Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: www.tandfonline.com/journals/lecr20

A projection-based approach for interactive fixed effects panel data models

Georg Keilbar, Juan M. Rodriguez-Poo, Alexandra Soberón & Weining Wang

To cite this article: Georg Keilbar, Juan M. Rodriguez-Poo, Alexandra Soberón & Weining Wang (2026) A projection-based approach for interactive fixed effects panel data models, Econometric Reviews, 45:1, 93-110, DOI: 10.1080/07474938.2025.2556702

To link to this article: https://doi.org/10.1080/07474938.2025.2556702

ECONOMETRIC REVIEWS
2026, VOL. 45, NO. 1, 93–110
https://doi.org/10.1080/07474938.2025.2556702

A projection-based approach for interactive fixed effects panel data models

Georg Keilbar[a], Juan M. Rodriguez-Poo[b], Alexandra Soberón[b], and Weining Wang[c]

[a]Humboldt-Universität zu Berlin, Berlin, Germany; [b]Department of Economics, University of Cantabria, Santander, Spain; [c]School of Economics, University of Bristol, Bristol, UK

ABSTRACT
This article introduces a straightforward sieve-based approach for estimation and inference of regression parameters in panel data models with interactive fixed effects. The method's key assumption is that factor loadings can be decomposed into an unknown smooth function of individual characteristics plus an idiosyncratic error term. Our estimator offers advantages over existing approaches by taking a simple partial least squares form, eliminating the need for iterative procedures or preliminary factor estimation. The limiting distribution exhibits a discontinuity that depends on how well our basis functions explain the factor loadings, as measured by the variance of the error factor loadings. As a consequence, conventional "plug-in" methods using the estimated asymptotic covariance can produce excessively conservative coverage probabilities. We demonstrate that uniformly valid non conservative inference can be achieved through the cross-sectional bootstrap method. Monte Carlo simulations confirm the estimator's strong performance in terms of mean squared error and good coverage results for the bootstrap procedure. An application to cross-country growth rates shows that higher consumption and government spending are associated with lower growth. Contrary to existing methods, we find that within OECD countries investment fosters growth, whereas a higher investment price level reduces it.

ARTICLE HISTORY
Received 24 February 2025
Accepted 27 August 2025

KEYWORDS
Cross-sectional dependence; large panels; principal components; semiparametric factor models; sieve approximation

1. Introduction

This article considers panel data models with interactive fixed effects, which are widely used to capture unobserved heterogeneity and cross-sectional dependence (CSD). These models assume that the error term, v_{it}, follows a latent factor structure of the form $v_{it} = \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + u_{it}$, where $\boldsymbol{\lambda}_i$ are individual-specific factor loadings and $\boldsymbol{f}_t$ are time-varying common factors. These models are particularly relevant when latent global shocks (i.e., financial conditions, technology diffusion, or geopolitical events) affect all individuals with heterogeneous intensities.

In macroeconomics, for instance, fluctuations in global trade or capital markets can induce strong CSD in cross-country growth regressions (Chudik et al., 2017; Lu and Su, 2016). In finance, unobserved risk or liquidity conditions affect asset returns beyond what is captured by standard pricing factors, such as those in the Fama-French model (e.g., small market capitalization and book-to-market ratios) (Bernanke, Boivin, and Eliasz, 2005; Fan, Ke, and Liao, 2021). Ignoring such latent structures can yield biased and inconsistent estimators of the structural parameters.

CONTACT Georg Keilbar georg.keilbar@hu-berlin.de Humboldt-Universität zu Berlin, Berlin, Germany.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07474938.2025.2556702.

To address the endogeneity problem arising from the correlation between covariates and latent components (i.e., $\boldsymbol{f}_t$ and/or $\boldsymbol{\lambda}_i$) in large panels (e.g., where both the cross-section, N, and the time dimension, T, are large), the literature has evolved along two primary methodological paths: (i) controlling for unobserved common factors $\boldsymbol{f}_t$ while treating factor loadings $\boldsymbol{\lambda}_i$ as nuisance parameters, and (ii) modeling loadings $\boldsymbol{\lambda}_i$ as smooth functions of some observed characteristics.

The first line of research includes influential approaches such as the common correlated effects (CCE) estimator of Pesaran (2006), which addresses the presence of unobserved common factors by augmenting the model with cross-sectional averages of both regressors and dependent variables. Another prominent approach is the principal component (PC) estimator, introduced by Bai (2009) and further refined by Moon and Weidner (2015, 2017). These PC-based techniques consistently estimate both latent factors and loadings and recover structural parameters by solving a non convex optimization problem. This first methodological strand has fostered the development of a rich and widely applicable literature with numerous extensions (see Sarafidis and Wansbeek (2012), Chudik and Pesaran (2015), Bai and Wang (2016), or Westerlund and Urbain (2015), among others). Despite their popularity, a key drawback is the agnostic treatment of loadings, which may lead to a loss of estimation efficiency, particularly in settings such as asset pricing, where firm-level attributes (such as size, leverage, profitability, or industry classification) are often informative about factor exposure.

Motivated by this empirical insight, the second strand of the literature addresses the endogeneity issue by modeling factor loadings $\boldsymbol{\lambda}_i$ as smooth functions of observable unit-level covariates. Notable contributions include Connor and Linton (2007), Desai and Storey (2012), Ma, Linton, and Gao (2021), and Cheng et al. (2024), who specify $\boldsymbol{\lambda}_i = \boldsymbol{g}(\boldsymbol{Z}_i)$, where $\boldsymbol{g}(\cdot)$ is an unknown smooth function and $\boldsymbol{Z}_i$ contains unit-level characteristics. However, this modeling approach imposes the restrictive assumption that the entire variation in $\boldsymbol{\lambda}_i$ must be explained by $\boldsymbol{Z}_i$, which increases the risk of model misspecification. To allow partial flexibility and mitigate the risk of misspecification, Fan, Liao, and Wang (2016) consider a pure factor model and propose a more flexible framework that decomposes factor loading into a systematic component and an idiosyncratic error, that is, $\boldsymbol{\lambda}_i = \boldsymbol{g}(\boldsymbol{Z}_i) + \boldsymbol{\gamma}_i$, where $\boldsymbol{\gamma}_i$ reflects the part of $\boldsymbol{\lambda}_i$ that cannot be explained by $\boldsymbol{Z}_i$. They propose the projected Principal Component Analysis (projected-PCA) method, which improves upon standard PCA by first projecting the data onto a sieve space defined by the basis functions of covariates.

This article contributes to the literature by developing a tractable methodology that extends the projected-PCA approach of Fan, Liao, and Wang (2016) to a more general panel data regression setting in which the factor loadings are characterized by a semiparametric structure,

$$\begin{cases} y_{it} &= \boldsymbol{X}_{it}^\top \boldsymbol{\beta} + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + u_{it}, \\ \boldsymbol{\lambda}_i &= \boldsymbol{g}(\boldsymbol{Z}_i) + \boldsymbol{\gamma}_i, \end{cases} \qquad i = 1, \dots, N; \quad t = 1, \dots, T, \tag{1}$$

where y_{it} is the response variable of individual i at time t, $\boldsymbol{X}_{it}$ is a Q-dimensional vector of covariates, $\boldsymbol{\beta}$ is the Q-dimensional vector of parameters to be estimated, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{iK})^\top$ and $\boldsymbol{f}_t = (f_{t1}, \dots, f_{tK})^\top$ are K-dimensional vectors of factor loadings and common factors, respectively, and u_{it} is the idiosyncratic error term which is assumed to have zero mean and to be independent of both covariates and factor structure. Furthermore, $\boldsymbol{Z}_i$ is a D-dimensional vector of additional covariates representing individual characteristics, $\boldsymbol{g}(\boldsymbol{Z}_i) = (g_1(\boldsymbol{Z}_i), \dots, g_K(\boldsymbol{Z}_i))^\top$ is a K-dimensional vector of unknown functions, and $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iK})^\top$ is a K-dimensional vector of errors. Throughout the article, we assume that $\{\boldsymbol{\gamma}_i\}_{i \le N}$ has zero mean and is independent of $\{\boldsymbol{Z}_i\}_{i \le N}$, and the number of factors K is finite and does not depend on the size of the cross-section N or the time dimension T.

Integrating all the information contained in (1), the resulting regression model can be written as follows

$$y_{it} = \boldsymbol{X}_{it}^\top \boldsymbol{\beta} + \boldsymbol{g}(\boldsymbol{Z}_i)^\top \boldsymbol{f}_t + \boldsymbol{\gamma}_i^\top \boldsymbol{f}_t + u_{it}, \qquad i = 1, \dots, N; \quad t = 1, \dots, T. \tag{2}$$

Building upon the core idea of Fan, Liao, and Wang (2016), the innovation of this article lies in developing a direct estimation procedure for $\boldsymbol{\beta}$ that relies on projecting the data onto the subspace generated by the sieve basis functions of the covariates $\boldsymbol{Z}_i$. This orthogonal projection intends to "project away" the unobserved factor loadings to eliminate asymptotically the bias related to the loadings and obtain consistent and asymptotically normal estimators of the $\boldsymbol{\beta}$'s in (2) without the need for computationally intensive procedures.

The estimation procedure proposed in this article offers several advantages. First, it is based on a simple OLS framework, which avoids the complexities of iterative procedures. Second, it does not require prior knowledge of the number of common factors and does not require knowledge or assumptions about them, making it robust to various specifications. In addition, the underlying limiting distribution is centered at zero. Finally, the proposed estimator reaches the semiparametric efficiency bound under certain conditions.

In deriving the asymptotic properties of the proposed estimator, we obtain a rate of convergence that crucially depends on the persistence of the $\boldsymbol{\gamma}_i$ terms in the composite error. If the variance of these idiosyncratic factor loadings is zero or approaches zero with a rate $\mathcal{O}(T^{-1})$, our estimator is $\sqrt{NT}$-consistent. Otherwise, the rate of convergence will be between $\sqrt{NT}$ and the worst case rate $\sqrt{N}$ which occurs if the variance of $\boldsymbol{\gamma}_i$ is of order $\mathcal{O}(1)$. Although we obtain asymptotic normality results for all scenarios, it is true that the case of vanishing factor loadings is the one most favorable to our estimator compared with existing approaches for dealing with interactive fixed effects.

The idiosyncratic factor loading term $\boldsymbol{\gamma}_i$ has also a profound impact on inference, as it introduces a discontinuity in the limiting distribution when its variance is near the boundary. In this case, the usual "plug-in" approaches would lead to valid but overly conservative inference. Similarly, ignoring the idiosyncratic part leads to invalid inference in the case of persistent variance. To achieve uniformly valid but non conservative inference, we resort to the cross-sectional bootstrap originally proposed by Kapetanios (2008). In this way, by stacking the time observations we are able to mimic the asymptotic distribution and conduct uniformly valid inference even in the presence of this type of discontinuities as shown by Liao and Yang (2018) and Fernández-Val et al. (2022). The issue of uniformity is an important topic for modeling panel data. Lu and Su (2023) consider a model with two-dimensional heterogeneity of varying degrees in the slope parameters and are interested in uniformly valid inference. Kock (2016) and Kock and Tang (2019) study uniform inference in high-dimensional panel regression contexts. Menzel (2021) showed that uniform non conservative inference is impossible under general dependence in more than one dimension. The novelty of our bootstrap procedure is that we resample cross-sectional units after projecting the data, i.e., partialing-out the modeled part of the factor loadings.

Our model setup is closely related to the one in Zhang, Zhou, and Wang (2021); however, we want to point out crucial distinctions. First and foremost, the main issue of interest in the above article is efficiency and they propose a GLS-type estimator that under broadly general conditions is oracle efficient. It is important to note that their asymptotic results require the consistency of the pooled OLS estimator in a first step which is not the case in our model setup. Second, Zhang, Zhou, and Wang (2021) assume that the factor loadings are fully explained by $\boldsymbol{Z}_i$, i.e., $\boldsymbol{\gamma}_i = 0$. Unfortunately, in the presence of error factor loadings, i.e., $\boldsymbol{\gamma}_i \neq 0$, the statistical properties of standard estimators for $\boldsymbol{\beta}$ remain unclear. Therefore, it is of interest to derive a new estimator to obtain consistency and asymptotic rates.

The rest of the article is organized as follows. In Section 2, we derive our projection-based interactive fixed effects estimator. Section 3 states our assumptions and studies the asymptotic properties of the proposed estimators. In Section 4, we validate the theoretical results in a simulation study. In Section 5, we apply our method to the identification of the determinants of economic growth. Lu and Su (2016) argued that the GDP growth rates per capita might not only be determined by observed factors but might also be influenced by latent factors or shocks. Our projection-based interactive fixed effects estimator is well-suited for such a setting. All proofs of the asymptotic results and further Monte Carlo results are relegated to a Supplementary Material document.

2. Estimation procedure

To nonparametrically estimate the unknown function $g_k(\cdot)$ without curse of dimensionality, it will be assumed that for each k, where $k = 1, \ldots, K$, $g_k(\cdot)$ is an additive function of the form

$$g_k(\boldsymbol{Z}_i) = \sum_{d=1}^{D} g_{kd}(Z_{id}), \quad i = 1, \ldots, N, \quad k = 1, \ldots, K. \tag{3}$$

For each k and d, the additive component $g_{kd}(\cdot)$ can be approximated by the sieve method. We define $\left\{\phi_1(Z_{id}), \ldots, \phi_{J_N}(Z_{id})\right\}$ as a set of basis functions (i.e., splines, Fourier series, wavelets), which spans a dense linear space of the functional space for $g_{kd}(\cdot)$. Then,

$$g_{kd}(Z_{id}) = \sum_{j=1}^{J_N} b_{j,kd}\phi_j(Z_{id}) + R_{kd}(Z_{id}), \quad k = 1, \ldots, K, \quad d = 1, \ldots, D, \tag{4}$$

where, for $j = 1, \ldots, J_N$, $\phi_j(\cdot)$'s are the sieve basis functions, $b_{j,kd}$'s are the sieve coefficients of the dth additive component of $g_k(\boldsymbol{Z}_i)$ corresponding to the kth factor loading, $R_{kd}(\cdot)$ is a "remainder function" that represents the approximation error, and J_N denotes the number of sieve terms which grows slowly as $N \to \infty$.

As it is well-known in the literature, under some regularity condition of the functional class, the approximation functions $\phi_j(\cdot)$ have the property that, as J_N grows, there is a linear combination of $\phi_j(\cdot)$ that can approximate $g_k(\cdot)$ arbitrarily well in the sense that the approximation error can be made arbitrarily small. Therefore, for a given $d = 1, \ldots, D$, the basic assumption for sieve approximation is that $\sup_z |R_{kd}(z)| \to 0$, as $J_N \to \infty$. In practice, an optimal choice for the smoothing parameter J_N can be based on cross-validation.

For the sake of simplicity, we take the same basis functions in (4) and, for each $k \leq K$, $d \leq D$ and $i \leq N$, let us define

$$\boldsymbol{b}_k^\top = (b_{1,k1}, \ldots, b_{J_N,k1}, \ldots, b_{1,kD}, \ldots, b_{J_N,kD}) \in \mathbb{R}^{J_N D},$$
$$\phi(\boldsymbol{Z}_i)^\top = (\phi_1(Z_{i1}), \ldots, \phi_{J_N}(Z_{i1}), \ldots, \phi_1(Z_{iD}), \ldots, \phi_{J_N}(Z_{iD})) \in \mathbb{R}^{J_N D},$$

so the above equation can be rewritten as

$$g_k(\boldsymbol{Z}_i) = \phi(\boldsymbol{Z}_i)^\top \boldsymbol{b}_k + \sum_{d=1}^{D} R_{kd}(Z_{id}). \tag{5}$$

Let $\boldsymbol{Z} = \left(\boldsymbol{Z}_1^\top, \ldots, \boldsymbol{Z}_N^\top\right)$ be an $N \times D$ matrix whose ith element is a D-dimensional vector of random variables as $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{iD})^\top$ and denote by $\boldsymbol{\mathcal{Z}}$ its support. Let also $\boldsymbol{G}(\boldsymbol{Z})$ be an $N \times K$ matrix of unknown functions, $g_k(\boldsymbol{Z}_i)$, $\Phi(\boldsymbol{Z}) = (\phi(\boldsymbol{Z}_1), \ldots, \phi(\boldsymbol{Z}_N))^\top$ be an $N \times J_N D$ matrix of basis functions, $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K)$ be an $J_N D \times K$ matrix of sieve coefficients, and $\boldsymbol{R}(\boldsymbol{Z})$ be an $N \times K$ matrix with the (i, k)th element $\sum_{d=1}^{D} R_{kd}(Z_{id})$. By considering (5) in matrix form, we obtain

$$\boldsymbol{G}(\boldsymbol{Z}) = \Phi(\boldsymbol{Z})\boldsymbol{B} + \boldsymbol{R}(\boldsymbol{Z}), \tag{6}$$

and substituting (6) into the matrix form of (2) leads to

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \Phi(\boldsymbol{Z})\boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{R}(\boldsymbol{Z})\boldsymbol{f}_t + \boldsymbol{v}_t, \qquad t = 1, \ldots, T, \tag{7}$$

where the residual term consists of two parts: the sieve approximation error, $\boldsymbol{R}(\boldsymbol{Z})\boldsymbol{f}_t$, and the error term, $\boldsymbol{v}_t$, that is an $N \times 1$ vector such as $\boldsymbol{v}_t = \boldsymbol{\Gamma}\boldsymbol{f}_t + \boldsymbol{u}_t$, where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N)^\top$ is an $N \times K$ matrix of unknown loading coefficients and $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})^\top$ is an $N \times 1$ vector of idiosyncratic errors. Finally, $\boldsymbol{X}_t$ is an $N \times Q$ matrix of covariates.

To obtain consistent estimators of $\boldsymbol{\beta}$ in (7) we propose a transformation that removes $\Phi(\boldsymbol{Z})\boldsymbol{B}\boldsymbol{f}_t$ and accounts for the error term, $\boldsymbol{\Gamma}\boldsymbol{f}_t$. A natural choice to remove $\Phi(\boldsymbol{Z})\boldsymbol{B}\boldsymbol{f}_t$ in (7) is to define the following projection matrix

$$\boldsymbol{P}_\Phi(\boldsymbol{Z}) \stackrel{\text{def}}{=} \Phi(\boldsymbol{Z})\left[\Phi(\boldsymbol{Z})^\top\Phi(\boldsymbol{Z})\right]^{-1}\Phi(\boldsymbol{Z})^\top. \tag{8}$$

Premultiplying both sides of (7) by $\boldsymbol{P}_\Phi(\boldsymbol{Z})$ and assuming that $(NT)^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t^\top\left[\boldsymbol{I}_N-\boldsymbol{P}_\Phi(\boldsymbol{Z})\right]\boldsymbol{X}_t$ is non singular, the following estimator for $\boldsymbol{\beta}$ is obtained,

$$\widehat{\boldsymbol{\beta}}=\left\{\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{X}_t^\top\left[\boldsymbol{I}_N-\boldsymbol{P}_\Phi(\boldsymbol{Z})\right]\boldsymbol{X}_t\right\}^{-1}\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{X}_t^\top\left[\boldsymbol{I}_N-\boldsymbol{P}_\Phi(\boldsymbol{Z})\right]\boldsymbol{y}_t. \tag{9}$$

3. Asymptotic properties

In this section, we analyze the main asymptotic properties of the estimator. First, we introduce some notation, definitions, and assumptions that will be necessary to derive the main results of this article. Later, we present the main large sample properties of these estimators. All proofs of these results are relegated to a Supplementary Material document.

3.1. Notation

Let $n=NT$. For two positive number sequences a_n and b_n, we say $a_n=\mathcal{O}(b_n)$ or $a_n\lesssim b_n$ (resp. $a_n\asymp b_n$) if there exists $C>0$ such that $a_n/b_n\le C$ (resp. $1/C\le a_n/b_n\le C$) for all large n, and say $a_n=\scriptstyle\mathcal{O}(b_n)$ if $a_n/b_n\to 0$ as $n\to\infty$. We set X_n and Y_n to be two sequences of random variables. Write $X_n=\mathcal{O}_p(Y_n)$ if for $\forall\epsilon>0$, there exists $C>0$ such that $\mathrm{P}(|X_n/Y_n|\le C)>1-\epsilon$ for all large n, and say $X_n=\scriptstyle\mathcal{O}_p(Y_n)$ if $X_n/Y_n\to 0$ in probability as $n\to\infty$. We use plim to denote the probability limit. Further, for a real matrix $\boldsymbol{A}$, let $\|\boldsymbol{A}\|_F=\mathrm{tr}^{1/2}(\boldsymbol{A}^\top\boldsymbol{A})$ and $\|\boldsymbol{A}\|_2=\lambda_{\max}^{1/2}(\boldsymbol{A}^\top\boldsymbol{A})$ denote its Frobenius and spectral norms, respectively. Let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of a square matrix, respectively. For a vector $\mathbf{v}$, let $\|\mathbf{v}\|$ denote its Euclidean norm.

3.2. Definitions and assumptions

Definition 3.1. *A function $h(\cdot)$ is said to belong to the class of additive functions $\mathcal{G}$, if: $h(\cdot)=\sum_{d=1}^{D}h_d(\cdot)$ and $h_d(\cdot)$ belongs to the Hölder class of functions*

$$\left\{h_d:|h_d^{(r)}(s)-h_d^{(r)}(t)|\le L|s-t|^\zeta\right\}$$

for some $L>0$, and for all s and t in the domain of $h_d(\cdot)$, where r stands for the r-th derivative of the real-valued function $h_d(\cdot)$ and $0<\zeta\le 1$.

For any scalar or vector function $\varphi(z)$, we use the notation $\Pi_{\mathcal{G}}[\varphi(z)]$ to denote the projection of $\varphi(z)$ onto the class of functions $\mathcal{G}$. That is, $\Pi_{\mathcal{G}}[\varphi(z)]$ is an element that belongs to $\mathcal{G}$ and is the closest function to $\varphi(z)$ among all the functions in $\mathcal{G}$. More specifically, we have

$$\begin{aligned}&\mathsf{E}\left\{[\varphi(z)-\Pi_{\mathcal{G}}(\varphi(z))][\varphi(z)-\Pi_{\mathcal{G}}(\varphi(z))]^\top\right\}\\&=\inf_{h\in\mathcal{G}}\mathsf{E}\left\{[\varphi(z)-h(z)][\varphi(z)-h(z)]^\top\right\},\end{aligned} \tag{10}$$

where the infimum is in the sense that

$$\begin{aligned}&\mathsf{E}\left\{[\varphi(z)-\Pi_{\mathcal{G}}(\varphi(z))][\varphi(z)-\Pi_{\mathcal{G}}(\varphi(z))]^\top\right\}\\&\le\mathsf{E}\left\{[\varphi(z)-h(z)][\varphi(z)-h(z)]^\top\right\},\end{aligned} \tag{11}$$

for all $h\in\mathcal{G}$, where for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A}\le\boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is negative semidefinite.

Denote $\theta(z)=\mathsf{E}[\boldsymbol{X}_t|\boldsymbol{Z}=z]$ and $m(z)$ is the projection of $\theta(z)$ onto $\mathcal{G}$, i.e., $m(z)=\mathsf{E}_{\mathcal{G}}[\theta(z)]$. For $t=1,\ldots,T$ we define $\boldsymbol{\xi}_t=\boldsymbol{X}_t-\boldsymbol{m}(\boldsymbol{Z})$, $\boldsymbol{\eta}(\boldsymbol{Z})=\theta(\boldsymbol{Z})-\boldsymbol{m}(\boldsymbol{Z})$, and $\boldsymbol{\varepsilon}_t=\boldsymbol{X}_t-\theta(\boldsymbol{Z})$, where $\boldsymbol{\xi}_t$, $\boldsymbol{\eta}(\boldsymbol{Z})$, and $\boldsymbol{\varepsilon}_t$ are $N\times Q$ matrices. Also, the following conditions about the data generating process, basis functions, factor loadings, and sieve approximation are required to obtain the large sample properties of the proposed estimator, $\widehat{\boldsymbol{\beta}}$.

Assumption 3.1. (Data generating process).

(i) *$\theta(z)$, $m(z)$, and $\eta(z)$ are bounded functions in $\mathcal{Z}$.*
(ii) $\sup_{z\in\mathcal{Z}} \mathsf{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top \,\middle|\, \boldsymbol{Z}=z\right) < C$*, for some $C>0$, $t=1,\ldots,T$.*
(iii) *Define* $\widetilde{V}_\xi = \operatorname{plim}_{N,T\to\infty} \frac{1}{NT}\sum_t \boldsymbol{\xi}_t^\top \boldsymbol{\xi}_t$*.$\widetilde{V}_\xi$ is finite and positive definite.*

Assumption 3.2. (Identification). *Almost surely, $T^{-1}\boldsymbol{F}^\top\boldsymbol{F} = \boldsymbol{I}_K$.*

Assumption 3.1 allows for correlation between $\boldsymbol{X}_{it}$ and $\boldsymbol{Z}_i$ through $\theta(\boldsymbol{Z})$, $m(\boldsymbol{Z})$, and $\eta(\boldsymbol{Z})$. This assumption is standard in semiparametric estimation techniques (see for example Assumption 2.1(ii) in Ahmad, Leelahanon, and Li (2005)). Also, Assumption 3.2 is commonly used in the estimation of factor models and enables to identify separately the factors $\boldsymbol{F}$ (see condition PC1 in Bai and Ng (2013)). Although this assumption is frequently employed in the factor literature, this does not imply that the true value of $\boldsymbol{F}$ must satisfy such a restriction. Indeed, when elements of $\boldsymbol{F}$ are randomly generated (e.g., from a normal distribution), one cannot ensure that the above condition holds unless $\boldsymbol{F}$ is normalized after all elements have been generated. Assumption 3.1 (iii) guarantees that $(NT)^{-1}\sum_{t=1}^T \boldsymbol{X}_t^\top\left[\boldsymbol{I}_N - \boldsymbol{P}_\Phi(\boldsymbol{Z})\right]\boldsymbol{X}_t$ is asymptotically non singular.

Assumption 3.3. (Sieve basis functions).

(i) *There are two positive constants, $c'_{\min}$ and $c'_{\max}$ such that, with probability approaching one (as $N\to\infty$),*

$$c'_{\min} < \lambda_{min}\left(N^{-1}\Phi(\boldsymbol{Z})^\top\Phi(\boldsymbol{Z})\right) < \lambda_{max}\left(N^{-1}\Phi(\boldsymbol{Z})^\top\Phi(\boldsymbol{Z})\right) < c'_{\max}.$$

(ii) $\max_{j\le J_N, i\le N, d\le D} \mathsf{E}[\phi_j(Z_{id})^2] < \infty$.

As already pointed out in Fan, Liao, and Wang (2016), $N^{-1}\Phi(\boldsymbol{Z})^\top\Phi(\boldsymbol{Z}) = N^{-1}\sum_{i=1}^N \phi\left(\boldsymbol{Z}_i\right)^\top \phi\left(\boldsymbol{Z}_i\right)$ and $\phi\left(\boldsymbol{Z}_i\right)$ is of order $J_N D$ much smaller than N. Thus, condition (i) can follow from a strong law of large numbers. This condition can be satisfied through proper normalizations of commonly used basis functions.

The following set of conditions is concerned with the accuracy of the sieve approximation.

Assumption 3.4. (Accuracy of sieve approximation).

(i) *For $k=1,\ldots,K$, $g_k(\cdot)\in\mathcal{G}$ and for $q=1,\ldots,Q$, $m_q(\cdot)\in\mathcal{G}$, where $m_q(\cdot)$ is the qth column of $m(\cdot)$.*
(ii) *For $k=1,\ldots,K$, $d=1,\ldots,D$, $q=1,\ldots,Q$, and $i=1,\ldots,N$, and let r and ζ be elements already stated in Definition* 3.1*. The sieve coefficients $\{b_{j,kd}\}_{j=1}^{J_N}$ and $\{c_{j,qd}\}_{j=1}^{J_N}$ satisfy, for $\kappa = 2(r+\zeta)\ge 4$ as $J_N\to\infty$,*

$$\sup_{z\in\mathcal{Z}_d}\left|g_{kd}(z) - \sum_{j=1}^{J_N} b_{j,kd}\phi_j(z)\right|^2 = \mathcal{O}\left(J_N^{-\kappa}\right),$$

$$\sup_{z\in\mathcal{Z}_d}\left|m_{qd}(z) - \sum_{j=1}^{J_N} c_{j,qd}\phi_j(z)\right|^2 = \mathcal{O}\left(J_N^{-\kappa}\right),$$

where $m_{qd}(z)$ is the d-th additive element of $m_q(z)$, $\mathcal{Z}_d$ is the support of the d-th element of $\mathcal{Z}$, and J_N is the sieve dimension.

(iii) $\max_{j,k,d} b_{j,kd}^2 < \infty$, $\max_{j,q,d} c_{j,qd}^2 < \infty$.

As it is remarked in Fan, Liao, and Wang (2016), Assumption 3.4 (ii) is satisfied by the use of common basis functions such as polynomial basis or B-splines. In particular, Lorentz (1986) and Chen (2007) show that (i) implies (ii) in this particular case.

The next assumption refers to the error factor loadings $\boldsymbol{\gamma}_i$, for $i = 1, \ldots, N$.

Assumption 3.5. (Error factor loadings).

(i) *$\{\boldsymbol{\gamma}_i\}_{i\le N}$ is independent of $\{\mathbf{Z}_i\}_{i\le N}$. Furthermore, conditionally on $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T$, $\{\boldsymbol{\gamma}_i\}_{i\le N}$ is independent of $\left\{\boldsymbol{\xi}_t\right\}_{t\le T}$ and $\mathsf{E}(\gamma_{ik}) = 0$ for $k = 1, \ldots, K$.*

(ii) $\max_{k\le K, i\le N} \mathsf{E}\left[g_k(\mathbf{Z}_i)^2\right] < \infty$. *Also, $\nu_N < \infty$ and*

$$\max_{k\le K, j\le N} \sum_{i\le N} |\,\mathsf{E}(\gamma_{ik}\gamma_{jk})| = \mathcal{O}(\nu_N),$$

where

$$\nu_N = \max_{k\le K} N^{-1} \sum_{i\le N} \mathsf{Var}(\gamma_{ik}),$$

(iii) *For some $\delta > 2$,*

$$\max_{i\le N; k\le K} \mathsf{E}\left|\gamma_{ik}\frac{1}{T}\sum_{t=1}^{T} \mathsf{E}\left(\xi_{itq}f_{kt}\middle|\,\boldsymbol{\Gamma}\right)\right|^{\delta} < \infty, \quad q = 1, \ldots, Q. \tag{12}$$

Note that in Assumption 3.5 (ii) we assume cross-sectional dependence of the error factor loadings. To show the consistency of the proposed estimator for simplicity, we can assume the independence of the factor loadings γ_{ik} from the random part of the covariates, $\mathbf{Z}_i$, but we do not need to impose a restrictive i.i.d. assumption.

Through the article, some regularity conditions about weak dependence and stationarity are assumed on the factors and the idiosyncratic terms. In particular, we impose strong mixing conditions. Let $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_T^\infty$ denote the σ-algebras generated by $\{(\boldsymbol{\xi}_t, \boldsymbol{f}_t, \boldsymbol{u}_t) : t \le 0\}$ and $\{(\boldsymbol{\xi}_t, \boldsymbol{f}_t, \boldsymbol{u}_t) : t \ge T\}$, respectively. Define the mixing coefficient

$$\alpha(T) = \sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_T^\infty} |\mathrm{P}(A)\mathrm{P}(B) - \mathrm{P}(AB)|.$$

Assumption 3.6. (Data generating process).

(i) *$\{\boldsymbol{\xi}_t, \boldsymbol{u}_t, \boldsymbol{f}_t\}_{t\le T}$ is strictly stationary, $\{\boldsymbol{u}_t\}_{t\le T}$ is independent of $\{\mathbf{Z}_i, \boldsymbol{\gamma}_i, \boldsymbol{\xi}_t, \boldsymbol{f}_t\}_{i\le N; t\le T}$ and $\mathsf{E}(u_{it}) = 0$ for all $i \le N$, $t \le T$.*

(ii) *For some $\delta > 2$,*

$$\max_{t\le T} \mathsf{E}\left|\xi_{itq}u_{it}\right|^{\delta} < \infty, \quad i = 1, \ldots, N; \quad q = 1, \ldots, Q, \tag{13}$$

$$\max_{t\le T}\max_{k\le K} \mathsf{E}\,|\xi_{itq}f_{tk}|^{\delta} = M_\delta < \infty, \quad i = 1, \ldots, N; \quad q = 1, \ldots, Q. \tag{14}$$

(iii) *Strong mixing: $\alpha(k) \le ak^{-\tau}$, where a is a positive constant and $\tau > \frac{\delta}{\delta-2}$.*

(iv) *Weak dependence: there is $C > 0$ so that*

$$\max_{j\le N}\sum_{i=1}^{N}|\,\mathsf{E}(u_{it}u_{jt})| < C,$$

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\,\mathsf{E}(u_{it}u_{js})| < C,$$

$$\max_{i\le N}(NT)^{-1}\sum_{l=1}^{N}\sum_{l'=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\mathsf{Cov}\,(u_{it}u_{lt}, u_{is}u_{l's})| < C.$$

Assumption 3.6 is standard in factor analysis (Bai, 2003; Stock and Watson, 2002; Fan, Liao, and Wang, 2016). Part (i) is standard in partially linear models (Ahmad, Leelahanon, and Li, 2005; Härdle, Liang, and Gao, 2000). The independence assumption between $\boldsymbol{u}_t$ and $\{\boldsymbol{Z}_i, \boldsymbol{\xi}_t\}$ can be relaxed by allowing for conditional independence. Part (iii) is a strong mixing condition for the weak temporal dependence of $\{\boldsymbol{\xi}_t, \boldsymbol{u}_t, \boldsymbol{f}_t\}$, whereas (iv) imposes weak cross-sectional dependence in $\{u_{it}\}_{i\le N, t\le T}$. This condition is usually satisfied when the covariance matrix of the error term u_{it} is sufficiently sparse under the strong mixing condition and it is commonly imposed for high-dimensional factor analysis.

3.3. Limiting theory

A very intuitive idea of the asymptotic behavior of our estimator can be obtained by plugging (2) in (9) that yields

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left[\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})\boldsymbol{X}_t\right]^{-1}\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})(\boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{u}_t),$$

where $\boldsymbol{M}_\Phi(\boldsymbol{Z}) = \boldsymbol{I}_N - \boldsymbol{P}_\Phi(\boldsymbol{Z})$ and $\boldsymbol{\Lambda} = \boldsymbol{G}(\boldsymbol{Z}) + \boldsymbol{\Gamma}$. As the reader can see from the above expression, there is a direct dependence of $\widehat{\boldsymbol{\beta}}$ on the unobserved factor loadings through $(NT)^{-1}\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})\boldsymbol{\Lambda}\boldsymbol{f}_t$. Nevertheless, using (1) and given that it can be proved that $(NT)^{-1}\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})\boldsymbol{G}(\boldsymbol{Z})\boldsymbol{f}_t = \mathcal{O}_p(1/\sqrt{NT})$ (see the proof of Theorem 3.1 in the Supplementary Material document), we have that

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left[\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})\boldsymbol{X}_t\right]^{-1}\sum_t \boldsymbol{X}_t^\top \boldsymbol{M}_\Phi(\boldsymbol{Z})(\boldsymbol{\Gamma}\boldsymbol{f}_t + \boldsymbol{u}_t) + \mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right). \tag{15}$$

In this situation, we can conclude that the limiting distribution of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ only depends on idiosyncratic terms (related to both the error term and the approximation error of the basis functions to the factor loadings). Under Assumptions 3.1–3.6, it is also possible to show (see Appendix A of the Supplementary Material document for a related proof) that, as both N and T tend to infinity,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathcal{O}_p\left(1/\sqrt{NT}\right) + \mathcal{O}_p\left(\sqrt{\nu_N/N}\right).$$

The interesting feature of this asymptotic bound is that the rate of convergence of $\widehat{\boldsymbol{\beta}}$ can be slower than $\sqrt{NT}$ depending on the behavior of ν_N (i.e., the variance of the error factor loadings). In order to clarify this, we further take a look at the asymptotic distribution of our estimator. The previous result on the consistency and the convergence rate is based on weak dependence in the error term and idiosyncratic factor loadings. To show asymptotic normality, we have to impose the stronger condition of cross-sectional independence while still allowing for weak dependence in the time dimension.

Assumption 3.7. $\left\{\boldsymbol{u}_i, \boldsymbol{\gamma}_i\right\}_{i\le N}$ *are independent and non identically distributed random variables across i.*

Then, the asymptotic distribution of the projection-based interactive fixed effects estimator $\widehat{\boldsymbol{\beta}}$ is provided in the following theorem.

Theorem 3.1. *(Limiting distribution). Let $\nu_N \sim T^{-\vartheta}$ for $\vartheta \geq 0$. Under assumptions 3.1–3.7 and if it is further assumed that, for $\kappa \geq 4$ and $\varrho \in \left(\frac{1}{\kappa}, \frac{1}{2}\right)$, $J_N \sim N^{\varrho}$ and $T/N^{\kappa\varrho-1} \to 0$, as both N and T tend to infinity, then for $\vartheta \in [0, 1)$,*

$$\sqrt{NT^{\vartheta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N\left(0, \widetilde{V}_{\xi}^{-1}\widetilde{V}_{\Gamma}\widetilde{V}_{\xi}^{-1}\right). \tag{16}$$

Under the same set of assumptions, if $\vartheta = 1$

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N\left(0, \widetilde{V}_{\xi}^{-1}\left(\widetilde{V}_{\Gamma} + \widetilde{V}_{u}\right)\widetilde{V}_{\xi}^{-1}\right), \tag{17}$$

and finally, if $\vartheta > 1$

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N\left(0, \widetilde{V}_{\xi}^{-1}\widetilde{V}_{u}\widetilde{V}_{\xi}^{-1}\right), \tag{18}$$

where

$$\widetilde{V}_{\xi} \stackrel{\text{def}}{=} \text{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T} \boldsymbol{\xi}_t^{\top}\boldsymbol{\xi}_t, \tag{19}$$

$$\widetilde{V}_{\Gamma} \stackrel{\text{def}}{=} \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{k\leq K;k'\leq K} \mathsf{E}\left[\gamma_{ik}\gamma_{ik'}\frac{1}{T^2}\sum_{t,s}\mathsf{E}\left(\boldsymbol{\xi}_{it}f_{tk}\middle|\boldsymbol{\Gamma}\right)\mathsf{E}\left(\boldsymbol{\xi}_{is}f_{sk'}\middle|\boldsymbol{\Gamma}\right)^{\top}\right], \tag{20}$$

$$\widetilde{V}_{u} \stackrel{\text{def}}{=} \lim_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathsf{E}(\boldsymbol{\xi}_t^{\top}\boldsymbol{u}_t\boldsymbol{u}_{t'}^{\top}\boldsymbol{\xi}_{t'}). \tag{21}$$

Note that ϑ is a parameter that reflects the strength of the relationship between $\boldsymbol{\lambda}_i$ and $\boldsymbol{Z}_i$ through the variance of $\boldsymbol{\gamma}_i$. Thus, when $\vartheta \approx 0$ the relationship is weak whereas when $\vartheta \gg 0$ this relationship becomes stronger.

The proof of Theorem 3.1 is provided in Appendix B.1 in the Supplementary Material document. The key component of the proof is following the Frisch-Waugh Theorem to partial-out the effect of the latent factors and corresponding loadings. We want to highlight that the relative rate requirements of N and T crucially depend on the smoothness parameter, κ. In particular, if $\kappa = 4$ we have the requirement that T/N tends to zero regardless of the choice for the sieve dimension ϱ. The constraints in the rates of growth imposed on N and T are similar to other assumptions used in similar literature such as in Ahmad, Leelahanon, and Li (2005). Note that Assumption 3.7 is introduced for the sake of simplicity. It is indeed used for the application of the corresponding central limit theorems (CLT) in the proof but it could be relaxed at the cost of a much cumbersome proof.

Remark 3.1. As we can observe from (15) and Theorem 3.1 the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ depends on interplay of two leading terms,

$$\sum_t \boldsymbol{X}_t^{\top}\boldsymbol{M}_{\Phi}(\boldsymbol{Z})\boldsymbol{\Gamma}\boldsymbol{f}_t + \sum_t \boldsymbol{X}_t^{\top}\boldsymbol{M}_{\Phi}(\boldsymbol{Z})\boldsymbol{u}_t.$$

The term $(NT)^{-1}\sum_t \boldsymbol{X}_t^{\top}\boldsymbol{M}_{\Phi}(\boldsymbol{Z})\boldsymbol{\Gamma}\boldsymbol{f}_t$ arises from the cross-sectional estimation, and it shows a rate of order $\mathcal{O}_p\left(N^{-1/2}\sqrt{\nu_N}\right)$ and the term $(NT)^{-1}\sum_t \boldsymbol{X}_t^{\top}\boldsymbol{M}_{\Phi}(\boldsymbol{Z})\boldsymbol{u}_t$ which has a leading term of order $\mathcal{O}_p\left((NT)^{-1/2}\right)$ (see **Proof of Theorem 3.1 (iii)** of the Supplementary Material). Indeed, the interaction between these two leading terms affects crucially the resulting rate of convergence of the limiting distribution. As it can be observed in Theorem 3.1, this rate is affected by the behavior of ν_N. This term reflects the strength of the relationship between the $\boldsymbol{\lambda}_i$'s and the $\boldsymbol{Z}_i$'s. When a relevant part of the variation of the loading coefficients $\boldsymbol{\lambda}_i$ is explained by $\boldsymbol{Z}_i$ (that is ν_N is close to zero) the observed characteristics capture almost all fluctuations of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, leading to a faster rate of convergence $\sqrt{NT}$. On the other hand, if ν_N is far from zero, then the fluctuations of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ can be explained mostly by cross-sectional variation, and therefore time series regression is not relevant to help to remove the correlation between loading coefficients and covariates when estimating the $\boldsymbol{\beta}$'s. In this case, the limiting distribution is determined by a cross-sectional CLT and hence the rate of convergence is slower (note that for $\nu_N = \mathcal{O}(1)$ the rate is $\sqrt{N}$).

Remark 3.2. From Theorem 3.1 it is also possible to identify specifications under which our estimator might outperform the PCA or the CCE estimators. If $\boldsymbol{\lambda}_i = \boldsymbol{g}(\boldsymbol{Z}_i) + \boldsymbol{\gamma}_i$ and $\nu_N \approx 0$ our estimator appears as more efficient as the others. If we further assume that the latent factor loadings can be completely explained by the nonparametric functions, i.e., $\boldsymbol{\Gamma} = 0$ as in Zhang, Zhou, and Wang (2021), and given the idiosyncratic error terms are i.i.d. with $\mathsf{Var}(u_{it}) = \sigma^2$, our estimator is semiparametrically efficient in the sense that the inverse of the asymptotic variance of $\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ equals the semiparametric efficiency bound. From the result of Chamberlain (1992) the semiparametric efficiency bound for the inverse of the asymptotic variance of an estimator of $\boldsymbol{\beta}$ is

$$\mathcal{J}_0 = \inf_{g \in \mathcal{G}} \mathsf{E}\left\{\left[\boldsymbol{X}_{it} - \boldsymbol{g}(\boldsymbol{Z}_i)\right] \mathsf{Var}\,(u_{it})^{-1} \left[\boldsymbol{X}_{it} - \boldsymbol{g}(\boldsymbol{Z}_i)\right]^\top\right\}. \tag{22}$$

Under the i.i.d. Assumption, (22) can be rewritten as

$$\begin{aligned} \mathcal{J}_0 &= \frac{1}{\sigma^2} \inf_{g \in \mathcal{G}} \mathsf{E}\left\{\left[\boldsymbol{X}_{it} - \boldsymbol{g}(\boldsymbol{Z}_i)\right]\left[\boldsymbol{X}_{it} - \boldsymbol{g}(\boldsymbol{Z}_i)\right]^\top\right\} \\ &= \frac{1}{\sigma^2} \mathsf{E}\left\{\left[\boldsymbol{X}_{it} - \boldsymbol{m}(\boldsymbol{Z}_i)\right]\left[\boldsymbol{X}_{it} - \boldsymbol{m}(\boldsymbol{Z}_i)\right]^\top\right\} \\ &= \frac{1}{\sigma^2} \mathsf{E}\left\{\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}^\top\right\}. \end{aligned}$$

Note that the inverse of the last expression coincides with the asymptotic variance of $\sqrt{NT}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ when the error terms are uncorrelated and homoskedastic. Then, $\widehat{\boldsymbol{\beta}}$ is a semiparametrically efficient estimator under these assumptions.

Remark 3.3. It is possible to estimate the latent factors and loading coefficients from the regression residuals using the Projected-PCA method of Fan, Liao, and Wang (2016). Let $\widetilde{\boldsymbol{y}}_t = \boldsymbol{y}_t - \boldsymbol{X}_t\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{Y}} = (\widetilde{\boldsymbol{y}}_1, \ldots, \widetilde{\boldsymbol{y}}_T)$. Now the matrix of factors $\boldsymbol{F}$ and $\boldsymbol{G}(Z)$ can be recovered from the projected matrix of residuals $\boldsymbol{P}_\Phi(\boldsymbol{Z})\widetilde{\boldsymbol{Y}}$. The asymptotic properties and the resulting convergence rates remain unaffected by the need to estimate the regression coefficients in a first step. We provide details on the estimation of the latent factors and loadings in Section C of the Supplementary Material document.

3.4. Uniformly valid inference via the cross-sectional bootstrap

The results of Theorem 3.1 have important implications for conducting inference on the estimated regression parameters. In particular, since the variance of the idiosyncratic part of the factor loadings decides which of the two terms will be the leading one, it will ultimately determine the convergence rate of our estimator. As a consequence, the asymptotic distribution of $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ has a discontinuity when the variance of the factor loadings is close to the boundary. In financial econometrics studies, this issue is often circumvented by assuming weak heterogeneity as a default setting (Connor and Linton, 2007; Connor, Hagmann, and Linton, 2012a). However, recent studies such as Fan, Liao, and Wang (2016) find empirical evidence for the case of strong heterogeneity. Further, usual plug-in approaches based on estimated asymptotic covariance matrices will lead to misleading conclusions if $\nu_N = \mathcal{O}(T^{-1})$, as they provide confidence intervals that are too wide because the asymptotic covariance matrix is overestimated, and the coverage probabilities will be too conservative (Liao and Yang, 2018; Fernández-Val et al., 2022). Similarly, simply ignoring the cross-sectional term will lead to under-coverage in the strong heterogeneity case.

Uniformly valid inference for panel data models is an important topic beyond the specifics of our model setup. For instance, Lu and Su (2023) observe a similar issue in a panel model with two-dimensional heterogeneity in the regression parameters. In their case, the issue is caused by the level of temporal and cross-sectional heterogeneity in the slope coefficients.

Fortunately, the uniformity issue can be solved by using the cross-sectional bootstrap proposed by Kapetanios (2008). Besides achieving uniformly valid inference, the approach is both intuitive and easy to implement. The basic idea is to sample with replacement cross-sectional units while keeping the entire time series of the sampled individual units unchanged. By doing this, the resampling scheme directly mimics the cross-sectional variations in $\boldsymbol{\Gamma}$, regardless of the underlying level of heterogeneity. The consequence is a uniformly valid inference. This is in direct contrast to Andrews (2000), who found that the usual bootstrap will lead to inconsistency when a parameter is on the boundary of the support. The reason why this problem does not occur in our case is that we do not explicitly model the variance of the idiosyncratic factor loadings as a parameter, i.e., it does not appear in the loss function of our least squares problem.

A crucial assumption for the bootstrap validity is that the data is cross-sectionally independent. In fact, Menzel (2021) showed that uniform non conservative inference is impossible under general dependence in more than one dimension. Recently, De Vos and Stauskas (2024) studied the theoretical properties of the cross-sectional bootstrap for the CCE approach of Pesaran (2006) and proposed a bias-correction procedure in the asymptotic regime $N/T \to \rho < \infty$. The uniform validity of the bootstrap procedure in settings similar to ours was recently shown in Liao and Yang (2018) and Fernández-Val et al. (2022). The specific aspect of our procedure is that we only resample cross-sectional units after projecting the data, i.e., removing the effect of $\boldsymbol{Z}_i$ on the factor loadings.

As a positive side effect, the cross-sectional bootstrap is able to keep the dependence in the time dimension. Therefore, the inference is also robust toward serial dependence in the idiosyncratic error term and in the latent factors. In the following, we summarize the steps of the cross-sectional bootstrap procedure.

Step 1: Choose a confidence level α, and the number of bootstrap samples, B.

Step 2: Regress y_{it} and X_{itq} on $\Phi(\boldsymbol{Z})$, and obtain residuals, $\dot{\boldsymbol{y}}_t \overset{\text{def}}{=} [\boldsymbol{I}_N - \boldsymbol{P}_\Phi(\boldsymbol{Z})]\boldsymbol{y}_t$ and $\dot{\boldsymbol{X}}_t \overset{\text{def}}{=} [\boldsymbol{I}_N - \boldsymbol{P}_\Phi(\boldsymbol{Z})]\boldsymbol{X}_t$ for $t = 1, \dots, T$.

Step 3: Calculate $\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^T \dot{\boldsymbol{X}}_t^\top \dot{\boldsymbol{X}}_t\right)^{-1} \sum_{t=1}^T \dot{\boldsymbol{X}}_t^\top \dot{\boldsymbol{y}}_t$.

Step 4: For $b = 1, \dots, B$, draw a sample of N cross-sectional units with replacement while keeping the unit's entire time series unchanged. Denote the resulting matrices of regressors and vectors of dependent variables by $\dot{\boldsymbol{X}}_{b,t}^*$ and $\dot{\boldsymbol{y}}_{b,t}^*$, respectively.

Step 5: Obtain the bootstrap estimate $\widehat{\boldsymbol{\beta}}_b^* = \left(\sum_{t=1}^T \dot{\boldsymbol{X}}_{b,t}^{*\top} \dot{\boldsymbol{X}}_{b,t}^*\right)^{-1} \sum_{t=1}^T \dot{\boldsymbol{X}}_{b,t}^{*\top} \dot{\boldsymbol{y}}_{b,t}^*$.

Step 6: Calculate the $(1-\alpha)$-confidence interval for the j-th component of $\boldsymbol{\beta}_j$,

$$CI_\alpha(\boldsymbol{\beta}_j) = \widehat{\boldsymbol{\beta}}_j \pm q_{\alpha,j},$$

where $q_{\alpha,j}$ is the $(1-\alpha)$-quantile of the bootstrap distribution of $\left|\widehat{\boldsymbol{\beta}}_{b,j}^* - \widehat{\boldsymbol{\beta}}\right|$. Or, more generally, for $v \in \mathbb{R}^Q$,

$$CI_\alpha(v^\top \boldsymbol{\beta}) = v^\top \widehat{\boldsymbol{\beta}} \pm q_{\alpha,v},$$

where $q_{\alpha,v}$ is the $(1-\alpha)$-quantile of the bootstrap distribution of $\left|v^\top(\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}})\right|$.

For the bootstrap validity, we need to assume the existence of a consistent estimator of the variance of $v^\top \widehat{\boldsymbol{\beta}}$.

Assumption 3.8. *Denote* $V_{\boldsymbol{\beta},v} = \lim_{N,T\to\infty} \mathsf{Var}[\sqrt{NT^{\vartheta}}v^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$, *if* $\vartheta \in [0,1)$ *and* $V_{\boldsymbol{\beta},v} = \lim_{N,T\to\infty} \mathsf{Var}[\sqrt{NT}v^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$, *if* $\vartheta \geq 1$. *There exists a consistent estimator* $V_{\boldsymbol{\beta},v,n}$, *satisfying* $V_{\boldsymbol{\beta},v}^{-1/2} - V_{\boldsymbol{\beta},v,n}^{-1/2} = \mathcal{O}_p(1)$.

The following theorem provides the bootstrap validity, uniformly over settings with varying degrees of variability in the idiosyncratic factor loadings.

Theorem 3.2. *(Bootstrap Validity). Let* $\{\mathrm{P}_T : T \geq 1\} \subset \mathcal{P}$ *be sequences of probability laws. Let the conditions of our Theorem 3.1 hold uniformly over these sequences. Further assume that* u_{it} *and* $\boldsymbol{\gamma}_i$ *are cross-sectionally independent. Then we have, uniformly for all* $\{\mathrm{P}_T : T \geq 1\} \subset \mathcal{P}$, *and for a confidence level* $1 - \alpha$,

$$\mathrm{P}_T\left(v^\top \boldsymbol{\beta} \in CI_\alpha(v^\top \beta)\right) \to 1 - \alpha.$$

The proof of Theorem 3.2 can be found in Appendix B.2 in the Supplementary Material document. An essential part of the proof is to show that the asymptotic expansion of the bootstrap version of the estimator is identical to that of the original estimator.

4. Numerical studies

In this section, we evaluate the finite-sample performance of our estimator in a simulation study. We are interested both in the estimation accuracy of the parameter vector, $\boldsymbol{\beta}$, and the empirical coverage probabilities of the cross-sectional bootstrap procedure. Throughout the study, we fix the number of factors, $K = 3$, the dimension of the time-invariant variable is set to $D = 2$, and the dimension of covariates is set to $Q = 2$. The true regression coefficients are $\boldsymbol{\beta} = (2, -1)^\top$. The time-invariant variables are generated by i.i.d. $Z_{id} \sim U[-1, 1]$. The covariates are generated by setting $X_{itq} = \boldsymbol{a}_{iq}^\top \boldsymbol{f}_t + 2(\sqrt{g_1(\boldsymbol{Z}_i)}, \ldots, \sqrt{g_K(\boldsymbol{Z}_i)})^\top \boldsymbol{b}_q + \pi_{itq}$, where $\pi_{itq} \sim N(0, 1)$ i.i.d., $a_{iqk} \sim U[-0.5, 0.5]$ and $b_{qk} \sim U[-1, 1]$. We generate the latent factors, $(f_{k1}, \ldots, f_{kT})$, as MA(∞) processes with algebraic decay and under independence across factors for all k. The factor loadings are set to $\lambda_{ik} = g_k(\boldsymbol{Z}_i) + \gamma_{ik}$, where $g_1(z) = \sin(2z_1)^3 + \cos(z_2^2)$, $g_2(z) = -\tan(z_1^2) + 2\cos(z_2 + 1)$ and $g_3(z) = z_2^3 - \sin(3z_1)$.

Finally, for the idiosyncratic error term, we consider the case of i.i.d. standard normal u_{it} as well as the case of weak temporal dependence, in which $(u_{i1}, \ldots, u_{iT})$ are generated from a MA(∞) process with algebraic decay parameter. To be precise, we set $u_{it} = \sum_{s=1}^{\infty} 5^{-s} e_{i,t-s}$ with i.i.d. standard normal innovations, $e_{i,s}$, for all i. We use the same MA(∞) process for the generation of the the latent factors. For the idiosyncratic part of the factor loadings, we consider four settings. First, in the strong heterogeneity case generate $\boldsymbol{\gamma}_i \sim N(0, 0.5)$ (i.e., $\nu_N = \mathcal{O}(1)$). Second, we consider the special case $\nu_N = 0$. Third, we consider the weak factor case in which $\boldsymbol{\lambda}_i^\top \boldsymbol{f}_t = \mathcal{O}(T^{-1/2})$ (i.e., $\nu_N = \mathcal{O}(T^{-1})$). Finally, we generate the factor loadings under the setting that the nonparametric functions have no explanatory power, $\boldsymbol{g}(\cdot) = 0$, but $\nu_N = \mathcal{O}(1)$. Throughout this numerical study, we rely on B-spline basis functions and we select $J_N = \lceil N^{1/3} 1.5 \rceil$. For each setting 500, Monte Carlo runs are conducted.

We compare the performance of our projection-based interactive fixed effects (P-IFE) estimator for $\boldsymbol{\beta}$ with the principal component-based interactive fixed effects (PC-IFE) estimator of Bai (2009) in the i.i.d. case and with a bias-corrected version of the same estimator (bc-PC-IFE) in the serially dependent case. For these comparisons, we rely on the R package `phtt` (Bada and Liebl, 2014). The number of factors is selected according to the PC1 criterion in Bai and Ng (2002). As performance measures, we consider the root mean square error (*RMSE*). The simulation results under Gaussian disturbances for different values of ν_N, N, and T are reported in Table 1. The *RMSE* of our P-IFE can be effectively reduced with increasing sample size. For the strong heterogeneity case, we observe an advantage of the PC-IFE for small and medium samples. For $N = 500$, this advantage is reversed and the P-IFE has a higher accuracy. In the other two settings for ν_N, we can see that the P-IFE outperforms its competitors in almost all cases. The

Table 1. *RMSE* of the P-IFE estimator and the PC-IFE estimator under i.i.d. Gaussian error terms.

			β_1		β_2	
ν_N	N	T	P-IFE	PC-IFE	P-IFE	PC-IFE
$\nu_N = \mathcal{O}(1)$	50	10	0.0670	0.0582	0.0690	0.0605
	100	10	0.0440	0.0437	0.0449	0.0446
	50	50	0.0385	0.0234	0.0399	0.0254
	100	50	0.0278	0.0195	0.0282	0.0204
	200	100	0.0161	0.0142	0.0171	0.0139
	500	100	0.0100	0.0117	0.0100	0.0118
$\nu_N = 0$	50	10	0.0401	0.0569	0.0452	0.0581
	100	10	0.0286	0.0438	0.0292	0.0436
	50	50	0.0205	0.0249	0.0196	0.0243
	100	50	0.0124	0.0208	0.0138	0.0210
	200	100	0.0065	0.0148	0.0065	0.0138
	500	100	0.0038	0.0136	0.0039	0.0139
$\nu_N = \mathcal{O}(T^{-1})$	50	10	0.0468	0.0585	0.0458	0.0538
	100	10	0.0319	0.0380	0.0308	0.0381
	50	50	0.0193	0.0197	0.0200	0.0196
	100	50	0.0139	0.0150	0.0138	0.0154
	200	100	0.0065	0.0078	0.0066	0.0077
	500	100	0.0039	0.0058	0.0041	0.0057
$\nu_N = \mathcal{O}(1)$	50	10	0.0646	0.0493	0.0708	0.0568
$\boldsymbol{g}(\cdot) = 0$	100	10	0.0456	0.0349	0.0429	0.0349
	50	50	0.0417	0.0184	0.0402	0.0189
	100	50	0.0289	0.0127	0.0271	0.0125
	200	100	0.0169	0.0064	0.0168	0.0062
	500	100	0.0099	0.0038	0.0101	0.0038

outperformance is best visible for settings with large sample sizes in the case of $\nu_N = 0$. In particular, for $N = 500$ the *RMSE* of our P-IFE is less than a third of that of the PC-IFE. As expected, the performance of the P-IFE is worse than the PC-IFE if the effect of the nonparametric functions on the factor loadings is absent, $\boldsymbol{g}(z) = 0$.

The results for serially dependent error terms are displayed in Table 2. Again, we can observe that the PC-IFE outperforms the P-IFE in the $\nu_N = \mathcal{O}(1)$ case for small and medium sample sizes. Also similar to the i.i.d. case, the P-IFE has the lower *RMSE* in all settings for $\nu_N = 0$ and $\nu_N = \mathcal{O}(T^{-1})$, whereas the *RMSE* is higher in the case of no explanatory power of the nonparametric functions. As a robustness check, we also consider t-distributed error terms in the Supplementary Material document. See Table A1 for the results. We also consider a more complicated additional data-generating process with a larger number of factors, $K = 10$. The results are displayed in Table A2 for i.i.d. errors and Table A3 for serially correlated errors. The most notable difference to the results of the first DGP is that the P-IFE outperforms the PC-IFE even in the strong heterogeneity case for settings with a small time dimension, $T = 10$. Again, the P-IFE dominates in all settings for $\nu_N = 0$ and $\nu_N = \mathcal{O}(T^{-1})$.

In the following, we look at the performance of the cross-sectional bootstrap procedure and show its validity in finite samples. As a comparison, we look at the empirical coverage of the PC-IFE estimator. By Corollary 1 in Bai (2009), under the assumption of i.i.d. error terms, $\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{\text{PC-IFE}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(0, \sigma^2 \boldsymbol{D}^{-1})$, where $\boldsymbol{D} = \text{plim}(NT)^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i^\top \boldsymbol{Z}_i$, $\boldsymbol{Z}_i = \boldsymbol{M}_F \boldsymbol{X}_i - 1/N \sum_{k=1}^{N} \boldsymbol{M}_F \boldsymbol{X}_k a_{ik}$, $\boldsymbol{M}_F = \boldsymbol{I}_N - \boldsymbol{F}\boldsymbol{F}^\top / T$ and $a_{ik} = \boldsymbol{\lambda}_i^\top (\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})^{-1} \boldsymbol{\lambda}_k$. We construct confidence intervals based on the asymptotic distribution with an estimated covariance matrix based on estimated factors and factor loadings.

Table 3 shows that the empirical coverage of our cross-sectional bootstrap procedure approaches the nominal coverage level as N and T increase. As is often the case, we can observe slight under-coverage in small samples. However, the issue becomes virtually absent in settings with the largest sample size. We want to highlight that these findings hold for all settings for the variance of the idiosyncratic factor loadings, ν_N. We have thus provided evidence for the uniform validity of the bootstrap procedure in finite samples. In the Supplementary Material document, we show that the cross-sectional bootstrap

Table 2. *RMSE* of the P-IFE estimator and the bias-corrected PC-IFE estimator under serially dependent Gaussian error terms (MA(∞)).

			β_1		β_2	
ν_N	N	T	P-IFE	bc-PC-IFE	P-IFE	bc-PC-IFE
$\nu_N = \mathcal{O}(1)$	50	10	0.0704	0.0629	0.0648	0.0626
	100	10	0.0438	0.0423	0.0463	0.0442
	50	50	0.0403	0.0252	0.0412	0.0252
	100	50	0.0286	0.0189	0.0285	0.0201
	200	100	0.0160	0.0124	0.0165	0.0126
	500	100	0.0102	0.0112	0.0101	0.0114
$\nu_N = 0$	50	10	0.0444	0.0604	0.0435	0.0586
	100	10	0.0297	0.0456	0.0288	0.0451
	50	50	0.0196	0.0261	0.0202	0.0262
	100	50	0.0124	0.0211	0.0132	0.0218
	200	100	0.0064	0.0152	0.0063	0.0151
	500	100	0.0040	0.0131	0.0040	0.0134
$\nu_N = \mathcal{O}(T^{-1})$	50	10	0.0463	0.0578	0.0455	0.0560
	100	10	0.0324	0.0374	0.0286	0.0365
	50	50	0.0199	0.0191	0.0204	0.0190
	100	50	0.0138	0.0157	0.0136	0.0153
	200	100	0.0065	0.0083	0.0070	0.0087
	500	100	0.0040	0.0066	0.0040	0.0063
$\nu_N = \mathcal{O}(1)$	50	10	0.0682	0.0505	0.0673	0.0525
$g(\cdot) = 0$	100	10	0.0455	0.0365	0.0475	0.0364
	50	50	0.0405	0.0191	0.0428	0.0194
	100	50	0.0274	0.0134	0.0272	0.0140
	200	100	0.0171	0.0069	0.0154	0.0070
	500	100	0.0105	0.0045	0.0102	0.0042

Table 3. Empirical coverage of the cross-sectional bootstrap confidence intervals vs. empirical coverage of the asymptotic confidence intervals of Bai (2009) for variable X_1.

			P-IFE			PC-IFE		
ν_N	N	T	90%	95%	99%	90%	95%	99%
$\nu_N = \mathcal{O}(1)$	50	10	0.856	0.902	0.966	0.846	0.898	0.974
	100	10	0.856	0.932	0.982	0.836	0.902	0.970
	50	50	0.880	0.922	0.984	0.810	0.870	0.950
	100	50	0.860	0.922	0.978	0.726	0.804	0.914
	200	100	0.858	0.932	0.982	0.636	0.702	0.820
	500	100	0.900	0.956	0.982	0.514	0.580	0.678
$\nu_N = 0$	50	10	0.852	0.920	0.972	0.846	0.900	0.958
	100	10	0.842	0.912	0.972	0.782	0.858	0.948
	50	50	0.858	0.916	0.982	0.808	0.882	0.948
	100	50	0.872	0.940	0.980	0.750	0.824	0.916
	200	100	0.888	0.958	0.992	0.630	0.708	0.816
	500	100	0.880	0.948	0.990	0.494	0.556	0.656
$\nu_N = \mathcal{O}(T^{-1})$	50	10	0.902	0.946	0.978	0.796	0.898	0.960
	100	10	0.880	0.932	0.982	0.812	0.886	0.950
	50	50	0.840	0.910	0.976	0.816	0.880	0.952
	100	50	0.878	0.932	0.984	0.766	0.858	0.930
	200	100	0.902	0.944	0.988	0.750	0.838	0.938
	500	100	0.882	0.950	0.990	0.662	0.740	0.850

procedure is also robust toward t-distributed errors and serially correlated error terms. See Tables A4 and A5. Our uniform bootstrap procedure naturally adapts to the data, particularly to potential serial dependence and a varying degree of heterogeneity in the factor loadings.

Looking at the coverage of the asymptotic distribution of the PC-IFE estimator, we can observe under-coverage in all settings. Moreover, the coverage does not improve with increasing sample size. On the contrary, the coverage is worst for the setting with $N = 500$ and $T = 100$.

Table 4. Summary statistics and data sources of dependent and independent variables.

Variable	Description	Mean	Median	Min	Max	Data
Growth	Annual GDP growth per capita	2.96	2.54	–67.29	141.63	PWT
Young	Age dependency ratio	54.13	49.92	14.92	107.40	WDI
Fert	Fertility rate	3.23	2.69	1.09	7.7	WDI
Life	Life expectancy	68.30	71.21	26.17	84.36	WDI
Pop	Population growth	1.70	1.51	-6.54	19.14	PWT
Invpri	Price level of investment	0.54	0.50	0.01	7.98	PWT
Con	Consumption share	0.64	0.65	0.09	1.56	PWT
Gov	Government consumption share	0.17	0.17	0.01	0.75	PWT
Inv	Investment share	0.22	0.22	0.00	0.92	PWT

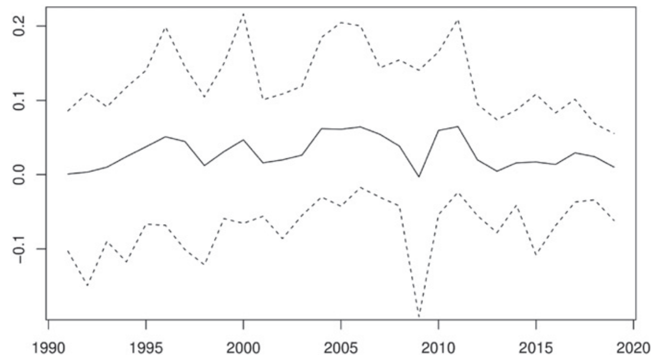

Figure 1. Time series of average annual real GDP growth rate per capita (solid line) and time series of 5% and 95%-quantiles (dashed lines).

5. Determinants of economic growth

The aim of this section is to show the performance of our estimator in empirical analysis. More precisely, we will apply our estimator in the analysis of the determinants of economic growth. We refer to Durlauf, Johnson, and Temple (2005) for a comprehensive review of the growth literature. While many studies focus on a cross-sectional analysis (see for instance Barro (1991)), there are also numerous studies employing a panel data approach with country-specific fixed effects (Acemoglu et al., 2019; Islam, 1995). However, Lu and Su (2016) argue that economic growth rates might not be solely determined by observable regressors, but could also be influenced by latent factors or shocks. Our projection-based interactive fixed effect estimator is well suited as it is flexible enough to model such latent factors.

The yearly data on GDP growth rates and the country-specific characteristics are obtained from the Penn World Table (PWT) and the World Bank World Development Indicators (WDI). Our sample contains 129 countries in a period from 1991–2019, $N = 129$ and $T = 29$. Countries with incomplete data availability or which did not exist yet in 1991 are excluded from our analysis. Our dependent variable is the real GDP growth rate per capita. The set of regressors is identical to the regressors in Lu and Su (2016). Summary statistics of all dependent and independent variables can be found in Table 4. Figure 1 shows the time series of the mean growth rates, averaged over all countries in our sample. We also visualize the time series of the cross-sectional 5% and 95%-quantiles of the growth rates in the same figure. For the time-invariant characteristics used for modeling the systematic part of the factor loadings, we take the longitude and latitude of the respective country[1].

We first fit our projection-based interactive fixed effects model using the complete sample of $N = 129$ countries. To be consistent with the simulation section, we use B-spline basis functions with $J_N = \lceil N^{1/3} 1.5 \rceil$. The estimation results can be found in Table 5. We report the estimated coefficients and the 95% confidence interval based on the cross-sectional bootstrap with 1000 bootstrap iterations. As

[1]Data obtained from developers.google.com/public-data/docs/canonical/countries_csv.

Table 5. Estimation results for the P-IFE and the PC-IFE based on the whole sample. *,** ,*** indicate the significance at 5%, 1% and 0.1% level.

	P-IFE		PC-IFE	
	Estimate	95%-CI	Estimate	95%-CI
Con	−0.0508**	[−0.0841, −0.0188]	−0.0518***	[−0.0757, −0.0279]
Gov	−0.0697*	[−0.1374, −0.0125]	−0.1389***	[−0.1830, −0.0948]
Inv	0.0129	[−0.0420, 0.0592]	0.0503**	[0.0178, 0.0828]
Invpri	0.0076	[−0.0101, 0.0240]	0.0067	[−0.0042, 0.0176]
Young	0.0007*	[0.0001, 0.0014]	0.0009***	[0.0005, 0.0013]
Fert	−0.0080	[−0.0182, 0.0025]	−0.0158***	[−0.0223, −0.0093]
Life	0.0001	[−0.0008, 0.0010]	0.0000	[−0.0006, 0.0006]
Pop	−0.0046	[−0.6075, 0.7627]	−0.2226	[−0.4696, 0.0244]

Table 6. Estimation results for the P-IFE and the PC-IFE based on OECD sample. *,** ,*** indicate the significance at 5%, 1% and 0.1% level.

	P-IFE		PC-IFE	
	Estimate	95%-CI	Estimate	95%-CI
Con	−0.1078***	[−0.1685, −0.0447]	−0.0725***	[−0.1048, −0.0402]
Gov	−0.0343	[−0.1859, 0.0713]	0.0110	[−0.0413, 0.0633]
Inv	0.1142*	[0.0117, 0.2226]	0.0405	[−0.0105, 0.0915]
Invpri	−0.0550*	[−0.0939, −0.0097]	−0.0082	[−0.0242, 0.0078]
Young	0.0011	[−0.0001, 0.0021]	0.0008	[−0.0001, 0.0017]
Fert	−0.0086	[−0.0289, 0.0081]	−0.0206*	[−0.0369, −0.0043]
Life	−0.0004	[−0.0031, 0.0012]	−0.0025***	[−0.0038, −0.0012]
Pop	−0.8195	[−1.6591, 0.2194]	0.3263	[−0.1794, 0.8320]

a comparison, we also report the estimated coefficients and confidence intervals following the PC-IFE approach of Bai (2009). We obtain negative significant coefficients for consumption share and government consumption share and a positive significant coefficient for the age dependency ratio at the 5% confidence level. These results are similar to the estimation based on the PC-IFE. However, investment share and fertility rate also become significant for the PC-IFE. The remaining variables are insignificant for both estimation procedures.

We now restrict our analysis to the subset of countries that are members of the OECD (Organisation for Economic Cooperation and Development). See Table 6 for the estimation results. Similar to the previous results, both approaches find a negative significant effect on the consumption share. However, our P-IFE identifies a positive effect on investment share and a negative effect on the investment price level. Both variables are insignificant for the PC-IFE approach. Moreover, PC-IFE additionally finds significant effects on the fertility rate and life expectancy.

6. Conclusions

In this article, a new estimator for the regression parameters in a panel data model with interactive fixed effects has been proposed. The main novelty of this approach is that factor loadings are approximated through nonparametric additive functions, and it is then possible to partial out the interactive effects. Therefore, the new estimator adopts the well-known partial least squares form, and there is no need to use iterative estimation techniques to compute it. It turns out that the limiting distribution of the estimator has a discontinuity when the variance of the idiosyncratic parameter of the factor loading approximation is near the boundaries. The discontinuity makes the usual "plug-in" inference based on the estimated asymptotic covariance matrix problematic since it can lead to either over- or under-coveraging probabilities. We show that non conservative uniformly valid inference can be achieved by cross-sectional bootstrap. A Monte Carlo study indicates good performance in terms of mean squared error and bootstrap coverage. We apply our methodology to analyze the determinants of growth rates in OECD countries.

Funding

Juan M. Rodriguez-Poo and Alexandra Soberon acknowledge financial support from the I+D+i project Ref. PID2019-105986GB-C22 financed by MCIN/AEI/10.13039/ 501100011033. In addition, this work was also funded by the I+D+i project Ref. TED2021-131763A-I00 financed by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationUE/PRTR. Georg Keilbar acknowledges gratefully the support from the Deutsche Forschungsgemeinschaft via the IRTG 1792 "High Dimensional Nonstationary Time Series". Weining Wang's research is supported through the project "IDA Institute of Digital Assets", CF166/15.11.2022, contract number CN760046/23.05.2023, financed under the Romania's National Recovery and Resilience Plan, Apel nr. PNRR-III-C9-2022-I8; and the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym: DIGITAL, Project No. 101119635. The research is partially supported by the ESRC (Grant Reference: ES/T01573X/1)

References

Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy* 127(1):47–100. doi: 10.1086/700936

Ahmad, I., Leelahanon, S., Li, Q. (2005). enquoteEfficient estimation of a semiparametric partially linear varying coefficient model. *The Annals of Statistics* 33:258–283.

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68(2):399–405. doi: 10.1111/1468-0262.00114

Bada, O., Liebl, D. (2014). phtt: panel data analysis with heterogeneous time trends in R. *Journal of Statistical Software* 59:1–33.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71(1):135–171. doi: 10.1111/1468-0262.00392

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

Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221. doi: 10.1111/1468-0262.00273

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

Bai, J., Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics* 8(1):53–80. doi: 10.1146/annurev-economics-080315-015356

Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics* 106(2):407. doi: 10.2307/2937943

Bernanke, B. S., Boivin, J., Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120(1):387–422. doi: 10.1162/qjec.2005.120.1.387

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica* 60(3):567. doi: 10.2307/2951584

Chen, X. (2007). Chaper 76: Large sample sieve estimation of semi-nonparametric models. In: *Handbook of Econometrics*, Vol. 6. Amsterdam: North Holland, pp. 5549–5632.

Cheng, T., Dong, C., Gao, J., Linton, O. (2024). GMM estimation for high-dimensional panel data models. *Journal of Econometrics* 244(1):105853. doi: 10.1016/j.jeconom.2024.105853

Chudik, A., Mohaddes, K., Pesaran, M. H., Raissi, M. (2017). Is there a debt-threshold effect on output growth?. *Review of Economics and Statistics* 99(1):135–150. doi: 10.1162/REST_a_00593

Chudik, A., Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics* 188(2):393–420. doi: 10.1016/j.jeconom.2015.03.007

Connor, G., Hagmann, M., Linton, O. (2012). Efficient semiparametric estimation of the fama-french model and extensions. *Econometrica* 80:713–754.

Connor, G., Linton, O. (2007). Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance* 14(5):694–717. doi: 10.1016/j.jempfin.2006.10.001

Desai, K. H., Storey, J. D. (2012). Cross-dimensional inference of dependent high-dimensional data. *Journal of the American Statistical Association* 107:135–151.

De Vos, I., Stauskas, O. (2024). Cross-section bootstrap for CCE regressions. *Journal of Econometrics* 240:105648.

Durlauf, S. N., Johnson, P. A., Temple, J. R. (2005). Growth econometrics. *Handbook of Economic Growth* 1:555–677.

Fan, J., Ke, Y., Liao, Y. (2021). Augmented factor models with applications to validating market risk factors and forecasting bond risk premia. *Journal of Econometrics* 222(1):269–294. doi: 10.1016/j.jeconom.2020.07.002

Fan, J., Liao, Y., Wang, W. (2016). Projected principal component analysis in factor models. *The Annals of Statistics* 44:219–254.

Fernández-Val, I., Gao, W. Y., Liao, Y., Vella, F. (2022). Dynamic heterogeneous distribution regression panel models, with an application to labor income processes. *arXiv preprint arXiv:2202.04154*.

Härdle, W., Liang, H., Gao, J. (2000). *Partially Linear Models*. Heidelberg: Physica-Verlag.
Islam, N. (1995). Growth empirics: a panel data approach. *The Quarterly Journal of Economics* 110(4):1127–1170. doi: 10.2307/2946651
Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal* 11(2):377–395. doi: 10.1111/j.1368-423X.2008.00243.x
Kock, A. B. (2016). Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models. *Journal of Econometrics* 195(1):71–85. doi: 10.1016/j.jeconom.2016.06.001
Kock, A. B., Tang, H. (2019). Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. *Econometric Theory* 35(2):295–359. doi: 10.1017/S0266466618000087
Liao, Y., Yang, X. (2018). *Uniform Inference and Prediction for Conditional Factor Models with Instrumental and Idiosyncratic Betas, Departamental Working Papes*, Rutgers University, Deparment of Economics, New Brunswick, NJ.
Lorentz, G. (1986). *Approximation of Functions*, 2nd ed. New York: Chelsea Publishing.
Lu, X., Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 190(1):148–175. doi: 10.1016/j.jeconom.2015.09.005
Lu, X., and Su, L. (2023). Uniform inference in linear panel data models with two-dimensional heterogeneity. *Journal of Econometrics* 235:694–719.
Ma, S., Linton, O., Gao, J. (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics* 222(1):295–323. doi: 10.1016/j.jeconom.2020.07.003
Menzel, K. (2021). Bootstrap with cluster–dependence in two or more dimensions. *Econometrica* 89(5):2143–2188. doi: 10.3982/ECTA15383
Moon, H. R., Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83(4):1543–1579. doi: 10.3982/ECTA9382
Moon, H. R., Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33:158–195.
Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4):967–1012. doi: 10.1111/j.1468-0262.2006.00692.x
Sarafidis, V., Wansbeek, TOM. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31(5):483–531. doi: 10.1080/07474938.2011.611458
Stock, J. H., Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460):1167–1179. doi: 10.1198/016214502388618960
Westerlund, J., Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics* 185(2):372–377. doi: 10.1016/j.jeconom.2014.09.014
Zhang, L., Zhou, W., Wang, H. (2021). A semiparametric latent factor model for large scale temporal data with heteroscedasticity. *Journal of Multivariate Analysis* 186:104786.