Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Robust Estimation of Large Panels with Factor Structures

Marco Avarucci & Paolo Zaffaroni

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/01621459.2022.2050244

Taylor & Francis
Taylor & Francis Group

Robust Estimation of Large Panels with Factor Structures

Marco Avarucci[a] and Paolo Zaffaroni[b,c]

[a]University of Glasgow, Glasgow, UK; [b]Imperial College Business School, London, UK; [c]University of Rome La Sapienza, Rome, Italy

ABSTRACT

This article studies estimation of linear panel regression models with heterogeneous coefficients using a class of weighted least squares estimators, when both the regressors and the error possibly contain a common latent factor structure. Our theory is robust to the specification of such a factor structure because it does not require any information on the number of factors or estimation of the factor structure itself. Moreover, our theory is efficient, in certain circumstances, because it nests the GLS principle. We first show how our unfeasible weighted-estimator provides a bias-adjusted estimator with the conventional limiting distribution, for situations in which the OLS is affected by a first-order bias. The technical challenge resolved in the article consists of showing how these properties are preserved for the feasible weighted estimator in a double-asymptotics setting. Our theory is illustrated by extensive Monte Carlo experiments and an empirical application that investigates the link between capital accumulation and economic growth in an international setting. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2020
Accepted February 2022

KEYWORDS
Factor structure; GLS; Panel; Robustness; Weighted least squares estimation

1. Introduction

This article studies estimation of linear panel regression models with heterogeneous coefficients using a class of weighted least squares estimators, when both the regressors and the error possibly contain a common latent factor structure. Factor models represent one of the most popular and successful ways to capture cross-sectional and temporal dependence, especially when facing a large number of units (N) and time periods (T). However, in our context, the possibility of a common factor structure in both regressors and error leads to an endogeneity problem, making estimation of panel data models by ordinary least squares (OLS) invalid.

In principle, the endogeneity would be trivially resolved if one could hypothetically observe the latent common factors and use them in the OLS regression. Second, an efficiency improvement over the OLS estimator can be achieved if the true variance-covariance matrix of the idiosyncratic regression error is known, leading to the generalized least squares (GLS) estimator. Both of these two steps are not feasible in practice, as we do not observe the latent common factors nor we have the covariance matrix of the idiosyncratic error.

The contribution of this article is to show how to achieve these two goals, despite the latency of the factors and of the variance-covariance matrix of the idiosyncratic regression error. Unlike other approaches, we do not advocate preliminary estimation of the factor structure and, instead, we treat factors and their loadings as nuisance parameters rather than objects of interest. In particular, we construct a feasible weighted estimator that resolves the endogeneity without *any* knowledge of the common factors, qualifying our methodology as *robust*. This appears to be the strongest feature of our methodology.

Moreover, the statistical properties of our estimator are preserved under several circumstances of interest that can arise in the data.[1] This is relevant as many existing estimation procedures designed to tackle the endogeneity problem in panel regressions, such as the popular common correlated effects (hereafter CCE) estimator of Pesaran (2006), become invalid, that is, first-order biased, when any of the above circumstances apply. Moreover, the large majority of the contributions do not allow heterogeneity of the regression coefficients (see, e.g., Bai 2009). Resolving these problems, that is constructing an estimator robust to endogeneity when alternative procedures fail and allowing for heterogeneous slope coefficients, represents our advancement in the literature of panel estimation with interactive fixed effects.

Our work also advances the literature of GLS estimation along two dimensions. In fact, first, under specific circumstances on the idiosyncratic error, the weighting scheme adopted by our estimator achieves full efficiency, meaning that it is asymptotically equivalent to a feasible GLS estimator. We qualify this property as *quasi-efficiency*, given that these circumstances on the idiosyncratic error are not generally warranted. Second, our work also contributes to the literature that demonstrates how efficient estimation techniques not only lead to an improvement of precision but, most importantly, may resurrect the required asymptotic properties, in terms of bias, rate of convergence, and distribution, when these fails in inefficient approaches (see Phillips 1991; Robinson and Hidalgo 1997).

CONTACT Paolo Zaffaroni p.zaffaroni@imperial.ac.uk Imperial College Business School, London, UK.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

[1]These includes the case of of weak endogeneity, when the loadings to the latent factors in the regressors and in the errors are cross-correlated, when the latent factors' space cannot be recovered from cross-sectional averages of the observable variates, a special case of which being when the number of latent factors is arbitrarily larger than the number of heterogenous regressors.

Section C of the supplementary materials offers a complete overview of these two vast streams of literature on estimation of panel with interactive fixed effects and efficient estimation of regression models.

We first consider the *unfeasible* weighted least squares estimator for the regression coefficients, assuming that both the covariance matrix of the idiosyncratic error and the latent factors are known. This unfeasible estimator is asymptotically equivalent to the (unfeasible) GLS estimator applied to the variates after these have been projected on the space orthogonal to the common factors, according to the Frisch–Waugh–Lovell theorem.[2] It turns out that, regardless of the possibility of endogeneity (i.e., when regressors and error are correlated), the unfeasible estimator is $T^{1/2}$-consistent and asymptotically normal. The unfeasible weighted estimator does not only represent a more efficient estimator than OLS but provides an *automatic* biased-adjusted estimator with desirable asymptotic properties, when OLS is invalid due to the failure of strict exogeneity. This result is due to an important insight, namely that the column space of the weighting matrix is, asymptotically in T, orthogonal to the column space of the matrix of common factors, that is, it spans the null space of the common latent factors. Most importantly, for our theory to go through, the norm of the product between the weighting matrix and the factors' matrix is manifested at the fast rate $O(T^{-1/2})$.

The substantial challenge arises when considering a feasible version of the weighted least squares estimator. Our main insight is that the feasible weighting matrix can be constructed using the sample *cross-sectional* second-moment matrix of the OLS residuals. Lack of consistency of the OLS estimator for the regression coefficients would suggest that our feasible weighting scheme is invalid, and, indeed, the feasible weighting matrix does not converge to the unfeasible weighting matrix (in the sense of element-by-element convergence).

The *surprising* crucial result established here is that the feasible weighting matrix maintains the asymptotic orthogonality to the latent common factors, established for the unfeasible weighting matrix, and at the same $O(T^{-1/2})$-rate. As a consequence, we are able to show that the feasible estimator for the regression coefficients is $T^{1/2}$-consistent and asymptotically normal, as *both* N, T diverge to infinity, making a complete use of the panel dimension. This holds under the same circumstances for which the OLS estimator is first-order biased. This is the most intriguing aspect of our theory. Finally, an iterative evaluation of our feasible estimator permits getting close to the unfeasible weighted least squares estimator, and thus, to GLS efficiency, under certain homogeneity assumptions for the idiosyncratic error covariance matrix.

Our asymptotic distribution theory requires $T^2/N \to 0$, whereas the milder $T/N \to 0$ ensures consistency. This relative speed spells out a neat dichotomy in terms of the role of T and N: the faster rate of divergence for N is asked to estimate accurately the weighting matrix required by our estimator, which in turn mitigates the asymptotic bias. Instead, the slower divergence of T controls the asymptotic variance of the weighted least squares estimator, dictating ultimately the estimator's convergence to its limiting distribution.

This article proceeds as follows. Section 2 illustrates the panel model and the estimators studied in the article, with the regularity assumptions described in Section 3. Our main asymptotic results are in Section 4. Monte Carlo experiments and an empirical application, which investigates whether capital accumulation enhances economic growth for a panel of OCED and non-OECD countries, are presented in Section 5. Section 6 concludes. The technical proofs, the literature survey, and further material, are relegated to the supplemental materials.

2. Methodology: Model and Estimators

The notation adopted throughout the article is collected in Appendix B, supplementary materials but we also define quantities throughout the article when needed. Assume that the observed variables obey a linear regression model with S common regressors $\mathbf{d}_t = [d_{t1}, \dots, d_{tS}]'$ and K heterogeneous regressors $\mathbf{x}_{it} = [\mathrm{x}_{it1} \dots \mathrm{x}_{itK}]'$. Following the convenient specification put forward by Pesaran (2006), the model for the ith unit can be expressed, in matrix form, as

$$\mathbf{y}_i = \mathbf{D}\boldsymbol{\alpha}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i, \tag{2.1}$$

for an observed $T \times 1$ vector $\mathbf{y}_i = [\mathrm{y}_{i1}, \dots, \mathrm{y}_{iT}]'$, an observed $T \times S$ matrix $\mathbf{D} = [\mathbf{d}_1 \dots \mathbf{d}_T]'$ of common regressors, an observed $T \times K$ matrix $\mathbf{X}_i = [\mathbf{x}_{i1} \dots \mathbf{x}_{iT}]'$ of heterogenous regressors, an unobserved $T \times 1$ vector $\mathbf{u_i} = [\mathrm{u}_{i1} \dots \mathrm{u}_{iT}]'$ of regression errors, and slope coefficients $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$. The presence of common observed regressors are not necessary for our results, and, in fact, complicate somewhat the asymptotic analysis but could be relevant in practical applications, so we provide a unified theory. In turn, the error vector satisfies the following latent factor structure:

$$\mathbf{u}_i = \mathbf{F}\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \qquad \text{with } \mathbb{E}\boldsymbol{\varepsilon}_i = \mathbf{0} \text{ and } \Xi_i := \mathbb{E}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i', \tag{2.2}$$

for an unobserved $M \times 1$ vector of factor loadings $\mathbf{b}_i$, an unobserved $T \times M$ matrix of common factors $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_T]'$, and an unobserved $T \times 1$ vector of idiosyncratic errors $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1} \dots \varepsilon_{iT}]'$. We postulate that the heterogeneous regressors satisfy

$$\mathbf{X}_i = \mathbf{D}\Delta_i + \mathbf{F}\Gamma_i + \mathbf{V}_i, \tag{2.3}$$

for an unobserved $S \times K$ matrix of factor loadings $\Delta_i = [\boldsymbol{\delta}_{i1} \dots \boldsymbol{\delta}_{iS}]'$ with $\boldsymbol{\delta}_{il} = [\delta_{il1} \dots \delta_{ilK}]'$, an unobserved $M \times K$ matrix of factor loadings $\Gamma_i = [\boldsymbol{\gamma}_{i1} \dots \boldsymbol{\gamma}_{iM}]'$ with $\boldsymbol{\gamma}_{il} = [\gamma_{il1} \dots \gamma_{ilK}]'$, and an unobserved $T \times K$ matrix of idiosyncratic errors $\mathbf{V}_i = [\mathbf{v}_{i1} \dots \mathbf{v}_{iT}]'$ with $\mathbf{v}_{it} = [v_{it1}, \dots v_{itK}]'$. Formulation (2.3) can be substantially relaxed, for instance without imposing a linear factor structure in the $\mathbf{X}_i$, although certainly at the cost of less primitive assumptions. However, (2.3) represents a simple yet powerful way to allow dependence between the regressors and the errors of (2.1) through the latent factor structure.[3] In

[2]Our simulation exercise shows that these two unfeasible estimators are asymptotically equivalent when T is large.

[3]As explained below, by exploiting the linearity assumption embedded in (2.2)–(2.3), the weighting scheme of our FWLS estimator implies that the idiosyncratic components of $\mathbf{X}_i$ and $\mathbf{u}_i$, that is $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$, respectively, dominate the limiting properties of the FWLS estimator, as the terms involving $\mathbf{F}$ become asymptotically negligible.

this respect, an important generalisation consists of allowing the $\mathbf{X}_i$ to depend *nonlinearly* on the $\mathbf{F}$, a simple example of which being when the $\mathbf{X}_i$ are function of polynomials in the $\mathbf{F}$ (see Appendix F of the supplementary materials). The maintained condition here is that K, S, and M *do not vary* with T and N. However, most importantly, one does not need to know *virtually* anything about M, not even an upper bound, as our methodology will work regardless of whether M is either *smaller, equal, or bigger* than K.[4] Although model (2.1) is written as a single regression across time for a given unit i, we assume that a panel of observations $\{\mathbf{y}_1 \dots \mathbf{y}_N, \mathbf{X}_1 \dots \mathbf{X}_N\}$ is available and fully used within our methodology.

Our main objective is to estimate the heterogeneous slope coefficients $\boldsymbol{\beta}_i$ of (2.1). We defer discussion on estimation of the slope coefficients $\boldsymbol{\alpha}_i$ of the common regressors to Section E of the supplementary materials. As explained below, throughout our analysis we always net out the effect of the common regressors by premultiplying the data by $\mathscr{M}_{\mathbf{D}} := \mathbf{I}_T - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{+}\boldsymbol{D}'$, where $\mathbf{I}_T$ is the identity matrix of dimension T and $\mathbf{A}^+$ denotes the Moore–Penrose inverse of any matrix $\mathbf{A}$. This allows us to avoid making any assumptions on $\boldsymbol{\Delta}_i$. Hence, one obtains

$$\mathscr{M}_{\mathbf{D}}\mathbf{y}_i = \mathscr{M}_{\mathbf{D}}\mathbf{X}_i\boldsymbol{\beta}_i + \mathscr{M}_{\mathbf{D}}\mathbf{u}_i. \tag{2.4}$$

We consider three different estimators for the parameters $\boldsymbol{\beta}_i$, namely the OLS and the *unfeasible* and *feasible* weighted least squares estimators, hereafter indicated as the WLS and FWLS estimators, respectively. Here we explain how to construct the estimators and, subsequently, provide the intuition on how the weighted estimators mitigate the endogeneity bias and, possibly, achieve efficiency.

Regarding the OLS estimator for $\boldsymbol{\beta}_i$,

$$\hat{\boldsymbol{\beta}}_i^{\mathrm{OLS}} := (\mathbf{X}_i'\mathscr{M}_{\mathbf{D}}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathscr{M}_{\mathbf{D}}\mathbf{y}_i. \tag{2.5}$$

We now describe the weighted least squares estimators.

Proceeding along the lines of (Magnus and Neudecker 1988, sec. 11—chap. 13) we define the WLS estimator, whose weighting matrix is based on taking the generalized inverse of $\mathscr{M}_{\mathbf{D}}\mathbf{S}_i\mathscr{M}_{\mathbf{D}}$ with $\mathbf{S}_i := \mathbf{F}\boldsymbol{B}\mathbf{F}' + \boldsymbol{\Xi}_i$, as follows[5]

$$\hat{\boldsymbol{\beta}}_i^{\mathrm{WLS}} := \left(\mathbf{X}_i'\mathscr{M}_{\mathbf{D}}\left(\mathscr{M}_{\mathbf{D}}\mathbf{S}_i\mathscr{M}_{\mathbf{D}}\right)^{+}\mathscr{M}_{\mathbf{D}}\mathbf{X}_i\right)^{-1} \mathbf{X}_i'\mathscr{M}_{\mathbf{D}}(\mathscr{M}_{\mathbf{D}}\mathbf{S}_i\mathscr{M}_{\mathbf{D}})^{+}\mathscr{M}_{\mathbf{D}}\mathbf{y}_i. \tag{2.6}$$

The matrix $(\mathscr{M}_{\mathbf{D}}\mathbf{S}_i\mathscr{M}_{\mathbf{D}})$ is singular, leading to the complicate expression for $\hat{\boldsymbol{\beta}}_i^{\mathrm{WLS}}$ of (2.6), but, by Lemma K.2 in the supplemental materials , $(\mathscr{M}_{\mathbf{D}}\mathbf{S}_i\mathscr{M}_{\mathbf{D}})^{+} = \mathbf{D}_\perp\left(\mathbf{D}_\perp'\mathbf{S}_i\mathbf{D}_\perp\right)^{-1}\mathbf{D}_\perp'$, where $\mathbf{D}_\perp$ spans the null space of $\mathbf{D}$, namely the $T \times (T-S)$ full rank matrix such that $\mathscr{M}_{\mathbf{D}} = \mathbf{D}_\perp\mathbf{D}_\perp'$, where $\mathbf{D}_\perp'\mathbf{D}_\perp = \mathbf{I}_{T-S}$.[6]

By substitution, setting for simplicity,

$$\begin{aligned} \boldsymbol{y}_i &:= \mathbf{D}_\perp'\mathbf{y}_i, \quad \boldsymbol{\mathcal{X}}_i := \mathbf{D}_\perp'\mathbf{X}_i, \quad \boldsymbol{\epsilon}_i := \mathbf{D}_\perp'\boldsymbol{\varepsilon}_i, \\ \boldsymbol{\mathcal{F}} &:= \mathbf{D}_\perp'\mathbf{F}, \quad \boldsymbol{u}_i := \mathbf{D}_\perp'\mathbf{u}_i, \quad \boldsymbol{\mathcal{V}}_i := \mathbf{D}_\perp'\mathbf{V}_i, \end{aligned} \tag{2.7}$$

one obtains

$$\begin{aligned} \hat{\boldsymbol{\beta}}_i^{\mathrm{WLS}} &= \left(\mathbf{X}_i'\mathbf{D}_\perp\left(\mathbf{D}_\perp'\mathbf{S}_i\mathbf{D}_\perp\right)^{-1}\mathbf{D}_\perp'\mathbf{X}_i\right)^{-1} \\ &\quad \mathbf{X}_i'\mathbf{D}_\perp\left(\mathbf{D}_\perp'\mathbf{S}_i\mathbf{D}_\perp\right)^{-1}\mathbf{D}_\perp'\mathbf{y}_i \\ &= \left(\boldsymbol{\mathcal{X}}_i'\boldsymbol{\mathcal{S}}_i^{-1}\boldsymbol{\mathcal{X}}_i\right)^{-1}\boldsymbol{\mathcal{X}}_i'\boldsymbol{\mathcal{S}}_i^{-1}\boldsymbol{y}_i, \end{aligned} \tag{2.8}$$

where we set

$$\boldsymbol{\mathcal{S}}_i := \mathbf{D}_\perp'\mathbf{S}_i\mathbf{D}_\perp = \boldsymbol{\mathcal{F}}\mathbf{B}\boldsymbol{\mathcal{F}}' + \boldsymbol{\mathfrak{X}}_i \quad \text{with} \quad \boldsymbol{\mathfrak{X}}_i := \mathbf{D}_\perp'\boldsymbol{\Xi}_i\mathbf{D}_\perp. \tag{2.9}$$

This means that the WLS has now the more conventional expression (2.8) of a weighted estimator for the model

$$\boldsymbol{y}_i = \boldsymbol{\mathcal{X}}_i\boldsymbol{\beta}_i + \boldsymbol{u}_i, \qquad \text{with} \qquad \boldsymbol{u}_i = \boldsymbol{\mathcal{F}}\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2.10}$$

without involving Moore–Penrose matrices, that is, with respect to a nonsingular weighting matrix $\boldsymbol{\mathcal{S}}_i^{-1}$.[7] Premultiplying the data by $\mathbf{D}_\perp'$ reduces the sample size by S time series observations since now the $\boldsymbol{y}_i$ and the $\boldsymbol{\mathcal{X}}_i$ have $T - S$ rows. Likewise, considering again model (2.10), an equivalent representation to (2.5) is $\hat{\boldsymbol{\beta}}_i^{\mathrm{OLS}} = (\boldsymbol{\mathcal{X}}_i'\boldsymbol{\mathcal{X}}_i)^{-1}\boldsymbol{\mathcal{X}}_i'\boldsymbol{y}_i$.

For the special case $S = 0$, that is when the common regressors $\mathbf{D}$ are absent from the model, the above formulas continue to be valid, by setting $\mathscr{M}_{\mathbf{D}} = \mathbf{D}_\perp = \mathbf{I}_T$, yielding, for example, $\hat{\boldsymbol{\beta}}_i^{\mathrm{OLS}} = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i$ and $\hat{\boldsymbol{\beta}}_i^{\mathrm{WLS}} = (\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{y}_i$.

Considering model (2.10), our proposed FWLS estimator is given by

$$\hat{\boldsymbol{\beta}}_i^{\mathrm{FWLS}} := \left(\boldsymbol{\mathcal{X}}_i'\hat{\boldsymbol{\mathcal{S}}}_N^{-1}\boldsymbol{\mathcal{X}}_i\right)^{-1}\boldsymbol{\mathcal{X}}_i'\hat{\boldsymbol{\mathcal{S}}}_N^{-1}\boldsymbol{y}_i, \tag{2.11}$$

where

$$\hat{\boldsymbol{\mathcal{S}}}_N := N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_i', \quad \text{with } \hat{\boldsymbol{u}}_i := \boldsymbol{y}_i - \boldsymbol{\mathcal{X}}_i\hat{\boldsymbol{\beta}}_i^{\mathrm{OLS}} = \mathscr{M}_{\boldsymbol{\mathcal{X}}_i}\boldsymbol{u}_i. \tag{2.12}$$

Again, the special case for the FWLS estimator when $S = 0$ in (2.1), that is when no common regressors are present, is obtained replacing $\mathbf{D}_\perp$ with $\mathbf{I}_T$ in (2.11) and (2.12). Finally, our methodology also applies to the case of cross-sectional regressions with factor structure (see, Andrews 2005).[8]

3. Assumptions

We now present our assumptions, which, thanks to the specification of model (2.1)–(2.3), appear relatively primitive. Recall that, throughout our exposition, we always assume that $N > T$ holds, implied by condition $T/N = o(1)$, as demanded by our theory.

[4]Our results can be extended to the case of M (slowly) increasing with T as a suitable rate.

[5]To motivate the expression for $\mathbf{S}_i$, notice that it can be obtained as the limit of the cross-sectional average of the $\mathbb{E}(\mathbf{u}_i\mathbf{u}_i'|\{\mathbf{z}_t\})$, as $N \to \infty$, for the special homogeneity case $\Xi_i = \Xi$. Here $\boldsymbol{B}$ defines the limit of $N^{-1}\sum_{i=1}^{N}\mathbf{b}_i\mathbf{b}_i'$, which we assume to be positive definite. Finally, pre and postmultiplication by $\mathscr{M}_{\mathbf{D}}$ is a consequence of netting out the effect of the common regressors $\mathbf{D}$.

[6]Note that $\mathbf{D}_\perp$ is assumed full rank when T is large enough, and $\mathbf{S}_i \geq \Xi_i > 0$, because the latter has its minimum eigenvalue (uniformly) bounded away from zero, ensuring that $(\mathbf{D}_\perp'\mathbf{S}_i\mathbf{D}_\perp)^{-1}$ is always well-defined a.s., despite $\mathbf{S}_i$ being random.

[7]Given the analogies between GLS and Seemingly Unrelated Regression (SUR) estimator, one can in principle construct a SUR-type WLS estimator (see Appendix H, supplementary materials for details).

[8]Details are available upon request.

Assumption 3.1 (idiosyncratic innovation ε_{it}). The $N \times 1$ vector $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t} \dots \varepsilon_{Nt})'$ satisfies the following equation:

$$\boldsymbol{\varepsilon}_t = \boldsymbol{R}\boldsymbol{a}_t, \qquad \text{for} \quad t = 1, \dots T, \tag{3.1}$$

where the $N \times N$ matrix of constants $\boldsymbol{R} = [r_{ij}]$, with r_{ij} not varying with N for given i and j, has column- and row-norms satisfying $\|\boldsymbol{R}\|_{\text{row}} + \|\boldsymbol{R}\|_{\text{col}} < \infty$, $\min_i \sum_{j=1}^{N} |r_{ij}| > \kappa$ for some $0 < \kappa < \infty$ (not always the same), and the elements of the $N \times 1$ vector $\boldsymbol{a_t} = (a_{1t}, a_{2t}, \dots, a_{Nt})'$ follow a linear process,

$$a_{it} = \sum_{s=0}^{\infty} \phi_{is}\eta_{i,t-s}, \quad \max_i \sum_{s=0}^{\infty} s^2|\phi_{is}| < \infty, \quad \text{with} \quad \phi_{i0} = 1, \tag{3.2}$$

where the elements of $\boldsymbol{\eta}_t = (\eta_{1t}, \dots, \eta_{Nt})'$ make an iid sequence across i and t with $\mathbb{E}\eta_{it} = 0$, $\mathbb{E}|\eta_{it}|^{12} < \infty$ and the smallest eigenvalue of their covariance matrix satisfying $\lambda_1(\mathbb{E}\boldsymbol{\eta}_t\boldsymbol{\eta}_t') \geq \kappa > 0$. Moreover, for every complex number z satisfying $|z| \leq 1$,

$$\min_i |\phi_i(z)| > \kappa, \qquad \text{where} \quad \phi_i(z) := \sum_{s=0}^{\infty} \phi_{is}z^s. \tag{3.3}$$

Remark 3.1. Assumption 3.1 is similar to Assumptions 1 and 2 in Pesaran and Tosetti (2011), and, with some variations, this form of cross-sectional and time dependence has been adopted also by Moon and Weidner (2015, 2017), and Onatski (2015). The assumption above turns out to be extremely convenient for establishing the asymptotic distribution of the WLS and FWLS estimators along the lines of Theorem 1 in Robinson and Hidalgo (1997). Notice that it implies that the sequence of the $\boldsymbol{R}$ matrices, as N increases, makes a nested family of matrices.

Remark 3.2. Assumption 3.1 implies that, for every $2 \leq h, \ell \leq 12$,

$$\max_{i_1} \max_{t_1} \sum_{i_2 \dots i_\ell = 1}^{N} \sum_{t_2 \dots t_h = 1}^{T} |\text{cum}_h(\varepsilon_{i_1 t_1}, \varepsilon_{i_2 t_2} \dots, \varepsilon_{i_\ell t_h})| < \infty,$$

where the $\text{cum}_h(\varepsilon_{i_1 t_1}, \varepsilon_{i_2 t_2} \dots, \varepsilon_{i_\ell t_h})$ are the cumulants of order h of $\varepsilon_{i_1 t_1}, \varepsilon_{i_2 t_2} \dots, \varepsilon_{i_\ell t_h}$.

Remark 3.3. By (Brockwell and Davis 1991, Lemma 4.5.3) (3.3) implies that the eigenvalues of the covariance matrices of $\boldsymbol{a_j} = (a_{j1}, \dots a_{jT})'$ are bounded and greater than κ for every j. Easy calculations give $\Xi_i = \sum_{j=1}^{N} r_{ij}^2 \mathbb{E}\boldsymbol{a}_j\boldsymbol{a}_j'$, implying that its smallest and largest eigenvalues satisfy $\min_i \lambda_1(\Xi_i) > \kappa$ and $\max_i \lambda_T(\Xi_i) < \infty$.

Assumption 3.2 (regressor innovation $\mathbf{V_i}$). The sequence $\{v_{itk}\}$ is covariance stationary, with zero mean, $\max_i \max_k \max_t \mathbb{E}|v_{itk}|^{14} < \infty$, and they satisfy, for every $2 \leq h, \ell, s \leq 14$, and $2 \leq j \leq h$

$$\max_{k_1 \dots k_s} \max_{i_1} \max_{t_1} \sum_{i_2 \dots i_\ell = 1}^{N} \sum_{t_2 \dots t_h = 1}^{T} (1 + t_j^2)|\text{cum}_h(v_{i_1 t_1 k_1}, \dots, v_{i_\ell t_h, k_s})| < \infty,$$

where the $\text{cum}_h(v_{i_1 t_1 k_1}, \dots, v_{i_\ell t_h, k_s})$ are the cumulants of order h of $v_{i_1 t_1 k_1}, \dots, v_{i_\ell t_h, k_s}$. Moreover, $\min_i \lambda_1\left(\mathbb{E}\mathbf{v}_{it}\mathbf{v}_{it}'\right) > \kappa$, where $\mathbf{v}_{it}' = (v_{it1}, \dots, v_{itK})$.

Remark 3.4. Assumption 3.2 implies that $T^{-1}\mathbf{V}_{\boldsymbol{i}}'\mathbf{V}_{\boldsymbol{i}} \xrightarrow{p} \Sigma_{\mathbf{V}_i'\mathbf{V}_i}$, where $\xrightarrow{p}$ means convergence in probability, with $\min_i \lambda_1\left(\Sigma_{\mathbf{V}_i'\mathbf{V}_i}\right) > \kappa$. It follows that $\max_i \left\|\left(\mathbf{V}_{\boldsymbol{i}}'\mathbf{V}_{\boldsymbol{i}}/T\right)^{-1}\right\| = O_p(1)$ as T diverges, where $\|\cdot\|$ is the Frobenius norm.

Remark 3.5. The v_{itk} can be interpreted as the high-rank components of the regressors x_{itk}, adopting Moon and Weidner (2015) terminology, as opposed to the $\mathbf{D}$, which represents the low-rank components. For instance, if for each k, the v_{itk} are generated as ε_{it} in Assumption 3.1, one obtains $\boldsymbol{V_k} = [v_{itk}]_{t,i=1}^{T,N} = O_p(\sqrt{\max(N,T)})$ for every k (see the discussion in Moon and Weidner (2015, Appendix 1) and Onatski (2015)). In contrast, $(\sum_{i=1}^{N} \|\mathbf{D}\|^2)^{1/2} = (N\|\mathbf{D}\|^2)^{1/2} = O_p(\sqrt{NT})$.

Assumption 3.3 (latent and observed factors). Set $\mathbf{Z} := [\mathbf{D}, \mathbf{F}] = [z_{t,r}]$ for $1 \leq t \leq T$ and $1 \leq r \leq M + S < \infty$. Then,

$$\frac{\mathbf{Z}'\mathbf{Z}}{T} \xrightarrow{p} \Sigma_{\mathbf{Z}'\mathbf{Z}}, \quad \text{with} \quad \Sigma_{\mathbf{Z}'\mathbf{Z}} := \begin{bmatrix} \Sigma_{\mathbf{D}'\mathbf{D}} & \Sigma_{\mathbf{D}'\mathbf{F}} \\ \Sigma_{\mathbf{F}'\mathbf{D}} & \Sigma_{\mathbf{F}'\mathbf{F}} \end{bmatrix} > 0, \tag{3.4}$$

and $\Sigma_{\mathbf{D}'\mathbf{D}} > 0$, $\Sigma_{\mathbf{F}'\mathbf{F}} > 0$, where $>$ means positive definiteness. Moreover, we assume $\mathbb{E}\|\mathbf{z}_t\|^4 < \infty$, where $\mathbf{z}_t = [z_{t,1}, \dots, z_{t,M+S}]'$.

Remark 3.6. Equation (3.4) implies that $\left\|\left(T^{-1}\mathbf{F}'\mathscr{M}_{\mathbf{D}}\mathbf{F}\right)^{-1}\right\| = O_p(1)$ (see Lütkepohl 1996, Result (4), sec. 9.11.2).

Remark 3.7. The factors $\mathbf{D}$ and $\mathbf{F}$ are allowed to be cross-correlated as well as serially correlated, although not perfectly collinear. For instance, the joint dynamics of $\mathbf{Z}$ could be described by a multivariate stationary ARMA. We are ruling out trending behaviours in $\mathbf{D}$ and $\mathbf{F}$, although our results can be suitably modified to accommodate trends.

Assumption 3.4 (regressors). For every i, the $T \times (K + S)$ matrix $\mathbf{Z_i} := [\mathbf{D}, \mathbf{X_i}]$ has full column rank. Moreover, for T, N satisfying $1/T + T/N = o(1)$, the matrix $N^{-1}\sum_{i=1}^{N} \mathscr{M}_{\mathbf{Z_i}}\mathbf{u}_i\mathbf{u}_i'\mathscr{M}_{\mathbf{Z_i}}$ has full rank with probability approaching one.

Remark 3.8. Assumption 3.4 requires enough cross-sectional heterogeneity of the $\mathbf{X}_{\boldsymbol{i}}$ across units. Simple manipulations show that, with probability approaching one,

$$\hat{\boldsymbol{\mathcal{S}}}_N = \mathbf{D}_\perp'\left(\frac{1}{N}\sum_{i=1}^{N} \mathscr{M}_{\mathbf{Z_i}}\mathbf{u}_i\mathbf{u}_i'\mathscr{M}_{\mathbf{Z_i}}\right)\mathbf{D}_\perp = \frac{1}{N}\sum_{i=1}^{N} \mathscr{M}_{\boldsymbol{\mathcal{X}}_i}\boldsymbol{u}_i\boldsymbol{u}_i{}'\mathscr{M}_{\boldsymbol{\mathcal{X}}_i} > 0,$$

with $\hat{\boldsymbol{\mathcal{S}}}_N$ as defined in (2.12).

Assumption 3.5 (loadings). $\Gamma_{\mathbf{i}}$ and $\mathbf{b}_i$ are nonrandom such that $\max_i \|\Gamma_{\mathbf{i}}\| < \infty$ and $\max_i \|\mathbf{b}_i\| < \infty$ and, as $N \to \infty$,

$$\boldsymbol{B_N} := \frac{1}{N}\sum_{i=1}^{N} \mathbf{b}_i\mathbf{b}_i' \to \mathbf{B} > 0. \tag{3.5}$$

Additionally,

$$\mathbf{A}_N := \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{I}_M - \Gamma_{\mathbf{i}}\Psi_{\mathbf{i}}{}^{-1}\Gamma_{\mathbf{i}}{}'\frac{\mathbf{F}'\mathscr{M}_{\mathbf{D}}\mathbf{F}}{T}\right)$$
$$\mathbf{b}_i\mathbf{b}_i'\left(\mathbf{I}_M - \frac{\mathbf{F}'\mathscr{M}_{\mathbf{D}}\mathbf{F}}{T}\Gamma_{\mathbf{i}}\Psi_{\mathbf{i}}{}^{-1}\Gamma_{\mathbf{i}}{}'\right) > 0 \qquad (3.6)$$

with

$$\Psi_i := \Gamma_{\mathbf{i}}{}'\frac{\mathbf{F}'\mathscr{M}_{\mathbf{D}}\mathbf{F}}{T}\Gamma_{\mathbf{i}} + \Sigma_{\mathbf{V}_i'\mathbf{V}_i}. \qquad (3.7)$$

Remark 3.9. Assumption 3.5 characterizes the behavior of the finite-dimensional matrices, namely $\mathbf{A}_N$ and $\boldsymbol{B}_N$, that play a key role in the behavior of the weighting matrix of the FWLS estimator. We differ from the usual weighted OLS estimation setup, as the FWLS weighting matrix can never be a diagonal matrix, unless $\mathbf{b}_i = \mathbf{0}_M$, that is an $M \times 1$ vector of zeros.

Condition (3.5) implies that the factor structure (2.2) is strong, as defined in Pesaran and Tosetti (2011). This is commonly assumed in the factor models literature. However, there is a major difference between the factor models literature and this article. In that literature, the goal is to estimate the factors, and their loadings, and to do so one needs to eliminate the effect of the idiosyncratic component, by means of the averaging induced by the (static or dynamic) principal components. In this article, the aim is the opposite, namely to get rid of the factors, because they represent the source of the bias in the OLS estimator for $\boldsymbol{\beta}_{i0}$. This is based on projection arguments, rather than on averaging arguments which would not work here, exploiting the properties of the weighting matrix $\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}{}^{-1}$, which we demonstrate to (asymptotically) span the space *orthogonal* to the factors. This is the key result that drives the asymptotic results for the FWLS estimator.

Remark 3.10. The technical condition (3.6) is used in the proof of Theorem 4.2. As shown in Appendix O.2, Lemma O.2(iv), supplementary materials the matrices in brackets are of full rank. Hence, (3.6) will be satisfied when there is enough cross-sectional heterogeneity in the sample. Finally, our results will not change if loadings are assumed random and cross-sectionally independent from other parameters.

Assumption 3.6 (independence). The (d_{tp}, f_{tq}), v_{isk}, ε_{ju} are mutually independent for every i, j and t, s, u, and p, q, k.

Remark 3.11. We are not allowing for any correlation between any entries of $\boldsymbol{\varepsilon_j}$ and $\mathbf{X_i}$. This rules out the possibility that $\mathbf{X_i}$ contains a weakly exogenous component, and in this respect, we are similar to Pesaran (2006) and Bai (2009). The implications from generalizing this assumption, in particular when considering dynamic panels in which one element of $\mathbf{X_i}$ represents the lagged dependent variable, are briefly discussed in Section G of the supplementary materials.

Remark 3.12. Assumptions 3.2, 3.3, and 3.6 and Remark 3.6 imply that $T^{-1}\mathbf{X_i'X_i} \xrightarrow{p} \Sigma_{\mathbf{X_i'X_i}} > 0$ and $T^{-1}\mathbf{X_i'}\mathscr{M}_{\mathbf{D}}\mathbf{X_i} \xrightarrow{p} \Sigma_{\mathbf{X_i'}\mathscr{M}_{\mathbf{D}}\mathbf{X_i}} > 0$, for every i. Hence, $\|(T^{-1}\mathbf{X_i'X_i})^{-1}\| = O_p(1)$ and $\|(T^{-1}\mathbf{X_i'}\mathscr{M}_{\mathbf{D}}\mathbf{X_i})^{-1}\| = O_p(1)$.

Remark 3.13. Assumption 3.6 implies that sometimes $\boldsymbol{\mathcal{V}_i}$ and $\boldsymbol{\epsilon_i}$ can be interchanged with $\mathbf{V_i}$ and $\boldsymbol{\varepsilon}_t$ without affecting the results; that is, they give rise to the same limits. For instance, $\Sigma_{\boldsymbol{\mathcal{V}_i'\mathcal{V}_i}} = \boldsymbol{\Sigma}_{\mathbf{V}_i'\mathbf{V}_i}$, $\Sigma_{\boldsymbol{\epsilon_i'\epsilon_i}} = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}$, and $\Sigma_{\boldsymbol{\mathcal{V}_i'\epsilon_i}} = \boldsymbol{\Sigma}_{\mathbf{V}_i'\boldsymbol{\varepsilon}}$.

4. Methodology: Asymptotics

The following two theorems enunciate the asymptotic properties of the OLS, WLS, and FWLS estimators, respectively, where $\xrightarrow{d}$ denotes convergence in distribution.

Theorem 4.1. When Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 hold, as $T \to \infty$, for any N,

(i) (OLS estimator)

$$T^{1/2}(\Sigma_{\mathbf{i}}{}^{\mathrm{OLS}})^{-1/2}(\hat{\boldsymbol{\beta}}_i^{\mathrm{OLS}} - \boldsymbol{\beta}_i - \hat{\boldsymbol{\tau}}_i^{\mathrm{OLS}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$$

where the bias term satisfies

$$\hat{\tau}_i^{\mathrm{OLS}} := (\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{X}_i})^{-1}\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{F}}\mathbf{b}_i \to_p \tau_i^{\mathrm{OLS}} := \Sigma_{\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{X}_i}}{}^{-1}\Sigma_{\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{F}}}\mathbf{b}_i, \qquad (4.1)$$

and the asymptotic covariance matrix equals

$$\Sigma_{\mathbf{i}}{}^{\mathrm{OLS}} := \Sigma_{\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{X}_i}}{}^{-1}\Sigma_{\boldsymbol{\mathcal{X}_i}'\mathfrak{X}_{\mathbf{i}}\boldsymbol{\mathcal{X}_i}}\Sigma_{\boldsymbol{\mathcal{X}_i}'\boldsymbol{\mathcal{X}_i}}{}^{-1}. \qquad (4.2)$$

(ii) (WLS estimator)

$$T^{1/2}(\Sigma_i^{\mathrm{WLS}})^{-1/2}\left(\hat{\boldsymbol{\beta}}_i^{\mathrm{WLS}} - \boldsymbol{\beta}_i\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_K), \qquad (4.3)$$

where the asymptotic covariance matrix equals

$$\Sigma_i^{\mathrm{WLS}} := \Sigma^{-1}_{\boldsymbol{\mathcal{V}_i}'\mathfrak{X}_i^{-1}\boldsymbol{\mathcal{V}_i}}.$$

Remark 4.1. The OLS (and the WLS) estimator is functionally independent of N, which plays no role in the asymptotic analysis of Theorem 4.1. The OLS estimator is affected by a *first-order bias*. To simplify the exposition, we centered the OLS estimator around the sample bias $\hat{\tau}_i^{\mathrm{OLS}}$, as opposed to the population bias τ_i^{OLS}. This has no relevant consequence as the scope of part (a) of the theorem is simply to bring attention to the existence of the bias plaguing the OLS estimator. Thus, the OLS estimator is asymptotically unbiased if either $\mathbf{b}_i = \mathbf{0}$ or $\Gamma_{\mathbf{i}} = \mathbf{0}$ or, alternatively, for diagonal $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{F}}'\boldsymbol{\mathcal{F}}}$ as well as with $\Gamma_{\mathbf{i}}$ and $\mathbf{b}_i$ satisfying $\gamma_{il}b_{il} = 0$ for every l and i.

Remark 4.2. To gauge the intuition behind our WLS estimator, suppose that $S = 0$ (no common regressors), and $M = K = 1$ (one latent factor and one heterogeneous regressor) where the common factor vector equals the unit vector, that is, $\mathbf{F} = \boldsymbol{\iota}_T$, implying that model (2.1) can be written as

$$\mathbf{y}_i = \mathbf{X_i}\boldsymbol{\beta_i} + \mathbf{u}_i \text{ with } \mathbf{u}_i = b_i\boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_i \text{ and } \mathbf{X_i} = \gamma_i\boldsymbol{\iota}_T + \mathbf{v}_i. \quad (4.4)$$

Then, $\mathbf{S}_i = B\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \mathbf{I}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \mathbf{I}_T$ where $N^{-1}\sum_{i=1}^N b_i^2 \to B = 1$ and we set $\mathbb{E}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' = \boldsymbol{\Xi}_i = \mathbf{I}_T$. By the Shermann–Morrison theorem (see Lemma L.2 in the supplementary materials) and simple algebraic steps,

$$\mathbf{S}_i^{-1}\boldsymbol{\iota}_T = \frac{\boldsymbol{\iota}_T}{(1 + \boldsymbol{\iota}_T'\boldsymbol{\iota}_T)} = \frac{\boldsymbol{\iota}_T}{(1 + T)}, \qquad (4.5)$$

that is, each element of the vector $\mathbf{S}_i^{-1}\boldsymbol{\iota}_T$ converges to zero at the fast rate $O(T^{-1})$.[9] This fast rate is the *key ingredient* driving the asymptotic properties of the WLS and FWLS estimators. In particular, examining the random component (which involves $\mathbf{u}_i$) of our WLS estimator $\hat{\boldsymbol{\beta}}_i^{\text{WLS}} = \boldsymbol{\beta}_i^{\text{WLS}} + \left(\mathbf{X}_i\mathbf{S}_i^{-1}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{u}_i$, one obtains, by (4.5),

$$\begin{aligned}\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{u}_i &= (\gamma_i\boldsymbol{\iota}_T + \mathbf{v}_i)'\mathbf{S}_i^{-1}(b_i\boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_i)\\ &= \mathbf{v}_i'\mathbf{S}_i^{-1}\boldsymbol{\varepsilon}_i + \frac{b_i\gamma_i T}{(1+T)} + \frac{(b_i\mathbf{v}_i + \gamma_i\boldsymbol{\varepsilon}_i)'\boldsymbol{\iota}_T}{(1+T)}\\ &= \mathbf{v}_i'\mathbf{S}_i^{-1}\boldsymbol{\varepsilon}_i + O(1) + O_p(T^{-1/2})\\ &= \mathbf{v}_i'\boldsymbol{\varepsilon}_i + O_p(1) + O(1) + O_p(T^{-1/2}),\end{aligned}$$

under our assumptions.[10] Therefore, $\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{u}_i$ is asymptotically equivalent to $\mathbf{v}_i'\boldsymbol{\varepsilon}_i$ implying that the weighting matrix $\mathbf{S}_i^{-1}$ annihilates (asymptotically) the effect of the latent factors $\mathbf{F} = \boldsymbol{\iota}_T$. This result demonstrates how the WLS estimator $\hat{\boldsymbol{\beta}}_i^{\text{WLS}}$ is asymptotically unbiased, because $\mathbf{v}_i'\boldsymbol{\varepsilon}_i = O_p(T^{1/2})$ and $\left(\mathbf{X}_i\mathbf{S}_i^{-1}\mathbf{X}_i\right)^{-1} = O_p(T^{-1})$ under our assumptions.

Considering now the general case $\boldsymbol{\Xi}_i \neq \mathbf{I}_T$, by the same steps outlined above one obtains $\mathbf{X}_i'\mathbf{S}_i^{-1}\mathbf{u}_i = \mathbf{v}_i'\boldsymbol{\Xi}_i^{-1}\boldsymbol{\varepsilon}_i + O_p(1)$ and $\mathbf{X}_i\mathbf{S}_i^{-1}\mathbf{X}_i = \mathbf{v}_i'\boldsymbol{\Xi}_i^{-1}\mathbf{v}_i + O_p(1)$, implying that the WLS estimator is also asymptotically equivalent to the (unfeasible) GLS estimator, derived under the assumption that $\boldsymbol{\Xi}_i$ is known.

The weighting matrix $\mathbf{S}_i$ represents one of the many suitable choices for a weighting matrix. The limiting properties of the WLS estimator remain unchanged if one replaces $\mathbf{B}$ with any other nonsingular positive definite matrix with bounded eigenvalues, although $\mathbf{B}$ emerges when considering the FWLS. In our example above, we can replace $B = 1$ with any other positive constant, and obtaining identical limiting properties of the WLS estimator. In contrast, the matrix $\boldsymbol{\Xi}_i$ is critical in order to achieve efficiency, as we have just outlined.

Further details are relegated to Section D of the supplemental materials, where we extend the above arguments to the case of generic latent factors $\mathbf{F}$ when observed common regressors are allowed for ($S > 0$), and when M and K can be greater than one.

Remark 4.3. The assumption that the same latent factors $\mathbf{F}$ enter into $\mathbf{X}_i$ and $\mathbf{u}_i$ is without loss of generality. In fact, assume $\mathbf{u}_i = \mathbf{G}\mathbf{b}_i + \boldsymbol{\varepsilon}_i$, with the rows of $\mathbf{G}$ correlated but not identical to the rows of $\mathbf{F}$. Then the bias takes the form

$$\tau_i^{\text{OLS}} = \Sigma_{\mathcal{X}_i'\mathcal{X}_i}{}^{-1}\Sigma_{\mathbf{F}'\mathscr{P}_F\mathbf{G}}\mathbf{b}_i,$$

exploiting the decomposition $\mathbf{G} = \mathscr{P}_F\mathbf{G} + \mathscr{M}_F\mathbf{G}$, where $\mathscr{P}_F := \mathbf{I}_T - \mathscr{M}_F$. Hence, the bias will only be nonzero due to the portion of $\mathbf{G}$ correlated with $\mathbf{F}$. The same consideration applies to the FWLS estimator. In Appendix F of the supplementary materials, we explore more in detail the implications of having different, yet correlated, factor structures for regressors and innovations.

We now present the main result of the article.

Theorem 4.2. When Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6 hold:
(i) As $1/T + T/N \to 0$,

$$\hat{\boldsymbol{\beta}}_i^{\text{FWLS}} \xrightarrow{p} \boldsymbol{\beta}_i; \tag{4.6}$$

(ii) As $1/T + T^2/N \to 0$, then

$$\begin{aligned}&\left(\mathcal{V}_i'\mathcal{C}_N{}^{-1}\mathfrak{X}_i\mathcal{C}_N{}^{-1}\mathcal{V}_i\right)^{-1/2}\left(\mathcal{V}_i'\mathcal{C}_N{}^{-1}\mathcal{V}_i\right)\\ &(\hat{\boldsymbol{\beta}}_i^{FWLS} - \boldsymbol{\beta}_i) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_K\right),\end{aligned} \tag{4.7}$$

where

$$\mathcal{C}_N := \mathbf{D}_\perp'\mathbf{C}_N\mathbf{D}_\perp, \qquad \mathbf{C}_N := \frac{1}{N}\sum_{i=1}^N(\boldsymbol{\Xi}_i + \Theta_{\mathbf{i}}) \tag{4.8}$$

with

$$\Theta_{\mathbf{i}} := \mathbb{E}\left[\mathbf{V}_i\Sigma_{\mathcal{X}_i'\mathcal{X}_i}{}^{-1}\Gamma_{\mathbf{i}}\Sigma_{\mathcal{F}'\mathcal{F}}\mathbf{b}_i\mathbf{b}_i'\Sigma_{\mathcal{F}'\mathcal{F}}\Gamma_{\mathbf{i}}'\Sigma_{\mathcal{X}_i'\mathcal{X}_i}{}^{-1}\mathbf{V}_i'\right]. \tag{4.9}$$

(iii) For any integer $\tau < T$ satisfying $1/\tau + \tau/T^{\frac{1}{4}} + T^2/N \to 0$

$$\left(\mathcal{X}_i'\hat{\mathcal{S}}_N{}^{-1}\mathcal{X}_i/T\right)^{-1} - \left(\mathcal{V}_i'\mathcal{C}_N{}^{-1}\mathcal{V}_i/T\right)^{-1} \to_p \mathbf{0}_{K\times K} \text{ and} \tag{4.10}$$

$$\begin{aligned}&\left(\hat{\mathcal{L}}_{i0} + \sum_{h=1}^{\tau}\ell(h,\tau)\left(\hat{\mathcal{L}}_{ih} + \hat{\mathcal{L}}_{ih}'\right)\right)\\ &- \left(\mathcal{V}_i'\mathcal{C}_N{}^{-1}\mathfrak{X}_i\mathcal{C}_N{}^{-1}\mathcal{V}_i/T\right) \to_p \mathbf{0}_{K\times K},\end{aligned} \tag{4.11}$$

setting $\ell(j,\tau) := 1 - [h/(\tau+1)]$, and $\hat{\mathcal{L}}_{ih} := T^{-1}\sum_{t=h+1}^T\hat{a}_{it}\hat{a}_{i,t-h}\hat{\omega}_{it}\hat{\omega}_{i,t-h}'$, $h = 0, 1, \ldots, \tau$, where $\hat{a}_{it}$ denotes tth element of $(T \times 1)$ vector $\hat{\boldsymbol{a}}_i := \hat{\mathcal{S}}_N{}^{-1/2}\hat{\boldsymbol{u}}_i^{\text{FWLS}}$, with $\hat{\boldsymbol{u}}_i^{\text{FWLS}} := \boldsymbol{y}_i - \mathcal{X}_i\hat{\boldsymbol{\beta}}_i^{\text{FWLS}}$ and $\hat{\omega}_{it}'$ denotes the denotes tth row of the $(T \times K)$ matrix $\widehat{\mathcal{W}}_i := \hat{\mathcal{S}}_N{}^{-1/2}\mathcal{X}_i$.

Remark 4.4. To simplify the exposition and the proofs, we have not differentiated the conditions required for consistency and asymptotic normality of the FWLS estimator, such as the moment conditions on the regressors and the error term. However, different rates for N and T are required for consistency and asymptotic normality of the FWLS estimator.

Remark 4.5. A mechanism, similar to the one described in Remark 4.2 for the WLS estimator, applies to the FWLS in terms of bias-reduction, in turn granted by annihilating the $\mathcal{F}$, as long as $\mathbf{S}_i^{-1}$ is replaced by *any* matrix of the form $(\mathcal{F}\mathbf{A}\mathcal{F}' + \mathbf{C})^{-1}$, where $\mathbf{A}$ and $\mathbf{C}$ are, as minimum conditions, nonsingular matrices (in general, functions of N and T) with uniformly bounded eigenvalues. The most challenging aspect of the proof to Theorem 4.2 is to establish that $\hat{\mathcal{S}}_N{}^{-1}$ represents (asymptotically) one of those matrices, where in particular its $\mathbf{C}$ matrix is approximately (in a precise sense defined below) behaving like $N^{-1}\sum_{i=1}^N\boldsymbol{\Xi}_i$, as a byproduct of the averaging implicit in the construction of $\hat{\mathcal{S}}_N$ (by (2.12)). Thus, such $\mathbf{C}$ matrix can never be equal to $\boldsymbol{\Xi}_i$, except (generically speaking) when $\boldsymbol{\Xi}_i = \boldsymbol{\Xi}$, an homogenity case across units of the $\boldsymbol{\varepsilon}_i$. Therefore, the FWLS estimator is not efficient in general, as evinced by the form of the asymptotic covariance matrix in Theorem 4.2.

[9] Although the vector $\mathbf{S}_i^{-1}\iota_T$ is $T \times 1$, and hence, growing with T, its norm $\| \mathbf{S}_i^{-1}\iota_T \|$ is $O(T^{-1/2})$, fastly decaying toward zero.

[10] In particular, $(b_i\mathbf{v}_i + \gamma_i\varepsilon_i)'\iota_T/(1+T)$ has mean zero and variance $O(T^{-1})$, and $\mathbf{v}_i'\mathbf{S}_i^{-1}\varepsilon_i = \mathbf{v}_i'\varepsilon_i - (\mathbf{v}_i'\iota_T)(\varepsilon_i'\iota_T)/(1+T) = \mathbf{v}_i'\varepsilon_i + O_p(1)$ because the two terms $(\varepsilon_i'\iota_T)$ and $(v_i'\iota_T)$ have mean zero and variance $O(T)$.

Remark 4.6. A multistep procedure can be envisaged that improves the precision of the FWLS and, under homogeneity of the covariance matrix of the error, achieves an estimator with an asymptotic distribution arbitrarily close to the asymptotic distribution of the WLS estimator. We shall call this procedure the *iterated*-FWLS estimator. The first step starts from the FWLS estimator, which we now denominate as $\hat{\boldsymbol{\beta}}_i^{(1)}$ for simplicity. We then construct the associated residuals $\hat{\boldsymbol{u}}_{\boldsymbol{i}}^{(1)} := \boldsymbol{y_i} - \boldsymbol{\mathcal{X}_i}\hat{\boldsymbol{\beta}}_i^{(1)}$. Recall that $\hat{\boldsymbol{\beta}}_i^{(1)} = \hat{\boldsymbol{\beta}}_i^{\text{FWLS}}$ is a consistent estimator for $\boldsymbol{\beta}_i$, and thus, the residuals $\hat{\boldsymbol{u}}_{\boldsymbol{i}}^{(1)}$ are close to the error $\boldsymbol{u_i}$, asymptotically. The second step entails constructing $\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(1)} := N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{u}}_{\boldsymbol{i}}^{(1)}\hat{\boldsymbol{u}}_{\boldsymbol{i}}^{(1)\prime}$ and using it to obtain $\hat{\boldsymbol{\beta}}_i^{(2)} := \left(\boldsymbol{\mathcal{X}_i}'\left(\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(1)}\right)^{-1}\boldsymbol{\mathcal{X}_i}\right)^{-1}\boldsymbol{\mathcal{X}_i}'\left(\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(1)}\right)^{-1}\boldsymbol{y_i}$.

In general the hth step entails constructing $\hat{\boldsymbol{\beta}}_i^{(h)} := \left(\boldsymbol{\mathcal{X}_i}'\left(\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(h-1)}\right)^{-1}\boldsymbol{\mathcal{X}_i}\right)^{-1}\boldsymbol{\mathcal{X}_i}'\left(\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(h-1)}\right)^{-1}\boldsymbol{y_i}$, where $\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{N}}^{(h-1)}$ is obtained based on $\hat{\boldsymbol{u}}_{\boldsymbol{i}}^{(h-1)} := \boldsymbol{y_i} - \boldsymbol{\mathcal{X}_i}\hat{\boldsymbol{\beta}}_i^{(h-1)}$.[11]

5. Numerical Results

5.1. Monte Carlo Analysis

We conduct a set of Monte Carlo experiments to evaluate the performance of our asymptotic results for the (iterated) FWLS estimator in finite samples, and compare it with the WLS and OLS estimators, and CCE estimator of Pesaran (2006), the latter being the most commonly used estimator for model (2.1).[12] We construct various data-generating processes (DGPs) that provide an exhaustive range of situations of interest.

Design. We generate the data according to

$$y_{it} = \alpha_{i0} + \beta_{i0}x_{it} + \mathbf{b}_{i0}'\mathbf{f}_t + \varepsilon_{it}, \tag{5.1}$$

where $K = 1$, and the single heterogeneous regressor satisfies $x_{it} = 0.5 + \boldsymbol{\gamma}_{i0}'\mathbf{f}_t + v_{it}$, implying the single observed common factor d_t equal to 1 for all observations. The parameters are generated according to $\beta_{i0} \sim NID(1, 1.04)$, and thus, heterogeneous across units, where NID means iid normally distributed, $\alpha_{i0} = -0.5$ for $i = 1, \ldots, N/2$ and $\alpha_{i0} = 0.5$ for $i = N/2 + 1, \ldots, N$, the M latent common factors according to $\mathbf{f}_t = 0.5\mathbf{f}_{t-1} + \sqrt{0.5}\boldsymbol{\eta}_{f,t}$, with $\boldsymbol{\eta}_{f,t} \sim NID(\mathbf{0}_M, \mathbf{I}_M)$, and the error terms to the y_{it} and x_{it} according to $\varepsilon_{it} = \rho_{i\varepsilon}\varepsilon_{it-1} + \eta_{i\varepsilon,t}$, with $\eta_{i\varepsilon,t} \sim NID(0, \sigma_{i\varepsilon}^2(1-\rho_{i\varepsilon}^2))$, and to $v_{it} = \rho_{iv}v_{it-1} + \eta_{iv,t}$, with $\eta_{iv,t} \sim NID(0, (1-\rho_{iv}^2))$, with $\rho_{i\varepsilon} \sim UID(0.05, 0.95)$, $\rho_{iv} \sim UID(0.05, 0.95)$, and $\sigma_{i\varepsilon}^2 \sim UID(0.5, 1.5)$, where UID means iid uniformly distributed.[13]

Finally, regarding the loadings, we consider four cases:

$$\text{DGP1.}\quad M = 2, \mathbf{b}_{i0} \sim \text{NID}\left(\begin{pmatrix}1\\1\end{pmatrix}, 0.2\mathbf{I}_2\right),$$
$$\boldsymbol{\gamma}_{i0} \sim \text{NID}\left(\begin{pmatrix}0.5\\0\end{pmatrix}, 0.5\mathbf{I}_2\right),$$
and mutually independent.

$$\text{DGP2.}\quad M = 2, \mathbf{b}_{i0} \sim \text{NID}\left(\begin{pmatrix}1\\1\end{pmatrix}, 0.2\mathbf{I}_2\right),\quad \boldsymbol{\gamma}_{i0} = \mathbf{b}_{i0}.$$

$$\text{DGP3.}\quad M = 4, \mathbf{b}_{i0} \sim \text{NID}\left(\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, 0.2\mathbf{I}_4\right),$$
$$\boldsymbol{\gamma}_{i0} \sim \text{NID}\left(\begin{pmatrix}0.5\\0\\0.5\\0\end{pmatrix}, 0.5\mathbf{I}_4\right),$$
and mutually independent.

$$\text{DGP4.}\quad M = 2, \mathbf{b}_{i0} \sim \text{NID}\left(\begin{pmatrix}0\\0\end{pmatrix}, 0.2\mathbf{I}_2\right),$$
$$\boldsymbol{\gamma}_{i0} \sim \text{NID}\left(\begin{pmatrix}0.5\\0\end{pmatrix}, 0.5\mathbf{I}_2\right),$$
and mutually independent.

The meaning of the above DGPs is as follows: DGP1 is the benchmark case; DGP2 represents the case of (perfect) cross-correlation between the loadings in the regressor and in the error term; DGP3 represents the case when $M = 4 > K = 1$, that is when the number of latent factors exceeds (substantially) the number of regressors; DGP4 represents the case when the factors affecting the error term are not pervasive.[14] The DGPs here follow the Monte Carlo design of Pesaran (2006). We have considered alternative Monte Carlo designs, such as the one in Karabiyik, Urbain, and Westerlund (2019), with extremely similar results. (Details are available upon request.)

Results. We consider 1000 Monte Carlo replications with sample sizes $(N, T) \in \{60, 200, 600\} \times \{30, 100, 300\}$, where $N > T$.[15] Under DGP1 of Table 1, all estimators work, in terms of absolute bias (expressed as the percentage change, in absolute value, of the estimate from the true value) and root-mean-squared-error, except for OLS. As predicted by our theory, the FLWS works particularly well when N large, especially when N and T are of comparable magnitude, and is commensurate to the CCE. The performance of the WLS and GLS remains similar to one another throughout all the DGPs. Tables 2, 3, and 4 show the performance of the estimators in the realistic scenarios of interest indicated by DGP2, DGP3, and DGP4. When the loadings to the latent factors, $\mathbf{b}_{i0}$ and $\boldsymbol{\gamma}_{i0}$, are (perfectly) cross-correlated (Table 2), the FLWS continue to perform well, whereas the CCE

[11] The Monte Carlo experiments, presented in the subsequent Section 5.1, suggest how, as h increases, the asymptotic distribution of the FWLS $\hat{\boldsymbol{\beta}}_i^{(h)}$ gets close to the one of the WLS.

[12] The CCE estimator builds on the intuition, mathematically formalised by Pesaran (2006), that cross-section averages of the observables unit-specific regressors $\mathbf{X}_i$ and the $\mathbf{y}_i$ span the space of the unobserved common factors $\mathbf{F}$. Thus, augmenting regression (2.1) with these averages mitigates the differential effects of $\mathbf{F}$, so long the so-called rank condition holds, namely that the $M \times (K+1)$ matrix $\mathbb{E}(\mathbf{b}_i, \Gamma_i)$ has full row-rank M.

[13] These cases imply that the ε_{it} are independent across units but results are available for when the ε_{it} are cross-sectionally dependent through a (weak) factor structure.

[14] The so-called rank condition, required by the CCE estimator, is satisfied for DGP1 but fails for DGP2, DGP3, and DGP4. Juodis, Karabiyik, and Westerlund (2021) show that in certain situations the pooled CCE estimator, that is, the generalization of the fixed effects OLS estimator (see (65) in Pesaran 2006), for the average slope is still consistent (despite the rank condition fails) although with a non-Normal asymptotic distribution.

[15] In the interest of space, we will only illustrate the results for the estimation of the regression coefficients β_{i0}. Overall, little difference arises between the simple and iterated FWLS estimators, especially in terms of root-mean-square error, when N is large (although they can differ when N is small), and therefore, we omit the results for the simple FWLS. Likewise, we omit the results for the estimation of the coefficients α_{i0}. Details are available upon request.

Table 1. DGP1 with heterogeneous AR(1) idiosyncratic error ϵ_{it}

	absolute bias (per cent)					rmse (per cent)					10%	5%	1%
(N, T)	FWLS	WLS	GLS	OLS	CCE	FWLS	WLS	GLS	OLS	CCE		FWLS	
(60 ,30)	18.17	11.13	11.23	47.71	14.73	21.86	13.35	13.49	50.54	17.71	0.103	0.056	0.015
(200,30)	12.77	11.16	11.27	44.75	14.15	15.62	13.68	13.82	48.00	17.29	0.112	0.062	0.018
(200,100)	9.14	5.82	5.84	42.18	8.01	11.13	7.08	7.10	42.67	9.65	0.103	0.054	0.012
(600,30)	11.71	11.11	11.24	45.30	14.04	14.23	13.53	13.69	49.16	17.25	0.116	0.065	0.019
(600,100)	6.76	5.80	5.82	43.13	7.86	8.16	7.00	7.02	44.24	9.56	0.106	0.055	0.013
(600,300)	5.18	3.30	3.30	42.48	4.63	6.25	3.98	3.98	42.36	5.61	0.103	0.053	0.011

NOTE: Columns 2–6 report the absolute bias (in percentage: $100 \times \mathrm{abs}(\tilde{\beta}_{i0}/\beta_{i0} - 1)$) of the FWLS (iterated), WLS, GLS, OLS, and CCE estimators of β_{i0} of model (5.1), columns 7–11 report the root mean square error (in percentage) for the same estimators, and columns 12–14 report the size of the *t-ratios* for the FWLS (iterated) estimator, corresponding to a nominal size of 10%, 5%, and 1%, respectively. We report averages, across 1000 Monte Carlo iterations *and* across the N units, for all the above quantities, that is $N^{-1}1000^{-1}\sum_{i=1}^{N}\sum_{m=1}^{1000}\Delta_{m,i}$ where $\Delta_{m,i}$ refers to any of the above quantities (absolute bias, rmse, and *t-ratios* sizes) for the ith unit and mth Monte Carlo iteration for a given (N, T) and a given estimator. The true β_{i0} are generated as $\mathrm{NID}(1, 0.04)$.

Table 2. DGP2 with heterogeneous AR(1) idiosyncratic error ϵ_{it}

	absolute bias (per cent)					rmse (per cent)					10%	5%	1%
(N, T)	FWLS	WLS	GLS	OLS	CCE	FWLS	WLS	GLS	OLS	CCE		FWLS	
(60,30)	19.28	11.93	11.93	68.88	21.74	22.03	13.50	13.49	65.70	24.13	0.102	0.055	0.015
(200,30)	13.59	11.66	11.54	66.54	20.97	16.26	13.93	13.82	65.07	24.06	0.124	0.07	0.021
(200,100)	9.34	5.99	5.97	67.53	17.17	11.13	7.12	7.09	64.64	18.50	0.103	0.053	0.012
(600,30)	12.35	11.33	11.23	65.31	20.81	14.99	13.79	13.70	65.20	24.11	0.133	0.078	0.024
(600,100)	6.85	5.83	5.81	66.74	17.18	8.28	7.05	7.02	65.26	18.56	0.11	0.057	0.014
(600,300)	5.17	3.30	3.30	66.87	15.82	6.25	3.99	3.98	64.95	16.23	0.103	0.053	0.011

NOTE: See the note to Table 1.

Table 3. DGP3 with heterogeneous AR(1) idiosyncratic error ϵ_{it}.

	absolute bias (per cent)					rmse (per cent)					10%	5%	1%
(N, T)	FWLS	WLS	GLS	OLS	CCE	FWLS	WLS	GLS	OLS	CCE		FWLS	
(60,30)	17.41	11.11	11.56	46.44	21.85	21.20	13.49	14.06	50.56	24.80	0.113	0.063	0.018
(200,30)	12.50	11.31	11.65	47.90	20.11	15.29	13.92	14.38	52.39	23.23	0.131	0.074	0.022
(200,100)	9.03	5.83	5.87	44.02	17.00	11.05	7.13	7.18	45.22	18.32	0.106	0.055	0.013
(600,30)	11.53	11.44	11.74	49.86	20.65	13.95	13.87	14.27	53.81	23.73	0.128	0.073	0.022
(600,100)	6.74	5.86	5.89	47.14	17.79	8.11	7.05	7.09	47.97	19.02	0.109	0.058	0.014
(600,300)	5.18	3.32	3.32	46.08	16.88	6.23	3.99	3.99	45.49	17.13	0.104	0.053	0.011

NOTE: See the note to Table 1.

Table 4. DGP4 with heterogeneous AR(1) idiosyncratic error ϵ_{it}.

	absolute bias (per cent)					rmse (per cent)					10%	5%	1%
(N, T)	FWLS	WLS	GLS	OLS	CCE	FWLS	WLS	GLS	OLS	CCE		FWLS	
(60,30)	17.77	10.52	11.23	21.94	17.04	21.32	12.59	13.49	24.33	20.01	0.107	0.058	0.016
(200,30)	12.42	10.53	11.27	21.05	20.05	15.16	12.82	13.82	23.80	23.01	0.118	0.066	0.019
(200,100)	9.08	5.70	5.84	18.36	16.42	11.04	6.92	7.10	19.39	17.60	0.104	0.054	0.013
(600,30)	11.46	10.49	11.24	20.92	18.35	13.87	12.70	13.69	23.74	21.52	0.128	0.074	0.023
(600,100)	6.72	5.69	5.82	18.19	14.45	8.11	6.86	7.02	19.32	15.89	0.111	0.059	0.014
(600,300)	5.17	3.28	3.30	17.26	13.06	6.24	3.95	3.98	17.45	13.56	0.104	0.054	0.012

NOTE: See the note to Table 1.

becomes severely biased. The same pattern is observed (Table 3) when the number of latent factors M is much larger than the number of observed regressors (K), and even when the factor structure in the error term of (5.1) is weak (Table 4). The last three columns of each table report the empirical sizes of the *t-ratios*, defined as $\sqrt{T}(\tilde{\beta}_{i0} - \beta_{i0})/\mathrm{SE}_{\beta_i}$ for the FWLS estimator (SE_{β_i} denotes the standard error of $\tilde{\beta}_{i0}$ as from Theorem 4.2-(iii)), which appear close to the nominal sizes of 1%, 5%, and 10%, across all tables in most cases, with some deterioration occurring when T remains small.

5.2. *Empirical Application: Capital Accumulation and GDP Growth*

We investigate the empirical relationship between investment in physical capital and long-run economic growth across countries, a topic of paramount importance for quantifying the effectiveness of development policy, and also representing a way to evaluate economic growth theories. Endogenous growth models, the so-called AK models (see, Romer 1986, among others), predict that that an increase in investment results in a permanently higher growth rate, implying a reduced form equation for per capita GDP growth equal, in its simplest form, to

$$\Delta y_{it} = \theta_1 + \theta_2 x_{it} + \beta_1 \Delta x_{it} + u_{it}, \tag{5.2}$$

indicating by x_{it} the logarithm of investment as a share of gross domestic product (GDP) for country i in year t, and by y_{it} the logarithm of output per worker for country i in year t, with Δy_{it} denoting the annual per capita GDP growth. Model (5.2) implies a steady-state GDP growth equal to $\Delta y_i = \theta_1 + \theta_2 x_i$.

In contrast, exogenous growth theories, variation of the celebrated Solow model (see Mankiw, Romer, and Weil 1992, among others), postulates that the long-run growth rate of output per worker is only determined by the exogenous technical progress but predict a positive correlation between investment and the level of per capita GDP, ruling out any association between investment and steady-state growth rates. These theories can be expressed by the reduced form equation

$$y_{it} = \theta_0 + \theta_1 t + \beta_1 x_{it} + u_{it}, \tag{5.3}$$

implying the constant steady-state GDP growth $\Delta y_i = \theta_1$. Now, a permanent increase in the investment-GDP ratio x_{it} predicts a higher level of output per worker y_{it} along the steady state growth path, but affects growth only during the transition to the new steady state.

The existing empirical evidence is mainly focused on the OECD countries and is mixed at best, with many studies concluding that there is a weak association between investment rates and long-run growth rates, casting doubt on the effectiveness of physical capital accumulation as a source of long-run economic growth (see Jones 1995; Bond, Leblebicioglu, and Schiantarelli 2010, among others). Most of the empirical studies consider the case of constant slopes in the estimated regression, implying that one can only learn about average effects, leaving out the substantial heterogeneity existing among countries. This is a consequence of these studies using cross-sectional and panel regressions, that unavoidably do not allow identification the heterogenous trend slopes. Moreover, capital accumulation and GDP growth are no doubt jointly determined, implying that standard techniques such as OLS are affected by simultaneous equation bias. Our estimation procedure, based on the FWLS approach, tackles both problem, offering a substantial improvement over the existing econometric techniques used so far in the empirical growth literature.

We adapt Bond, Leblebicioglu, and Schiantarelli (2010) (see their eq. (7)) and postulate the following linear relationship between the logarithm of GDP per capita y_{it}, and the logarithm of the investment share to GDP x_{it}, which nests both the endogenous and the exogenous growth model specifications (5.2) and (5.3):

$$\Delta y_{it} = \theta_{1i} + \theta_{2i} x_{it} + \beta_{1i} \Delta x_{it} + \beta_{2i} \Delta x_{it-1} + u_{it}, \tag{5.4}$$

or, equivalently, $\Delta y_{it} = \alpha_i d_t + \mathbf{x}_{it}' \boldsymbol{\beta}_i + u_{it}$ setting $d_t = 1, \mathbf{x}_{it} = (x_{it}, \Delta x_{it}, \Delta x_{it-1})'$. Importantly, (5.4) allows for heterogeneity across countries of the regression coefficients, a crucial feature given that we are using data of both OECD and Non-OECD countries, for which the assumption of constant growth rate across countries is untenable. Model (5.4) implies that the GDP per worker, for country i, grows at the country-specific rate of $\Delta y_i = \theta_{1i} + \theta_{2i} x_i$ in steady state, nesting both the endogenous (when $\theta_{2i} = 0$) and the exogenous (when $\theta_{2i} \neq 0$) cases. Therefore, the long-run effect of an increase in the investment share on the growth rate of output per worker, defined as the *growth effect*, is $\partial \Delta y_i / \partial x_i = \theta_{2i}$, and the the long-run effect of an increase in the investment (log) share on the (log) level of output per worker, defined as the *level effect*, is given by $\partial \Delta y_i / \partial \Delta x_i = (\beta_{1i} + \beta_{2i})$.[16]

We use the data extracted from both the Penn World Table 6.3 (PWT 6.3) regarding real GDP per worker and the share of total gross investment in GDP, both measured in constant-price international dollars, for $N = 151$ countries (OECD and Non-OECD countries), yearly observations from 1970 to 2007.[17] Estimation of model (5.4) is challenging due to the strong degree of persistence, both time series and cross-sectional. Taking first-difference of the (log) GDP per capita, as in (5.4), mitigates the first concern, whereas this is less concerning for the (log) investment share x_{it} as it represents a fraction. Figure 1, left panel, reports the p-values of the Augmented Dickey–Fuller test, across countries, for the log GDP (black line) and its firs-difference (red line), together the 5% level (blue line), showing the effectiveness of the taking the first-difference. We did not take the first-difference of the investment share x_{it} as the assumption of stationarity is plausible given that x_{it} represents a ratio bounded between zero and one.[18]

Regarding the degree of cross-sectional dependence, Figure 1, right panel, reports the largest 32 eigenvalues in descending order (normalized by setting the largest equal to 1) of y_{it} (black line) and x_{it} (red line), with both exhibiting a prominent factors structure. Moreover, the two sets of estimated factors, corresponding to the eigenvalues of Figure 1 (right panel), appear strongly cross-correlated, as indicated by their canonical correlations, the first five of which are equal to 0.976, 0.905, 0.875, 0.799, and 0.701 in descending order, denoting substantial cross-sectional dependence.[19]

Therefore, it is very likely that estimation by OLS of (5.4) is affected by (first-order) bias, making this an ideal setup for our FWLS estimator, as it allows the presence of a factor structure (with an arbitrary number of latent factors), both in the dependent and independent variates, together with arbitrary heterogeneity of the regression coefficients.

[16] Bond, Leblebicioglu, and Schiantarelli (2010) augment (5.4) with lagged Δy_is, to enrich the short term dynamics of the model, but the magnitude of the corresponding estimated parameters turns out to be negligible in terms of growth and level effects, and thus, we consider the more parsimonious model (5.4). The presence of the lagged Δx_{it} in (5.4) accounts for the dominant part of this dynamic effect.

[17] We excluded all countries with gaps over the chosen data period, countries for which oil production is dominant, and countries with reported negative gross investment over some years.

[18] Use of the Augmented Dickey–Fuller test shows that for the investment share the p-value is below 0.2 for about half of the countries whereas for log GDP frequency of countries with a p-value below 0.2 is only about 15%.

[19] Canonical correlations, introduced by Hotelling (1936), permit to quantify the degree of dependence between two *sets* of variates, and are defined as the $min[K_a, K_b]$ eigenvalues of the matrix $(\mathbf{A}'\mathbf{A})^{-\frac{1}{2}}\mathbf{A}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-\frac{1}{2}}$, denoting by $\mathbf{A}, \mathbf{B}$ the samples of data for the two sets of variates, respectively, of dimension $T \times K_a$ and $T \times K_b$, assuming zero mean for simplicity.

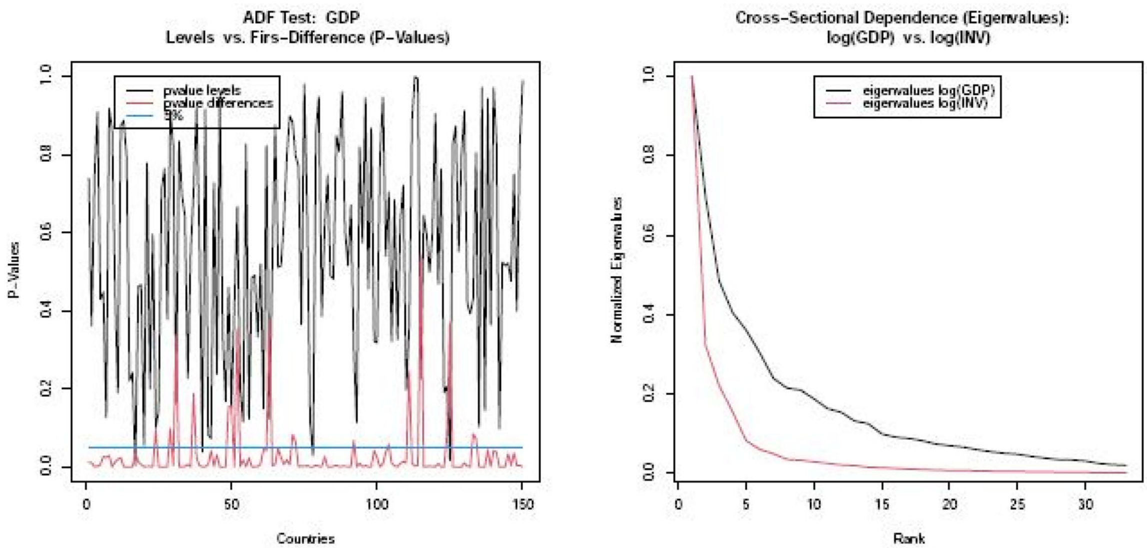

Figure 1. The left panel presents the cross-section across $N = 151$ countries of the p-values for the Augmented Dickey–Fuller test for log GDP, in level (black line) and in first-difference (red line). The blue line indicate the 5% level. The right panel presents the largest eigenvalues of the sample covariance matrix for the panel of log GDP (black line) and for the panel of log investment (red line), normalized by setting the maximum eigenvalue equal to unity.

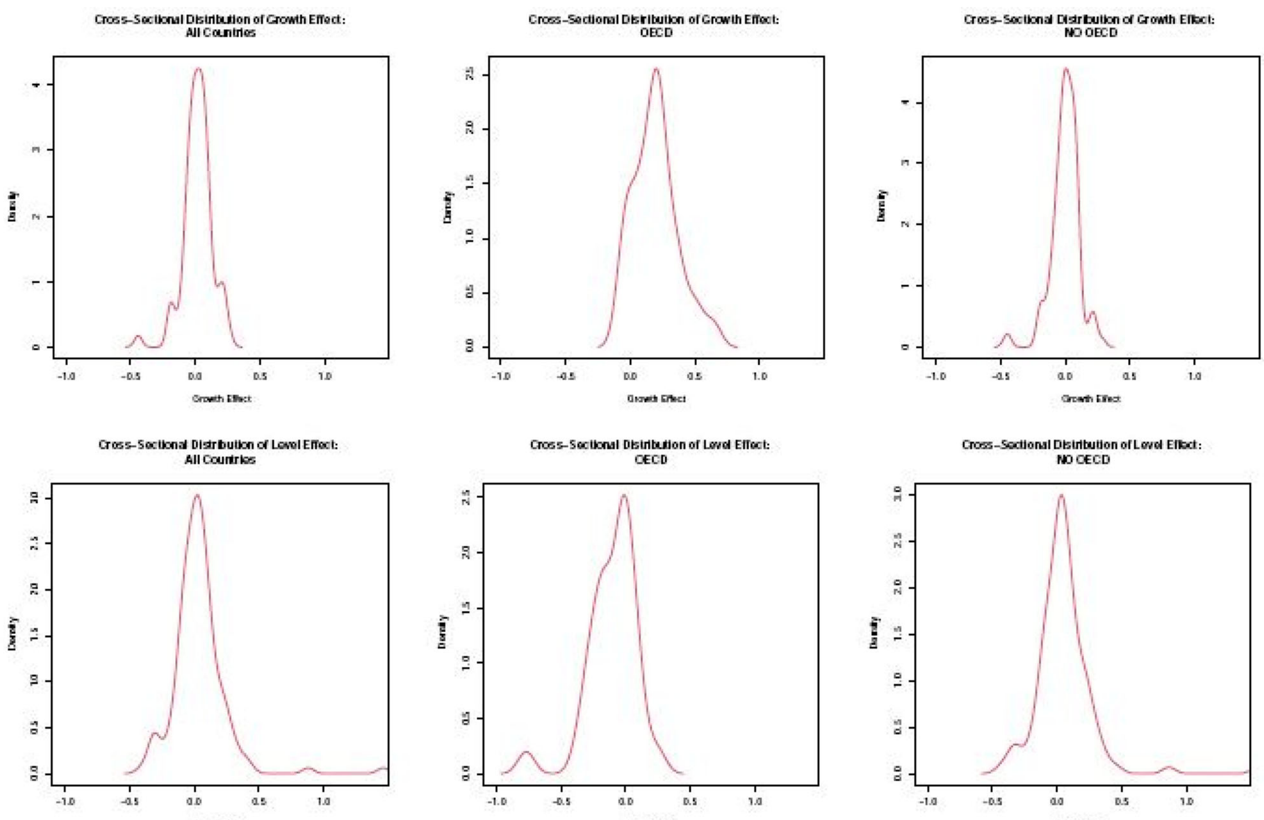

Figure 2. The top three panels present the estimated cross-sectional density, based on the nonparameteric Kernel estimator, of the *growth* effect (θ_{i1}) corresponding to all countries ($N = 151$—left panel), OECD countries ($N = 31$—middle panel), and Non-OECD countries ($N = 120$—right panel), and the bottom three panels present the estimated cross-sectional density, based on the nonparameteric Kernel estimator, of the *level* effect ($\beta_{i1} + \beta_{i2}$) corresponding to all countries ($N = 151$—left panel), OECD countries ($N = 31$—middle panel), and Non-OECD countries ($N = 120$—right panel).

Table 5 and Figure 2 report our empirical results. Generally speaking, we find robust evidence of a positive, often significant, relationship between investment and GDP growth, on average across countries. Figure 2 shows that there is a substantial degree of heterogeneity, and asymmetry, across countries, with the estimated cross-sectional densities appearing all shifted toward positive values (positive skewness) of the growth and level effects, except for the level effect of the OECD countries which appears to have negative skewness. The FWLS estimates reported in the top panel of Table 5 show clearly that some differences arise from comparing OECD and Non-OECD countries, with the growth effect being stronger with the former group

Table 5. Capital accumulation and GDP growth.

	ALL			OECD			Non-OECD		
	25% *perc* (t-*ratio*)	50% *perc* (t-*ratio*)	75% *perc* (t-*ratio*)	25% *perc* (t-*ratio*)	50% *perc* (t-*ratio*)	75% *perc* (t-*ratio*)	25% *perc* (t-*ratio*)	50% *perc* (t-*ratio*)	75% *perc* (t-*ratio*)
					FWLS				
$\tilde{\theta}_{1i}$ (growth effect)	−0.036 (−4.293)	0.024 (2.485)	0.079 (13.267)	0.083 (3.416)	0.185 (3.073)	0.271 (12.220)	−0.042 (−6.783)	0.009 (0.893)	0.068 (11.061)
$(\tilde{\beta}_{1i}+\tilde{\beta}_{2i})$ (level effect)	−0.071 (−7.844)	0.022 (1.618)	0.105 (11.221)	−0.186 (−7.216)	−0.049 (−1.240)	0.008 (0.373)	−0.065 (−5.936)	0.033 (1.998)	0.127 (11.282)
					OLS				
$\tilde{\theta}_{1i}$ (growth effect)	−0.023 (−4.373)	0.029 (3.584)	0.075 (14.415)	−0.038 (−1.779)	0.029 (0.743)	0.215 (5.166)	−0.025 (−4.227)	0.028 (3.316)	0.067 (14.125)
$(\tilde{\beta}_{1i}+\tilde{\beta}_{2i})$ (level effect)	−0.086 (−7.065)	0.029 (1.276)	0.194 (12.708)	−0.017 (−0.445)	0.147 (2.643)	0.261 (8.394)	−0.097 (−7.602)	0.011 (0.477)	0.151 (9.824
					CCE				
$\tilde{\theta}_{1i}$ (growth effect)	−0.031 (−6.950)	0.006 (0.840)	0.055 (9.563)	−0.036 (−2.375)	0.036 (1.693)	0.078 (6.171)	−0.025 (−5.414)	0.011 (1.439)	0.050 (9.254)
$(\tilde{\beta}_{1i}+\tilde{\beta}_{2i})$ (level effect)	−0.085 (−6.706)	0.041 (1.950)	0.173 (12.318)	0.124 (5.611)	0.228 (6.053)	0.325 (13.142)	−0.112 (−8.565)	−0.011 (−0.487)	0.131 (8.112)

NOTE: This table is made by three panels, corresponding to the FWLS estimates (top panel), the OLS estimates (middle panel), and the CCE estimate (bottom panel). Each panel presents the empirical quantiles (25%, median, and 75%) of the cross-sectional empirical distribution of the estimated parameters (corresponding *t-ratio* in brackets) of model (5.4), for the full sample ($N = 151$, columns two-four), for the OECD countries ($N = 31$, columns five–seven), and for the Non-OECD countries ($N = 120$, columns eight–ten). The *growth effect* parameter is defined as θ_{2i} and the *level effect* parameter is defined as $(\beta_{1i} + \beta_{2i})$.

and the level effect being stronger with the latter group, possibly because GDP growth for OECD countries is already at its steady-state level and can only be affected for shorter economic cycles. On average across countries, the median growth and level effects are positive and statistically significant, except for the level effect of OCED countries and the growth effect of Non-OECD countries, discussed above.[20] Our results indicate that, as predicted by growth theories, capital investment is an important, although not the only, ingredient of economic growth, with the results being widely heterogeneous across countries. The middle and bottom panels report the corresponding estimates obtained by OLS and CCE, respectively. What clearly emerges is that the results, both in terms of magnitude of the estimates and their *t-ratios*, for OECD countries based on the FWLS estimator appear to differ from the OLS and CCE results, which instead are more aligned with one another, suggesting that the latter are both affected by biases of similar magnitude and sign.[21]

6. Concluding Remarks

This article proposes weighted least square estimation for linear panels with a common factor structure in both the regressors and the regression error. The FWLS estimator is consistent and asymptotically normal when both the cross-section N and time series T dimensions diverge to infinity where under the same circumstances, the OLS is first-order biased. In summary, the FWLS estimator exhibits five main properties: first, it permits carrying out standard inference on the regression coefficients because based on conventional limiting distributions; second, it does not require any knowledge of the exact number of latent factors, or even an upper bound of such number; third, it is computationally easy to handle without invoking any nonlinear numerical optimizations; fourth, it works regardless of whether the regressors and the error are correlated; fifth, it can deliver efficient estimation under suitable conditions on the covariance matrix of the idiosyncratic error. Our results are corroborated by a set of Monte Carlo experiments and illustrated by an empirical application that investigates the effect of investment on GDP growth. Several generalizations, such as dynamic panels, cross-section regressions with time-varying coefficients, and different factor structures for regressors and error, are described in the supplementary materials.

[20] The *t-ratios* for the empirical quantiles of the parameters' estimates are computed using the traditional nonparameteric standard errors for empirical quantiles (see, Cramer 1946).

[21] A factor structure better captures the dependence, especially across the more integrated OECD countries, making this a particularly suitable case for using our FWLS estimator.

Supplementary Materials

Supplementary material contains additional materials, such as notation, extended literature review, the intuition behind the estimator and generalizations of the method, and the proofs of main theorems with the required lemmas.

Acknowledgments

The authors are extremely grateful to Joakim Westerlund, to the Associate Editor, and to the anonymous referees for comments that greatly improved the article.

References

Andrews, D. (2005), "Cross-Section Regression with Common Shocks," *Econometrica*, 73, 1551–1585. [3]

Bai, J. (2009), "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [1,5]

Bond, S., Leblebicioglu, A., and Schiantarelli, F. (2010), "Capital Accumulation and Growth: A New Look at the Empirical Evidence," *Journal of Applied Econometrics*, 25, 1073–1099. [9]

Brockwell, P., and Davis, R. (1991), *Time Series: Theory and Methods*. Springer Series in Statistics, Springer: New York. [4]

Cramer, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press. [11]

Hotelling, H. (1936), “Relations between Two Sets of Variates,” *Biometrika*, 28, 321–377. [9]

Jones, C. I. (1995), “Time Series Tests of Endogenous Growth Models,” *The Quarterly Journal of Economics*, 110, 495–525. [9]

Juodis, A., Karabiyik, H., and Westerlund, J. (2021), “On the Robustness of the Pooled CCE Estimator,” *Journal of Econometrics*, 220, 325–348. [7]

Karabiyik, H., Urbain, J. P., and Westerlund, J. (2019), “CCE Estimation of Factor Augmented Regression Models with More Factors than Observables,” *Journal of Applied Econometrics*, 34, 268–284. [7]

Lütkepohl, H. (1996), *Handbook of Matrices*, Chichester: Wiley. [4]

Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Probability and Mathematical Statistics, London: Wiley. [3]

Mankiw, N. G., Romer, D., and Weil , D. N. (1992), “A Contribution to the Empirics of Growth,” *Quarterly Journal of Economics*, 107, 407–437. [9]

Moon, H. R., and Weidner, M. (2015), “Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects,” *Econometrica*, 83, 1543–1579. [4]

Moon, H. R., and Weidner, M. (2017), “Dynamic Linear Panel Regression Models with Interactive Fixed Effects,” *Econometric Theory*, 33, 158–195. [4]

Onatski, A. (2015), “Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models,” *Journal of Econometrics*, 186, 388–406. [4]

Pesaran, M. H. (2006), “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure.” *Econometrica*, 74, 967–1012. [1,2,5,7]

Pesaran, M. H., and Tosetti, E. (2011), “Large Panels with Common Factors and Spatial Correlations,” *Journal of Econometrics*, 161, 182–202. [4,5]

Phillips, P. C. B. (1991), “Optimal Inference in Cointegrated Systems,” *Econometrica*, 59, 283–306. [1]

Robinson, P. M., and Hidalgo, F. J. (1997), “Time Series Regression with Long-Range Dependence,” *Annals of Statistics*, 25, 77–104. [1,4]

Romer, P. M. (1986), “Increasing Returns and Long-Run Growth,” *Journal of Political Economy*, 94, 1002–1037. [8]