi_k + \Gamma$, $\Gamma_{it} = \sum_{i=1}^{N} \lambda_i f_i \delta_{t}$ and $E_0 = \sum_{k=1}^{K} \beta_k E_k + E$. Notice that $E_0$ and $E_i$ 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 $E_0, E_1, ..., E_K$ have mean zero, hence they can be the sum of an error term with mean zero and a small remainder.

Let $\Pi_0 = (\Pi_{01}, ..., \Pi_{0T})$ ($N \times T(K + 1)$ matrix) and $\Pi_j = (\Pi_{j0}, ..., \Pi_{jT})$ ($N \times T(K)$ matrix). For $z = u, v$, we denote by $P_z$ the projector on the vector space spanned by the columns of $\Pi_z$ and $M_z$ the projector on the orthogonal of the vector space spanned by the columns of $\Pi_z$. Notice that $P_u$ (resp. $P_v$) corresponds also to the projector on the vector space spanned by the factor loadings (resp. factors). Let $r_z$ be the rank of $\Pi_z$. Notice that $\Pi_0$ and $\Pi_1$ are different matrices of the three-dimensional $K \times N \times T$ tensor $(\Pi_0 \Pi_1 \cdots \Pi_K)$. The tensor rank of this tensor is $\min_{\Pi_0} r_N r_z$ because it can be written as the sum of $r_N$ rank one tensors which correspond to the factors. Then, $r_u$ is what is called its mode-2 rank and $r_v$ its mode-3 rank. It is known from the tensor theory that mode-$k$ ranks are not necessarily equal. Also, we have $r_z \leq r_N$, $z = u, v$ because the mode-$k$ rank is lower than the tensor rank. For more details, we refer to Kolda and Bader (2009) and De Lathauwer et al. (2000, sec. 2.2).

The proposed estimator is as follows. In the first step, one estimates $M_u$ and $M_v$ by estimators $\hat{M}_u$ and $\hat{M}_v$. From there, the estimator of $\beta$ is

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^K} \left| \hat{E}_0 - \sum_{k=1}^{K} \beta_k \hat{E}_k \right|_2^2,$$  

(2.1)

where $\hat{E}_0 = \hat{M}_u Y \hat{M}_v$ and $\hat{E}_k = \hat{M}_u X_k \hat{M}_v$ for $k \in \{1, \ldots, K\}$.

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 $\tilde{\tau}_u = \text{rank}(I_N - \hat{M}_u), \tilde{\tau}_v = \text{rank}(I_T - \hat{M}_v)$ and $X_{it} = (X_{1it}, \ldots, X_{Kt})^T$.

**Lemma 2.1.** Let $\{\tilde{\lambda}_i\}_{i=1}^{N}$ (resp. $\{\tilde{f}_j\}_{j=1}^{T}$) be a family of vectors in $\mathbb{R}^{\hat{\tau}_u}$ (resp. $\mathbb{R}^{\hat{\tau}_v}$) such that $(\tilde{\lambda}_{ij} \cdots \tilde{\lambda}_{IN})^T$ (resp. $(\tilde{f}_{j1} \cdots \tilde{f}_{jT})^T$) is a generating family of the orthogonal of the null space of $\hat{M}_u$ (resp. $\hat{M}_v$). Then, it holds that

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^K} \min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{\hat{\tau}_u}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \langle X_{it} - \tilde{\lambda}_{ij} \phi_t - \tilde{f}_{jt} \rangle^2 \right).$$

Note that the rationale for using (2.1) as the second-step estimator is based on the following moments:

$$E \left[ \sum_{k=1}^{K} \beta_k M_u \hat{X}_{k} M_v Y - \sum_{k=1}^{K} \beta_k M_u \hat{X}_{k} M_v \right] = 0, \quad l = 1, \ldots, K,$$

$$E \left[ \sum_{k=1}^{K} \beta_k M_u \hat{E}_k M_v Y - \sum_{k=1}^{K} \beta_k M_u \hat{E}_k M_v \right] = 0, \quad l = 1, \ldots, K.$$

(2.2)

We do not assume that (2.2) is true, but assumption Assumption 3.1 implies that (2.2) holds asymptotically. The moments (2.2) have the nice property that $M_u$ and $M_v$ are multiplied by an error term ($E_i$ or $E_k$), which has small operator norm. Hence, any estimation error on $M_u$ or $M_v$ has a negligible impact on $\hat{\beta}$. In fact, one can show that, under some conditions, (2.2) is a Neyman orthogonality condition as in Chernozhukov et al. (2018). This explains why our estimator overperforms the factor-augmented estimator. It is the reason why we say that the estimator enjoys a double robustness property.

Instead, the factor-augmented regression literature (Greenaway-McGrevy et al. 2012) relies on the moments

$$E \left[ X_{it} M_v Y - \sum_{k=1}^{K} \beta_k X_{k} M_v \right] = 0, \quad l = 1, \ldots, K.$$

This time $M_v$ is not multiplied by an error term but by a regressor, which has large operator norm. Hence, an estimation error on $M_v$ does not have a negligible effect on the factor-augmented estimator.
3. Sufficient Assumptions for Asymptotic Normality

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

**Assumption 3.1.** The following holds:

i. There exists a $K \times K$ positive-definite matrix $\Sigma$ such that, for $k, l \in \{1, \ldots, K\}$, $[\tilde{E}_k, \tilde{E}_l] \sim (NT)^{-1/2} \Sigma_{kl}$;

ii. There exists $\sigma > 0$ such that $[\tilde{E}]^2 / (NT) \overset{p}{\rightarrow} \sigma^2$ and $\left(\tilde{E}_k \tilde{E}_l\right)_{k=1}^{NT} / (NT) \overset{d}{\rightarrow} N(0, \sigma^2 \Sigma)$.

This assumption ensures that the (unfeasible) OLS estimator of the regression of $E_0$ on $E_1, \ldots, E_K$ is well behaved. It is similar to hypotheses F in Bai (2009) and B in Greenaway-McGrevy et al. (2012). The difference with the assumptions of these articles is that the projectors used to define $\tilde{E}_0, \ldots, \tilde{E}_K$ are different to the ones in the present article. For instance, hypothesis B in Greenaway-McGrevy et al. (2012) makes Assumption 3.1 on $P_0 E_k$, $k = 1, \ldots, K$ and $P_0 E$ rather than $\tilde{E}_k$, $k = 1, \ldots, K$ and $E$. Neither Bai (2009) nor Greenaway-McGrevy et al. (2012) provided sufficient conditions for their assumptions. In Section S.1 of the supplementary material, we give our own sets of sufficient conditions. Remark that Moon and Weidner (2017, lem. B.2) provides a set of sufficient conditions for (3.1) similar to ours. Note also that condition (3.1) in Assumption 3.1 implicitly makes an homoscedasticity assumption. It can be seen from our proofs that homoscedasticity is not needed to obtain asymptotic normality. We only assume homoscedasticity because it simplifies the theory and the practice of the estimation of the asymptotic variance.

The last set of conditions concerns the performance of the estimators of the projectors $\tilde{M}_u$ and $\tilde{M}_v$. Let $\{u_N\}$ and $\{v_N\}$ be real-valued sequences such that $|\tilde{M}_u - M_u|^2 = \mathcal{O}(u_N)$ and $|\tilde{M}_v - M_v|^2 = \mathcal{O}(v_N)$. Let also $\{h_N\}$ and $\{\rho_N\}$ be real-valued sequences such that $\max_{k \in \{0, \ldots, K\}} |\Pi_k + E_k|^2 = \mathcal{O}(h_N)$ and $\max_{k \in \{0, \ldots, K\}} |E_k|_{op} = \mathcal{O}(\rho_N)$. For $k = 1, \ldots, K$, $|\Pi_k + E_k|^2 = |X_k|^2$ and, hence it is $\sqrt{NT}$ times the empirical second moment of $X_k$. Therefore, under reasonable assumptions on the cross-sectional and serial dependence of the variables, $h_N$ can be chosen equal to $\sqrt{NT}$. In the panel data literature, it is often assumed that $|E_k|_{op} = \mathcal{O}(\sqrt{NT})$. Such an assumption can hold under flexible cross-sectional and serial correlations. For a detailed discussion, see Moon and Weidner (2015, app. A.1). Hence, it makes sense to expect that $\rho_N = \sqrt{NT}$. The estimators satisfy the following assumption.

**Assumption 3.2.** The following holds:

i. $\tilde{M}_u$ and $\tilde{M}_v$ are symmetric almost surely;

ii. $\mathbb{P}(\tilde{r}_u = r_u) \rightarrow 1$ and $\mathbb{P}(\tilde{r}_v = r_v) \rightarrow 1$;

iii. $u_N \lor v_N \approx o(1)$ and $h_N^2 = \mathcal{O}(NT)$;

iv. $\sqrt{N}(u_N \lor v_N) h_N = \sigma(\sqrt{NT})$;

v. $u_N v_N h_N = o(\sqrt{NT})$;

It is difficult to understand the strength of Assumption 3.2 without examples of $u_N$ and $v_N$. Hence, we discuss it in Section 4, where we derive the properties of $\tilde{M}_u$ and $\tilde{M}_v$ when they are estimated by a method relying on PCA. The next theorem constitutes the main result of this article.

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

$$\sqrt{NT}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2 \Sigma^{-1}).$$

Also, for $k, l \in \{1, \ldots, K\}$, $\tilde{\Sigma}_{kk} = \tilde{E}_k \tilde{E}_l / (NT) \Rightarrow \Sigma_{kl}$ and $\tilde{\sigma}^2 = \tilde{E}_0 - \sum_{k=1}^K \tilde{\rho}_k \tilde{E}_k^2 / (NT) \Rightarrow \sigma^2$.

As will be seen in the next section, Assumption 3.2 allows that the estimator is asymptotically normal even when $1/\sqrt{NT} = o_p(u_N v_N)$, that is the error in estimating the projectors $M_u$ and $M_v$ is of order larger than the estimation error of $\beta$. This is due to novel proof techniques. In the proof of Theorem 3.1, when bounding $\|\langle(\hat{M}_u - \tilde{M}_u)X_1 M_v, E\rangle\|$, we use the fact that $M_u - \tilde{M}_u$ has low rank by Assumption 3.2 (3.2). This strategy also enables us to allow for nonstrong factors.

4. Estimation of the Projectors Using Principal Components Analysis

4.1. Convergence Results

The econometrician can use different methods to estimate the projectors $M_u$ and $M_v$. The approach in Beyhum and Gautier (2019) relies on a nuclear-norm penalized 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 Gagliardi and Gourieroux (2017). In this article, 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 Ahn and Horenstein (2013). For $z = u, v$, let us define

$$\tilde{r}_z \in \operatorname{argmax}_{\beta \in \{1, \ldots, \sqrt{N/\beta}\}} \frac{\sigma_j(Y_{z\beta})}{\sigma_{j+1}(Y_{z\beta})},$$

where $Y_u = (Y_1, X_1, \ldots, X_K)$ and $Y_v = (Y_1^\top, X_1^\top, \ldots, X_K^\top)$. It may be that there exists $r \in \{1, \ldots, \sqrt{N/\beta}\}$, $\sigma_j(Y_{z\beta}) = 0$. 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 $\infty$. The estimator in Ahn and Horenstein (2013) is of the form

$$\tilde{r}_z \in \operatorname{argmax}_{\beta \in \{1, \ldots, \sqrt{N/\beta}\}} \frac{\sigma_j(Y_{z\beta})}{\sigma_{j+1}(Y_{z\beta})},$$

where $d^* \in [0, 1]$. Therefore, the estimators in (4.1) correspond to the one in Ahn and Horenstein (2013) for a particular choice of $d^*$. Our theoretical analysis is different from the one of Ahn and Horenstein (2013), because it allows for nonstrong factors and a growing number of factors. Contrarily to the estimators
in Bai and Ng (2002), the advantage of the eigenvalue ratio estimator is that it does not require to choose a penalty level. Note also that an alternative estimator is the growth ratio estimator also introduced in Ahn and Horenstein (2013), we discuss it further in Section 5. To ensure consistency of the eigenvalue ratio estimator, we make the following assumption. Let $E_u = (E_0, \ldots, E_K)$ and $E_v = (E_0^T, \ldots, E_K^T)$.

**Assumption 4.1 (Eigenvalue Ratio).** For $z = u, v$, it holds that $1 \leq r_z \leq \sqrt{NT}$ almost surely, $|E_z|_{op} = Op(\sigma_{r_z}(\Pi_z))$ and there exists $C < 1$ such that

$$\mathbb{P}\left( \max_{j \in [1, \ldots, r_z-1]} \frac{\sigma_j(\Pi_z)}{\sigma_{r_z}(\Pi_z)} \leq C \frac{\sigma_{r_z}(\Pi_z)}{\sigma_{r_z+1}(\Pi_z)} \right) \rightarrow 1.$$

Notice that it is assumed that $r_z \geq 1$, $z = u, v$ (there is at least one nonzero factor and factor loading). In Section S.2 of the supplementary material, we show how to modify estimator (4.1) to handle the case $r_z = 0$. Let us give sufficient conditions for Assumption 4.1.

**Lemma 4.1.** For $z = u, v$, assume that $r_z \leq \sqrt{NT}$, $|E_z|_{op} = Op\left(\sqrt{\frac{N}{NT}}\right)$ has a finite deterministic limit in probability and there exists a sequence $(\varepsilon_N)$ such that $\sigma_{r_z}(\Pi_z) = Op(\varepsilon_N)$, $\sqrt{\frac{N}{NT}} = o(\varepsilon_N)$ and $\max_{j \in [1, \ldots, r_z-1]} \sigma_j(\Pi_z) / \sigma_{r_z}(\Pi_z) = op\left(\varepsilon_N / \sqrt{\frac{N}{NT}}\right)$, then Assumption 4.1 holds.

This lemma shows that our assumption allows for nonstrong factors and a growing number of factors. The condition $\sigma_{j}(\Pi_z) / \sigma_{j+1}(\Pi_z) = op\left(\varepsilon_N / \sqrt{\frac{N}{NT}}\right)$ implies that the singular values of $\Pi_z$ cannot decrease too quickly with $j \in \{1, \ldots, r_z\}$. The assumption that $|E_z|_{op} = Op\left(\sqrt{\frac{N}{NT}}\right)$ 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 Moon and Weidner (2015). Let us now state the main result regarding the eigenvalue ratio estimator.

**Lemma 4.2.** Under Assumption 4.1, we have $\mathbb{P}(\widehat{r}_u = r_u, \widehat{r}_v = r_v) \rightarrow 1$.

Given the estimators $\widehat{r}_u$ and $\widehat{r}_v$, we set $\widehat{M}_u = I_N - \sum_{j=1}^{r_u} u_i (Y_i u_j)^T$ and $\widehat{M}_v = I_T - \sum_{j=1}^{r_v} u_i (Y_j u_j)^T$. 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 $\mathbb{P}(\widehat{r}_z = r_z) \rightarrow 1$, we have $|\widehat{M}_u - M_u|_2 = Op(\sqrt{\frac{N}{NT}})$ and $|\widehat{M}_v - M_v|_2 = Op(\sqrt{\frac{N}{NT}})$ and $|\widehat{E}_z|_{op} = Op\left(\sqrt{\frac{N}{NT}}\right)$.

Using Theorem 3.1, we obtain the following corollary, which gives lower level conditions for asymptotic normality of our PCA-based estimator.

**Corollary 4.1.** Let us assume that $\max_{k \in \{0, \ldots, K\}} |\Pi_k + E_k|_2 = Op(\sqrt{NT})$ and $\max_{k \in \{0, \ldots, K\}} |E_k|_{op} = Op(\sqrt{NT})$ and $\max_{k \in \{0, \ldots, K\}} |E_k|_{op} = Op(\sqrt{NT})$.

Assumptions 3.1 and 4.1 hold. Then, if

$$r_N^2(N \vee T)^{3/2} = op\left(\left(\max_{j \in \{1, \ldots, r_N\}} \sigma_j(\Pi_z)\right)^2\right),$$

we have

$$\sqrt{NT}(\hat{\beta} - \beta) \overset{d}{\rightarrow} N\left(0, \sigma^2 \Sigma^{-1}\right).$$

**4.2. Strong Factors**

Now, we seek to relax condition (4.2) to the properties of the factors. Let $\Lambda = (\lambda_1, \ldots, \lambda_N, F = (f_1, \ldots, f_N)$, and $\Delta = (\delta_0, \ldots, \delta_K)$, where $\delta_0 = \delta + \sum_{k=1}^{K} \beta_k \delta_k$. One can show that when the factors are strong, $\sigma_j(\Pi_z) / \sqrt{NT}, j = 1, \ldots, r_z$ has a deterministic limit in probability. This result is stated in the following Lemma.

**Lemma 4.3.** Assume that $r_N = r$ is fixed and

i. There exists a $r \times r$ matrix $\Sigma_\Lambda$ such that $\Lambda \Lambda^T / N \overset{p}{\rightarrow} \Sigma_\Lambda$;

ii. There exists a $r \times r$ matrix $\Sigma_F$ such that $FF^T / T \overset{p}{\rightarrow} \Sigma_F$;

iii. $\Delta \Delta^T$ does not depend on $N$.

Then, for $z = u, v$, $\sigma_{r_z}(\Pi_z) / \sqrt{NT}, j = 1, \ldots, r_z$, has a strictly positive deterministic limit in probability.

The conditions in Lemma 4.3 correspond to a strong factor assumption in our model. Under these requirements, $(4.2)$ is satisfied as long as $(N \vee T) / (N \wedge T^2) = o(1)$. The latter corresponds to the condition for asymptotic normality of the debiased estimator in Bai (2009). As previously mentioned, this requirement is weaker than the conditions for asymptotic normality in Greenaway-McGreyvey et al. (2012) (which include $T/N = o(1)$ and $T/N^3 = o(1)$).

Let us further discuss the comparison with the factor augmented regression literature (Greenaway-McGreyvey et al. 2012). In the latter, the following model is assumed. There are $\tilde{r} \in \mathbb{N}$ factors with $\tilde{r}$ fixed. There exists factor loadings $\{\tilde{\lambda}_{ki}\}_{k=1}^{K}$, $i \in \{1, \ldots, N\}$ and factors $\tilde{f}_i, t \in \{1, \ldots, T\}$, all with support in $\mathbb{R}^{\tilde{r}}$, such that

$$Y_{it} = \sum_{k=1}^{K} \beta_{ki} X_{ki} + \sum_{j=1}^{\tilde{r}} \tilde{\lambda}_{0j} \tilde{f}_j + E_{it},$$

$$X_{ki} = \sum_{j=1}^{\tilde{r}} \tilde{\lambda}_{kj} \tilde{f}_{j} + E_{ki}, k \in \{1, \ldots, K\}.$$
Greenaway-McGrevy et al. (2012), our estimator is asymptotically normal under weaker conditions on \( N \) and \( T \) than the factor augmented estimator of Greenaway-McGrevy et al. (2012).

### 4.3. Nonstrong Factors

It is possible to relate the strength of the factors to the singular values of \( \Pi_2 \), \( z = u, v \), even in the case where the factors are not strong. The following lemma studies the case where the factors are not strong and the number of factors is allowed to grow to infinity with the sample size.

**Lemma 4.4.** Assume that there exists a constant \( C > 0 \) such that

i. For all \( N \in \mathbb{N} \), there exists a \( r_N \times r_N \) matrix \( \Sigma_{AN} \) with \( |\Sigma_{AN}|_{op} \leq C, 1/\sigma_{rN}(\Sigma_{AN}) \leq C \) and \( |N^{-1} \Lambda \Lambda^T - \Sigma_{AN}|_2 \xrightarrow{P} 0 \); 

ii. For all \( N \in \mathbb{N} \), there exists a sequence of \( r_N \times r_N \) matrices \( \Sigma_{FN} \) such that \( |\Sigma_{FN}|_{op} \leq C, 1/\sigma_{rN}(\Sigma_{FN}) \leq C \) and \( |T^{-1} FF^T - \Sigma_{FN}|_2 \xrightarrow{P} 0 \); 

iii. There exists a real-valued sequence \( \{w_N\} \) such that \( w_N \leq 1, |w_N^{-1} \Lambda \Lambda^T|_\infty = O_p(1) \);

Then, for \( z = u, v \) and \( j \in \mathbb{N} \), we have \( \sigma_j(\Pi_z)/\sqrt{w_N NT} = O_p(1) \).

Under the assumptions of **Lemma 4.4**, \( w_N \) controls the strength of the factors. Under the conditions of **Lemma 4.4**, \( (4.2) \) holds as long as \( r_N w_N (N \vee T)/(N \wedge T)^2 \rightleftharpoons o(1) \). This shows that our theory can allow for nonstrong factors and a growing number of factors.

### 5. Simulations

Our simulations investigate the effects of nonstrong factors on the performance of the estimators and demonstrate that our two-step procedure can have finite sample properties superior to those of the factor augmented estimator. We consider a data-generating process (henceforth, DGP) with a single regressor. The observed variables are generated by

\[
Y_{it} = X_{it} + \frac{1}{\sqrt{r_N}} \sum_{j=1}^{r_N} \lambda_{ij} f_{ij} + E_{it},
\]

\[
X_{it} = \frac{1}{\sqrt{r_N}} \sum_{j=1}^{r_N} \lambda_{ij} f_{ij} + E_{1it},
\]

where \( f_{it}, \lambda_{it}, E_{it}, \) and \( E_{1it} \) for all indices are mutually independent, \( f_{it} \sim N(1/2, 1), \lambda_{it} \sim N(3, 3) \) and \( E_{i1t}, \ldots, E_{iNt} \) are standard normals. In this simple DGP, we have \( \delta = \delta_1 = (1/\sqrt{r_N}) \). The DGP follows the assumptions of **Lemma 4.4** with \( w_N = r_N \). The matrix \( X_1 \) has a statistical factor structure with a low-rank component of rank \( r_N \). Recall that \( \hat{\beta}^{LS} \in \text{argmin}_{b \in \mathbb{R}} [Y - bX_1]_2^2 \) is the least-square estimator of the linear regression of the outcome on the regressors. \( \hat{\beta}^{FA} \in \text{argmin}_{b \in \mathbb{R}} [Y \hat{M}_{v} - bX_1 \hat{M}_{v}]_2^2 \) is the factor augmented regression estimator where \( \hat{M}_{v} \) is computed as in **Section 4** (using the eigenvalue ratio estimator (4.1) to estimate \( r_N \)). \( \hat{\beta}^{Bai} \) is the estimator of Bai (2009) and Moon and Weidner (2015). The number of factors for \( \hat{\beta}^{Bai} \) is set to 10 when \( N = Y = 50 \) and 20 when \( N = T = 150 \). Let \( \hat{\beta}^{ER} \) be the estimator (2.1), using the procedure of **Section 4** as the first-stage. We also introduce \( \hat{\beta}^{GR} \) which corresponds to estimator (2.1), using the procedure of **Section 4**, except that the number of factors is not estimated by the eigenvalue ratio estimator (4.1), but by a growth ratio estimator (see definition in **Appendix D.3**).

We study the effect of the number of factors for two different sample sizes. First, set \( N = T = 50 \), which roughly corresponds to the sample size of our empirical application. Our theoretical results assume that \( r_N \leq \sqrt{N \wedge T} \), which is 7 in our case. Hence, we report the results for \( r_N = 3, 5, \) and 7. Second, we use \( N = T = 150 \) and display the results for \( r_N = 5, 9, 12 \). 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. Table 3 presents the finite sample performance of the eigenvalue and growth ratio estimators for the estimation of \( r_N \) (the results for \( r_N \) are similar, and, therefore, omitted). The coverage is not reported for \( \hat{\beta}^{LS} \) because the latter is not asymptotically normal for the DGP that we consider. We use 7300 Monte-Carlo replications which allows for an

---

**Table 1.** \( N = T = 50 \).

| \( r_N \) | \( \hat{\beta}^{LS} \) Bias | \( \hat{\beta}^{LS} \) std | \( \hat{\beta}^{FA} \) Bias | \( \hat{\beta}^{FA} \) std | \( \hat{\beta}^{Bai} \) Bias | \( \hat{\beta}^{Bai} \) std | \( \hat{\beta}^{ER} \) Bias | \( \hat{\beta}^{ER} \) std | \( \hat{\beta}^{GR} \) Bias | \( \hat{\beta}^{GR} \) std |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0.96 | 0.03 | -2 \times 10^{-4} | 0.03 | 8 \times 10^{-4} | 0.83 | 0.02 | 2 \times 10^{-4} | 0.27 | 3 \times 10^{-4} |
| 5 | 0.97 | 0.07 | 1.5 \times 10^{-3} | 0.03 | 4 \times 10^{-3} | 0.81 | 0.04 | 1.95 \times 10^{-4} | 0.27 | 3 \times 10^{-4} |
| 7 | 0.97 | 0.45 | 3.1 \times 10^{-3} | 0.03 | 4 \times 10^{-4} | 0.31 | 0.04 | 1.8 \times 10^{-3} | 0.27 | 2 \times 10^{-4} |
Table 2. \( N = T = 150 \).

| \( r_N \) | \( \hat{\beta}^{LS} \) | \( \hat{\beta}^{TA} \) | \( \hat{\beta}^{Bai} \) | \( \hat{\beta}^{ER} \) | \( \hat{\beta}^{GR} \) |
|---------|---------------|---------------|---------------|---------------|---------------|
| 5 std Bias | 0.96 | -0.02 | 7 \times 10^{-5} | -4 \times 10^{-5} | -4 \times 10^{-5} |
| MSE | 3.2 \times 10^{-3} | 6.9 \times 10^{-3} | 8.4 \times 10^{-3} | 6.9 \times 10^{-3} | 6.9 \times 10^{-3} |
| Coverage | 0.32 | 0.84 | 0.94 | 0.94 |
| 9 Bias | 0.978 | -0.03 | 5 \times 10^{-4} | 4 \times 10^{-4} | 7 \times 10^{-5} |
| MSE | 2.4 \times 10^{-3} | 0.02 | 8.1 \times 10^{-3} | 0.02 | 7 \times 10^{-3} |
| Coverage | 0.95 | 1.3 \times 10^{-3} | 7 \times 10^{-5} | 4 \times 10^{-4} | 5 \times 10^{-5} |
| 12 Bias | 0.98 | 0.64 | 9 \times 10^{-4} | 0.31 | 1 \times 10^{-4} |
| MSE | 2 \times 10^{-3} | 0.41 | 8.2 \times 10^{-3} | 0.40 | 0.01 |
| Coverage | 0.96 | 0.58 | 7 \times 10^{-5} | 0.57 | 2 \times 10^{-4} |

Table 3. Probability of correct recovery of \( \hat{\beta}^{ru} \).

| \( N = T \) | \( r_N \) | Eigenvalue ratio | Growth ratio |
|-----------|---------|-----------------|--------------|
| 50 | 3 | 1.0000 | 1.0000 |
| 50 | 5 | 0.9952 | 1.0000 |
| 50 | 7 | 0.6347 | 1.0000 |
| 50 | 5 | 1.0000 | 1.0000 |
| 50 | 9 | 0.9996 | 1.0000 |
| 50 | 12 | 0.2679 | 1.0000 |

6. Empirical Application

We revisit the economic question of James (2015). The data are available at https://www.aeaweb.org/articles?id=10.1257/pol.20130211. This article investigates the effect on income tax of an exogenous increase of resource-based revenues (e.g., taxes paid by oil companies). The data consist of yearly nonresource and resource-based government revenues and private income for all U.S. states besides Alaska (49 subjects) between 1958 and 2008 (51 dates). The ratio of nonresource government revenues and private income is regressed on the ratio between resource-based government revenues and private income. Economic theory discussed in James (2015) predicted that state governments should react to an exogenous increase in resource-based revenues by lowering income tax.

A common factor influencing both nonresource and resource-based income may be the business cycle. When the macroeconomic output is high, energy consumption and private income both increase, leading to higher revenues (even relative to private income, because the tax scheme is usually progressive). This may create an omitted variable bias issue which can be tackled by our estimation approach. Our model allows the effect of business cycles differs by states (e.g., Hawaii is less affected by the shape of the mainland U.S. economy) and variable (because the tax schemes of nonresource and resource-based revenues differ).

The original article of James (2015) used two approaches to estimate the regression coefficient. First, a within instrument gives a value of −0.248, while an instrumental variable estimate yields −0.275. Using the eigenvalue ratio estimator of Section 4, we find \( \hat{r}_N = \hat{r}_V = 1 \) (the growth ratio estimator of Section D.3 leads to the same result). The coefficient estimate is −0.316 and the associated 95% confidence interval is [−0.698, 0.07], which contains the estimates of James (2015).

Appendix A: Proof of Lemma 2.1.

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

\[
\varphi(b) = \min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{r_{in}}, \phi_1, \ldots, \phi_T \in \mathbb{R}^{r_{t}} \atop l_1, \ldots, l_N \in \mathbb{R}^{r_{t}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Y_{it} - X_{it}^\top \hat{\beta}_i - \hat{\beta}_i^\top \hat{\phi}_i - \hat{\beta}_i^\top l_i \right)^2 .
\]

By algebra, we have

\[
\varphi(b) = \min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{r_{in}}, \phi_1, \ldots, \phi_T \in \mathbb{R}^{r_{t}} \atop l_1, \ldots, l_N \in \mathbb{R}^{r_{t}}} \times \sum_{i=1}^{N} \left| Y_{i1} - X_{i1}^\top \hat{\beta}_i - \left( \hat{\beta}_1, \ldots, \hat{\beta}_T \right)^\top l_i \right|^2 .
\]
Then, by definition of $\hat{\theta}_v$, it holds
\[
\varphi(b) = \min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{\tilde{n}_u}} \times \sum_{i=1}^{N} \left| \hat{\theta}_v \left( Y_i - X_i b - (\phi_1, \ldots, \phi_T)^\top \hat{\lambda}_i \right) \right|^2_2.
\]
Because $\hat{\theta}_v$ is symmetric, this implies
\[
\varphi(b) = \min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{\tilde{n}_u}} \times \left( Y - \sum_{k=1}^{K} b_k X_k - (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^\top (\phi_1, \ldots, \phi_T) \right)^2_2.
\]
Next, by definition of $\hat{\theta}_u$, we obtain
\[
\varphi(b) = \hat{\theta}_u \left( Y - \sum_{k=1}^{K} b_k X_k \right)^2_2.
\]
Hence, because the value of
\[
\min_{\phi_1, \ldots, \phi_T \in \mathbb{R}^{\tilde{n}_u}} \times \left( Y - \sum_{k=1}^{K} b_k X_k - (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^\top (\phi_1, \ldots, \phi_T) \right)^2_2
\]
is $\left| \hat{\theta}_u \left( Y - \sum_{k=1}^{K} b_k X_k \right) \right|^2_2$ when $(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^\top \phi_1 = (N - \hat{\lambda}_u)(Y_I - X_I b)$, we get
\[
\varphi(b) = \hat{\theta}_u \left( Y - \sum_{k=1}^{K} b_k X_k \right)^2_2.
\]

Appendix B: Proof of Theorem 3.1.

Proof of asymptotic normality:
Because $\hat{\theta}_u$ and $\hat{\theta}_v$ are symmetric, a solution to (2.1) satisfies, for $l = 1, \ldots, K$,
\[
\left( \hat{\theta}_u X_l \hat{\theta}_v, Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right) = 0,
\]

hence
\[
\left( \hat{\theta}_u X_l \hat{\theta}_v, E + \sum_{k=1}^{K} (\beta_k - \hat{\beta}_k) X_k \right)
= \left( (\hat{\theta}_u - \hat{\theta}_u) X_l \hat{\theta}_v, E + \sum_{k=1}^{K} (\beta_k - \hat{\beta}_k) X_k \right)
+ \left( \hat{\theta}_u X_l (\hat{\theta}_v - \hat{\theta}_v), E + \sum_{k=1}^{K} (\beta_k - \hat{\beta}_k) X_k \right)
+ \left( \hat{\theta}_u X_l (\hat{\theta}_v - \hat{\theta}_v), \Gamma + E + \sum_{k=1}^{K} (\beta_k - \hat{\beta}_k) X_k \right),
\]
so
\[
\sum_{k=1}^{K} (\beta_k - \hat{\beta}_k) \left( \left( \hat{\theta}_u X_l \hat{\theta}_v, X_k \right) - \left( \hat{\theta}_u - \hat{\theta}_u \right) X_l \hat{\theta}_v, X_k \right)
= \left( \hat{\theta}_u X_l \hat{\theta}_v, X_k \right) - \left( \hat{\theta}_u - \hat{\theta}_u \right) X_l \hat{\theta}_v, X_k \right).
\]

Let us show that $\left| \hat{\theta}_u X_l \hat{\theta}_v, X_k \right|$, which by Assumption 3.1 (3.1) diverges like $NT$, is the high-order term multiplying $(\beta_k - \hat{\beta}_k)$ in (B.1). This also yields the consistency of the estimator of the covariance matrix. For a matrix $M$ and $r \in \mathbb{N}$, let us define $|M|_r^2 = \sum_{k=1}^{r} \sigma_k (M)^2$. By symmetry of the projectors, (Giraud 2014, theor. C.5), and Assumption 3.2 (3.2) (which implies rank $(\hat{\theta}_u - \hat{\theta}_u) \leq 2 \sigma_i N$ w.p.a. 1), we have
\[
|\left( \hat{\theta}_u - \hat{\theta}_u \right) X_l \hat{\theta}_v, X_k |^2_2 \leq |\hat{\theta}_u - \hat{\theta}_u |^2_2 |X_l \hat{\theta}_v, X_k |^2_2 \leq \sqrt{2N} \left( \sqrt{\frac{n}{2}} + o_p(1) \right) |\hat{\theta}_u - \hat{\theta}_u |^2_2 |X_l \hat{\theta}_v, X_k |^2_2 \leq o_P \left( \sqrt{2N} \sigma_i N \right) = o_P (NT) \quad \text{(by Assumption 3.2 (3.2))}.
\]

We bound similarly $|\left( \hat{\theta}_u X_l (\hat{\theta}_v - \hat{\theta}_v), X_k \right)|$, and, for the fourth term, use that
\[
|\left( \hat{\theta}_u - \hat{\theta}_u \right) X_l \hat{\theta}_v, E |^2_2 \leq |\hat{\theta}_u - \hat{\theta}_u |^2_2 |X_l \hat{\theta}_v, E |^2_2 \leq \sqrt{2N} \left( \sqrt{\frac{n}{2}} + o_p(1) \right) |\hat{\theta}_u - \hat{\theta}_u |^2_2 |X_l \hat{\theta}_v, E |^2_2 \leq o_P \left( \sqrt{2N} \sigma_i N \right) = o_P (NT) \quad \text{(by Assumption 3.2 (3.2))}.
\]

Let us consider now the quantities on the right-hand side in (B.1). Notice that because $E = E_k = \sum_{k=1}^{K} (\beta_k - \hat{\beta}_k)^2 k$, it holds that $|E|_op = O_P (\sigma_i N)$. Proceeding like above, we have
\[
\left( \hat{\theta}_u X_l \hat{\theta}_v, E \right) = \left( \hat{\theta}_u - \hat{\theta}_u \right) X_l \hat{\theta}_v, E \leq \sqrt{2N} \left( \sqrt{\frac{n}{2}} + o_p(1) \right) |\hat{\theta}_u - \hat{\theta}_u |^2_2 \left( \sqrt{2N} \sigma_i N \right) = o_P (\sqrt{NT}) \quad \text{(by Assumption 3.2 (3.2))}.
\]
and treat similarly $|\left( \hat{\theta}_u X_l (\hat{\theta}_v - \hat{\theta}_v), E \right|$. With the same arguments as in (B.2), the absolute value of the last term of (B.1) is smaller than $u_N \sum_{k=1}^{K} \left| X_l \hat{\theta}_v, E \right|_2 \leq \sqrt{2N} \left( \sqrt{\frac{n}{2}} + o_p(1) \right) u_N |E|_op |E|_op \leq o_P \left( \sqrt{2N} u_N \sigma_i N \right) = o_P \left( \sqrt{NT} \right)$ 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 (B.1). By Assumption 3.2 (3.2), for all $k, l \in \{1, \ldots, K\}$, we have $\left( \hat{\theta}_u X_l \hat{\theta}_v, X_k \right) = \left( \hat{\theta}_u \hat{\theta}_v, X_k \right) + o_P (NT)$. Hence because of Assumption 3.1 (3.1), $\left( \hat{\theta}_u X_l \hat{\theta}_v, X_k \right)$ are the high-order terms on the left-hand side of (B.1). Similarly, by Assumption 3.1 (3.1), the high-order terms on the right-hand side of (B.1) are $\left( \hat{\theta}_u \hat{\theta}_v, E \right)$. As a result, $\hat{\beta}$ is asymptotically equivalent to the ideal estimator $\hat{\beta}$
\[
\hat{\beta} \in \text{argmin}_{\beta \in \mathbb{R}^K} \left| \hat{\theta}_u \left( Y - \sum_{k=1}^{K} \beta_k X_k \right) \right|^2_2.
\]

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

Proof of the consistency of $\hat{\sigma}$. We use
\[
NT \hat{\sigma}^2 = \left( Y - \sum_{k=1}^{K} \beta_k X_k \right) \hat{\theta}_u \left( Y - \sum_{k=1}^{K} \beta_k X_k \right) \hat{\theta}_v.
\]
Now, by the Cauchy–Schwarz inequality,
\[
\begin{align*}
\left\| Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, (\hat{M}_u - M_u) \right\|_2^2 & \leq Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, (\hat{M}_u - M_u) \left\| Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right\|_2^2 \\
& \leq Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, (\hat{M}_u - M_u) \left\| Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right\|_2^2 \\
& = o_p(\|\mu_\omega\|_\infty) \quad \text{(by the fact that } \hat{\beta} - \beta = o_p(1)) \\
& = o_p(NT) \quad \text{(by Assumption 3.2 (3.2)))}
\end{align*}
\]

Similarly, we can show that
\[
\left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, (\hat{M}_u - M_u) \left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right)_M \right) = o_p(NT)
\]

and
\[
\left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, M_u \left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right) \right) = o_p(NT).
\]

Hence, we have
\[
\begin{align*}
NT\sigma^2 & = \left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k, M_u \left( Y - \sum_{k=1}^{K} \hat{\beta}_k X_k \right)_M \right) + o_p(NT) \\
& = \left( Y - \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k - \sum_{k=1}^{K} \hat{\beta}_k X_k, M_u \right) \\
& \times \left( Y - \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k - \sum_{k=1}^{K} \hat{\beta}_k X_k, M_u \right)_M + o_p(NT) \\
& = \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k, M_u \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k \right)_M \right) \\
& \quad + \left( E_0, M_u \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k \right)_M \right) \\
& \quad + \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k, M_u E_K M_v \right) + \| \hat{E} \|_2^2 + o_p(NT).
\end{align*}
\]

Now, by the Cauchy–Schwarz inequality, Assumption 3.2 and the fact that $\hat{\beta} - \beta = o_p(1)$, one can show that
\[
\begin{align*}
\left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k, M_u \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k \right)_M \right) & = o_p(NT) \\
\left( E, M_u \left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k \right)_M \right) & = o_p(NT); \\
\left( \sum_{k=1}^{K} (\hat{\beta}_k - \beta_k) X_k, M_u E K M_v \right) & = o_p(NT).
\end{align*}
\]

We conclude the proof using Assumption 3.1.

Appendix C: Proof of results of Sections 4.2 and 4.3

C.1. Proof of Lemma 4.3.

We only prove the result for $\Pi_u$, the proof for $\Pi_v$ being similar. For $t \in \{1, \ldots, T\}$ and $k \in \{0, \ldots, K\}$, we use the notation $\psi_{tk} = (\psi_{t1}, \ldots, \psi_{tr_u} \delta_{xr_u})^\top = \psi_{t1}, \ldots, \psi_{tr_u}, \psi_{r_u K}. We also introduce $\Psi = (\psi_{10}, \ldots, \psi_{t0}, \ldots, \psi_{1K}, \ldots, \psi_{TK}). It holds that, for $j, j' \in \{1, \ldots, T\}$, $
\left( \psi_j \psi_j^\top \right)_{jj'}/T = \left( \Delta \psi_j \psi_j^\top /T \right)_{jj'}/T
\]

converges in probability to $\Sigma_\lambda \Sigma_{\Delta \Psi} \psi_j \psi_j^\top/ \Delta \Psi \psi_j^\top \lambda$. Next, let $U = (u_1, \ldots, u_{tr_u}, \Pi_u), \quad V = (v_1, \ldots, v_{tr_u}, \Pi_u)$ and $D$ be the $r_u \times r_u$ diagonal matrix for which $D_{jj} = \sigma_j (\Pi_u). We have $UDV^\top = \Lambda^\top \psi_j^\top \lambda U$, which implies $UDV^\top = \Delta \psi_j^\top \lambda U$. Hence, the $r_u \times r_u$ columns of $\Delta \Psi$ are eigenvectors of $\lambda^\top \psi_j^\top \lambda U$. Indeed, $\lambda U$ is a $r_u \times r_u$ matrix and for $x \in \mathbb{R}^{r_u}$, $\lambda U x \neq 0$ because the columns of $U$ (and, therefore, $Ux$) belong to the vector space generated by the rows of $\lambda$. Hence, the columns of $\Delta \Psi$ are linearly independent and, therefore, the diagonal elements of $\psi_j \psi_j^\top / \Delta \Psi \psi_j^\top \lambda$ (which has rank at most than $r_u$). Then, for $j \in \{1, \ldots, r_u\}$, we have
\[
\begin{align*}
\mathbb{P} \left( \left| \sigma_j \left( \Delta \psi_j \psi_j^\top \right)_{jj'/T} \right| > \xi \right) & \leq \mathbb{P} \left( \left| \lambda \psi_j \psi_j^\top \lambda U \right| > \xi \right) \\
& \leq \mathbb{P} \left( \left| \lambda \psi_j \psi_j^\top \lambda U \right| > \xi \right),
\end{align*}
\]

where the last inequality is due to Weyl’s inequality (Theorem C.6 in Giraud 2014). The result follows directly.

C.2. Proof of Lemma 4.4.

We only prove the result for $\Pi_u$, the proof for $\Pi_v$ being similar. We use the notations of the proof of Lemma 4.3. Let also $\Sigma_{\Delta \Psi} = \psi_j^\top \Delta \Psi / \psi_j^\top \lambda$. In the same manner as in the proof of Lemma 4.3, for $j \in \{1, \ldots, r_u\}$ and $\xi > 0$, we have
\[
\begin{align*}
\mathbb{P} \left( \max_{j \in \{1, \ldots, r_u\}} \left| \sigma_j \left( \psi_j^\top \psi_j \right)/w_{\psi_j} \right| > \xi \right) & \leq \mathbb{P} \left( \max_{j \in \{1, \ldots, r_u\}} \left| \sigma_j \left( \psi_j^\top \psi_j \right)/w_{\psi_j} \right| > \xi \right),
\end{align*}
\]

where the last inequality is due to Weyl’s inequality (Theorem C.6 in Giraud 2014). The result follows directly.
Therefore, to prove that \( \sigma_j \left( \frac{D_j^2}{w_N N T} \right) \) converges uniformly in \( j \) to \( \sigma_j \left( \Sigma_{\Lambda N} \Sigma_{\Delta F N} \right) \), it suffices to show that

\[
P \left( \left\| \frac{\Lambda A^T \Psi \Psi^T}{w_N N T} - \Sigma_{\Lambda N} \Sigma_{\Delta F N} \right\|_2 > \xi \right) \to 0.
\]

We have

\[
\left\| \frac{\Lambda A^T \Psi \Psi^T}{w_N N T} - \Sigma_{\Lambda N} \Sigma_{\Delta F N} \right\|_2 \leq \left\| \frac{\Lambda A^T}{N} - \Sigma_{\Lambda N} \right\|_2 \left\| \Sigma_{\Delta F N} \right\|_{\text{op}} + \left| \Sigma_{\Lambda N} \right| \left\| \Psi \Psi^T \right\|_2 \to 0,
\]

Remark that \( \left\| \Sigma_{\Delta F N} \right\|_{\text{op}} \leq \left\| w_N^{-1} \Lambda A^T \right\|_{\infty} \left\| \Sigma_{\Delta F N} \right\|_{\text{op}} = O(1) \) by Horn and Johnson (2012, exer. 7.5.P24 (b)) and

\[
\left\| \frac{\Psi \Psi^T}{w_N N T} - \Sigma_{\Delta F N} \right\|_2 \leq \sum_{j' = 1}^{N N} \left( \frac{\Lambda A^T}{w_N} \right)^2 \left( \frac{FF^T}{T} \right) \left\| \Psi \Psi^T \right\|_2 \leq \left\| \frac{\Psi \Psi^T}{w_N N T} - \Sigma_{\Delta F N} \right\|_2 \to 0.
\]

Hence, the right-hand side of (C.1) is an \( O(1) \) and we obtain that

\[
P \left( \max_{j \in \{1, \ldots, N N\}} \left| \sigma_j \left( \frac{D_j^2}{w_N N T} \right) - \sigma_j \left( \Sigma_{\Lambda N} \Sigma_{\Delta F N} \right) \right| > \xi \right) \to 0.
\]

To conclude the proof, just notice that \( \sigma_j \left( \Sigma_{\Lambda N} \Sigma_{\Delta F N} \right) \leq |\Sigma_{\Lambda N}|_{\text{op}} |\Sigma_{\Delta F N}|_{\text{op}} = O(1) \).

**Appendix D: Results on PCA**

**D.1. Some general results**

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

**Lemma D.1.** \( \tilde{A}_s - A \to_{\text{op}} \tilde{A} - A \).

**Proof.** We have \( \left| \tilde{A}_s - A \right|_{\text{op}} = \left| \tilde{A}_s - \tilde{A} + \tilde{A} - A \right|_{\text{op}} \leq \sum_{j=r+1}^{N} \tilde{\sigma}_j \left| \tilde{u}_j \tilde{v}_j^T \right|_{\text{op}} + \left| \tilde{Z}_s \right|_{\text{op}} \leq \sigma_{r+1} \tilde{Z}_s \to_{\text{op}} \tilde{Z}_{r+1} + \tilde{Z} \). Now, by Weyl's inequality (Theorem C.6 in Giraud (2014)), it holds that \( \tilde{\sigma}_{r+1} \leq \tilde{A} - A \to_{\text{op}} \tilde{Z} \). □

**Lemma D.2.** We have \( \left| \tilde{P}_r - P \right|_2 \leq 4 \sqrt{2} \left| \tilde{Z} \right|_{\text{op}} / \sigma_{r+1} \) almost surely.

**Proof.** Following the proof of Beyhum and Gautier (2019, prop. 10), we obtain \( P_{r+1}^{-1} \leq 2 \tilde{P}_r \) / \( 2 \tilde{Z}_r \). We conclude using

\[
\left| \tilde{M}_r A \right|_2 \leq \left| \tilde{M}_r (\tilde{A} - A) \right|_2 \leq \sqrt{2r} \tilde{A} - A \to_{\text{op}} \tilde{Z}_r \leq \sqrt{2r} \left| \tilde{Z} \right|_{\text{op}},
\]

by Lemma D.1 and the fact that \( \tilde{M}_r \) is a projector. □

**Lemma D.3.** The following holds:

i. For \( j \in \{1, \ldots, r\} \), \( \sigma_j - \tilde{\sigma}_j \leq \sigma_j + |\tilde{A} - A|_{\text{op}} \).

ii. For \( j \in \{r+1, \ldots, N\} \), \( \sigma_j - \tilde{\sigma}_j \leq \sigma_j + |\tilde{A} - A|_{\text{op}} \).

**Proof.** (D.3) follows from the fact that \( |\tilde{\sigma}_j - \sigma_j| \leq |\tilde{A} - A| \) \( \to_{\text{op}} \) \( \tilde{Z} \to_{\text{op}} \). Weyl's inequality also yields \( \sigma_j \leq \tilde{A} - A \to_{\text{op}} \) \( \tilde{Z} \to_{\text{op}} \). This implies the right-hand side of (D.3). To show the left-hand side of (D.3), from (7.3.13) in Horn and Johnson (2012), we obtain \( \sigma_j \to_{\text{op}} \tilde{Z} \).

**Lemma D.4.** Let \( Z \) be a \( K_1 N \times K_2 T \) random matrix, with \( K_1, K_2 \in \mathbb{N} \) and \( \nu_N = \text{op}(\sqrt{N} \vee T) \). Assume that \( |\tilde{Z}|_{\text{op}} = O \left( \sqrt{N \vee T} \right) \) and there exists \( \nu > 0 \) such that \( |\tilde{Z}|_{\text{op}} / (NT) \to_{\text{p}} \nu^2 \). Then, we have \( \frac{\nu}{\sqrt{NT}} \to_{\text{w.p.a. 1}} \) and \( \max_{j \in \{1, \ldots, \sqrt{NT} \}} |\tilde{Z}|_{\text{op}} / \sigma_{r+j}(Z) = O_p(1) \).

**Proof.** We have

\[
\frac{|\tilde{Z}|_{\text{op}}^2}{NT} \leq \frac{2}{NT} \frac{\nu}{\sqrt{NT}} \left( \frac{\nu}{\sqrt{NT}} \right)^2 = \frac{2}{NT} \left( \frac{\nu}{\sqrt{NT}} \right)^2 \to_{\text{p}} \nu^2
\]

and, therefore,

\[
P \left( \frac{|\tilde{Z}|_{\text{op}}^2}{\nu/\sqrt{NT}} \geq \frac{\nu}{2} \right) \to_{\text{p}} 1.
\]

Hence, we have

\[
P \left( \frac{|\tilde{Z}|_{\text{op}}^2}{\nu/\sqrt{NT}} \geq \frac{\nu}{2} \right) \to_{\text{p}} 1.
\]

Therefore, we obtain

\[
\frac{|\tilde{Z}|_{\text{op}}^2}{\nu/\sqrt{NT}} = O_p \left( \frac{|\tilde{Z}|_{\text{op}}^2}{\nu/\sqrt{NT}} \right) = O_p(1).
\]

This leads to

\[
\max_{j \in \{1, \ldots, \sqrt{NT} \}} \frac{|\tilde{Z}|_{\text{op}}}{\sigma_{r+j}(Z)} \leq \frac{|\tilde{Z}|_{\text{op}}}{\sigma_{r+j}(Z)} = O_p(1).
\]

□
D.2. Proof of results of Section 4

Proof of Lemma 4.1.

Because $\sqrt{\mathcal{N} \vee T} = O(z_N)$, it holds that $|E_{\log}| = O_p(\sigma_\ell(P_z))$. Then, we have $\sigma_{\ell r+1}(P_z) \leq |E_{\log}| = O_p\left(\sqrt{\mathcal{N} \vee T}\right)$ which implies $z_N/\sqrt{\mathcal{N} \vee T} = O_p\left(\sigma_\ell(P_z)/\sigma_{\ell r+1}(P_z)\right)$. Moreover, by Lemma D.4, we have $\max_{j \in \{1, \ldots, r_z-1\}} \left|E_{\log}\right|/\sigma_{\ell r+1}(P_z) = O_p(1)$. Because

$$\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(P_z)}{\sigma_{\ell r+1}(P_z)} = \sigma \left(z_N/\sqrt{\mathcal{N} \vee T}\right),$$

we obtain

$$P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(P_z)}{\sigma_{\ell r+1}(P_z)} \leq C \frac{\sigma_{\ell r}(P_z)}{\sigma_{\ell r+1}(P_z)} \Rightarrow 1.\right)$$

Proof of Lemma 4.2.

To prove Lemma 4.2, let us show that $P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{\ell r}(P_z)/\sigma_{\ell r+1}(P_z)} \Rightarrow 1.\right)$, where $j \in \{1, \ldots, r_z-1\}$, by Lemma D.3 (D.3) and Lemma D.1, we have $\sigma(Y_z) - 2 |E_{\log}| \leq \sigma_j(Y_z) \leq \sigma_j(Y_z) + 2 |E_{\log}| . Then, on the event $A = \{ \sigma_j(P_z) > 2 |E_{\log}| \}$, we obtain

$$\frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \leq \frac{\sigma_j(Y_z) + 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) - 2 |E_{\log}|} = \frac{\sigma_j(Y_z) + 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) - 2 |E_{\log}|},$$

$$\leq \frac{\sigma_j(Y_z) + 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) - 2 |E_{\log}|} \leq \frac{\sigma_j(Y_z) + 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) - 2 |E_{\log}|} \leq \frac{\sigma_j(Y_z) + 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) - 2 |E_{\log}|},$$

(D.1)

where the last equality is because $\sigma_j(P_z) \geq \sigma_{\ell r+1}(P_z) \geq \sigma_{\ell r}(P_z)$. Also, by Lemma D.3, on $A$, it holds that

$$\frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \geq \frac{\sigma_j(Y_z) - 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) + 2 |E_{\log}|} = \frac{\sigma_j(Y_z) - 2 |E_{\log}|}{\sigma_{\ell r+1}(P_z) + 2 |E_{\log}|},$$

(D.2)

Let us call $B$ the event

$$\left\{1 - 2 |E_{\log}|/\sigma_{\ell r}(P_z) \geq C, \right\},$$

where $C$ is the constant in Assumption 4.1. We have

$$P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \leq C, \right\} \geq P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \leq C, \right\} \cap A \cap B$$

$$\geq P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \leq C, \right\} \cap A \cap B \\\ (by \ D.1 \ and \ D.2)$$

$$\geq P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(Y_z)}{\sigma_{\ell r+1}(P_z)} \leq C, \right\} \cap A \cap B \Rightarrow 1,$$

where the last statement holds because $P(A) \rightarrow 1$, $P(B) \rightarrow 1$ (given that $|E_{\log}| = O_p(\sigma_\ell(P_z))$)

$$P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{\sigma_j(P_z)}{\sigma_{\ell r+1}(P_z)} \leq C, \right\} \rightarrow 1$$

by Assumption 4.1. Next, let us show that, $P\left(\max_{j \in \{1, \ldots, r_z-1\}} \frac{|E_{\log}|}{\sigma_{\ell r+1}(P_z)} \leq |E_{\log}|/\sigma_{\ell r+1}(P_z) \right) \Rightarrow 1$. By Lemma D.3 (D.3), we have, for all $j > r_z$,

$$\frac{|E_{\log}|}{\sigma_{\ell r+1}(P_z)} \leq |E_{\log}|/\sigma_{\ell r+1}(P_z) \leq 1.$$
Supplementary Materials

The supplementary material contains additional results discussed in the article and the MATLAB code used for the simulations.

Acknowledgments

The authors thank Domenico Giannone, Jihyun Kim, Pascal Lavergne, Thierry Magnac and Nour Meddahi, an associate editor and two referees of the Journal of Business & Economic Statistics for helpful comments and ideas.

Funding

Financial support from the European Research Council (ERC grant POEMH 337665) and from the French National Research Agency (ANR) (grant ANR-17-EURE-0010) is gratefully acknowledged.

References

Ahn, S. C., and Horenstein, A. R. (2013), “Eigenvalue Ratio Test for the Number of Factors,” Econometrica, 81, 1203–1227. [273, 274, 280] Bai, J. (2009), “Panel Data Models With Interactive Fixed Effects,” Econometrica, 77, 1229–1279. [271, 273, 274, 275] Bai, J., and Wang, P. (2016), “Econometric Analysis of Large Factor Models,” Annual Review of Economics, 8, 53–80. [271] Beyhum, J., and Gautier, E. (2019), “Square-Root Nuclear Norm Penalized Estimator for Panel Data Models With Approximately Low-Rank Unobserved Heterogeneity,” arXiv:1904.09192. [271, 273, 279] Carroll, J. D., and Chang, J-J. (1970), “Analysis of Individual Differences in Multidimensional Scaling Via an n-way Generalization of Eckart-Young Decomposition,” Psychometrika, 35, 283–319. [271] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018), “Double/Debiased Machine Learning for Treatment and Structural Parameters,” The Econometrics Journal, 21, C1–C68. [271, 272] Chudik, A., Pesaran, M. H., and Tosetti, E. (2011), “Weak and Strong Cross-Section Dependence and Estimation of Large Panels,” Econometrics Journal, 14. [271] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000), “A Multilinear Singular Value Decomposition,” SIAM Journal on Matrix Analysis and Applications, 21, 1253–1278. [272] Gagliardi, P., and Gourieroux, C. (2017), “Double Instrumental Variable Estimation of Interaction Models With Big Data,” Journal of Econometrics, 201, 176–197. [273] Giraud, C. (2014), Introduction to High-Dimensional Statistics, London, UK: Chapman and Hall/CRC. [277, 278, 279] Greenaway-McGrevy, R., Han, C., and Sul, D. (2012), “Asymptotic Distribution of Factor Augmented Estimators for Panel Regression,” Journal of Econometrics, 169, 48–53. [270, 271, 272, 273, 274, 275] Harshman, R. A. (1970), “Foundations of the Parafac Procedure: Models and Conditions for an ‘Explanatory’, Multifactor Model Analysis,” UCLA Working Papers in Phonetics, 16, 1–84. Available at https://www.psychology.uwo.ca/faculty/harshman/wpppfac0.pdf [271] Horn, R. A., and Johnson, C. R. (2012), Matrix Analysis, Cambridge: Cambridge University Press. [279] James, A. (2015), “U.S. State Fiscal Policy and Natural Resources,” American Economic Journal: Economic Policy, 7, 238–57. [276] Kapetanios, G., and Pesaran, M. H. (2005), “Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results With an Application to Modelling of Asset Returns,” Technical report, CEISifo working paper. [271] Kolda, T. G., and Bader, B. W. (2009), “Tensor Decompositions and Applications,” SIAM Review, 51, 455–500. [271, 272] Moon, H. R., and Weidner, M. (2015), “Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects,” Econometrica, 83, 1543–1579. [273, 274, 275, 276] Moon, H. R., and Weidner, M. (2017), “Dynamic Linear Panel Regression Models With Interactive Fixed Effects,” Econometric Theory, 33, 158–195. [273] ——— (2018), “Nuclear Norm Regularized Estimation of Panel Regression Models,” arXiv:1810.10987. [271] Pesaran, M. H. (2006), “Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure,” Econometrica, 74, 967–1012. [271] ——— (2015), Time Series and Panel Data Econometrics, Oxford: Oxford University Press. [270] Reese, S., and Westerlund, J. (2018), “Estimation of Factor-Augmented Panel Regressions With Weakly Influential Factors,” Econometric Reviews, 37, 401–465. [271]