Manuscript submitted to *The Econometrics Journal*, pp. 1–20.

Two-Stage Instrumental Variable Estimation of Linear Panel Data Models with Interactive Effects

Guowei Cui[†], Milda Norkutė[‡], Vasilis Sarafidis[§] and Takashi Yamagata[*]

[†] *School of Economics, Huazhong University of Science and Technology (HUST), Wuhan 430074, China.*

[‡] *Bank of Lithuania and Vilnius University, Totoriu g. 4, LT-01121, Vilnius, Lithuania.*

E-mail: MNorkute@lb.lt

[§] *Department of Economics, BI Norwegian Business School, 0484 Oslo, Norway.*

E-mail: vasilis.sarafidis@gmail.com

[*] *DERS, University of York, UK and ISER, Osaka University, Japan.*

E-mail: takashi.yamagata@york.ac.uk

Summary

This paper analyses the instrumental variables (IV) approach put forward by Norkute et al. (2021), in the context of static linear panel data models with interactive effects present in the error term and the regressors. Instruments are obtained from transformed regressors, thereby it is not necessary to search for external instruments. We consider a two-stage IV (2SIV) and a mean-group IV (MGIV) estimator for homogeneous and heterogeneous slope models, respectively. The asymptotic analysis reveals that: (i) the $\sqrt{NT}$-consistent 2SIV estimator is free from asymptotic bias that may arise due to the estimation error of the interactive effects, whilst (ii) existing estimators can suffer from asymptotic bias; (iii) the proposed 2SIV estimator is asymptotically as efficient as existing estimators that eliminate interactive effects jointly in the regressors and the error, whilst; (iv) the relative efficiency of the estimators that eliminate interactive effects only in the error term is indeterminate. A Monte Carlo study confirms good approximation quality of our asymptotic results.

Keywords: *Large panel data, Interactive effects, Common factors, Principal components analysis, Instrumental variables.*

1. INTRODUCTION

Panel data sets with large cross-section and time-series dimensions (N and T, respectively) have become increasingly available in the social sciences. As a result, regression analysis of large panels has gained an ever-growing popularity. A central issue in these models is how to properly control for rich sources of unobserved heterogeneity, including common shocks and interactive effects (see e.g. Sarafidis and Wansbeek (2020) for a recent overview).

Broadly speaking, there are two popular estimation approaches currently advanced in the field. The first one involves eliminating the interactive effects from the error term and the regressors *jointly*, in a single stage. Representative methods include the Common Correlated Effects approach of Pesaran (2006), which involves least-squares on a regression model augmented by cross-sectional averages (CA) of observables; and the Principal Components (PC) estimator considered first by Kapetanios and Pesaran (2005) and analysed subsequently by Westerlund and Urbain (2015). The second approach asymptotically eliminates the interactive effects from the error term only. The representative method is the Iterative Principal Components (IPC) estimator of Bai (2009), further developed by Moon and Weidner (2015, 2017), among many others. An attractive feature of CA (as well as PC) is that it permits estimation of models with heterogeneous slopes. On the other hand, an advantage of IPC is that it does not assume that regressors are subject to a factor structure.

In models with homogeneous slopes, Westerlund and Urbain (2015) showed that both CA and PC estimators suffer from asymptotic bias due to the incidental parameter problem (see Juodis et al. (2021) for additional results on the asymptotic properties of CA). A similar outcome was shown by Bai (2009) for the IPC estimator. Thus in all three cases, bias correction is necessary for asymptotically valid inferences. In addition, the CA estimator requires the so-called rank condition, which assumes that the number of factors does not exceed the rank of the (unknown) matrix of cross-sectional averages of the factor loadings. On the other hand, IPC involves non-linear optimisation, and so convergence to the global optimum might not be guaranteed (see e.g. Jiang et al. (2017)).

This paper analyses the instrumental variables (IV) approach put forward by Norkute et al. (2021) in the context of a static linear panel data model. Their approach differs from CA, PC and IPC because it asymptotically eliminates the interactive effects in the error term and in the regressors *separately*, in two stages. In particular, for models with homogeneous slopes, in the first stage the interactive effects are projected out from the regressors. Subsequently, the transformed regressors are used as instruments to obtain consistent estimates of the model parameters. This way, it is not necessary to search for external instruments. In the second stage, the interactive effects in the error term are eliminated using the first-stage residuals, and a second IV regression is run. That is, IV regression is performed in both of two stages. The resulting two-stage IV (2SIV) estimator is shown to be $\sqrt{NT}$-consistent and asymptotically normal. For models with heterogeneous slopes, we analyse a mean-group IV (MGIV) estimator and establish $\sqrt{N}$-consistency and asymptotic normality. The asymptotic results established in this paper are completely new, as we permit weak cross-section and time-series dependence in the idiosyncratic errors. The weak dependence assumption is typically employed by the static panel data literature, such as Bai (2009). In contrast, Norkute et al. (2021) focus on dynamic panels with interactive effects, assuming cross-sectional and serial independence of the idiosyncratic disturbances.

In addition, the present paper offers new insights into the literature by comparing and contrasting the asymptotic properties of 2SIV, IPC, PC and CA. Such a task was not considered by Norkute et al. (2021). To be more specific, we analytically show why the proposed two-stage approach makes the 2SIV estimator free from asymptotic bias, whilst under the same conditions IPC, PC and CA are subject to biases. In brief, the reason for the lack of asymptotic bias of 2SIV is that the factors in the regressors and the errors are estimated separately in two stages. This makes the endogeneity caused by the estimation errors of the interactive effects asymptotically negligible. Moreover, our analysis reveals that 2SIV is asymptotically as efficient as the bias-corrected versions of PC and CA, whereas the relative efficiency of the bias-corrected IPC estimator is indeterminate, in general. This is because the IPC estimator (i) does not necessarily eliminate the factors contained in the regressors; (ii) requires a transformation of the regressors, which is due to the estimation error of the interactive effects.

A Monte Carlo study confirms good approximation quality of our asymptotic results and competent performance of 2SIV and MGIV relative to existing estimators. Furthermore, the results demonstrate that the bias-corrections of IPC and PC can noticeably inflate the dispersion of the estimators in finite samples. We apply our methodology to study the effect of climate shocks on economic growth using an unbalanced panel of 125 countries over the period 1961-2003. The implications of our results are different from those obtained in existing literature.

A Stata algorithm that implements our approach, has been recently developed by Kripfganz and Sarafidis (2021) and is available to all Stata users.[1]

The remainder of this paper is organised as follows. Section 2 introduces a panel data model with homogeneous slopes and interactive effects, and describes the set of assumptions employed. Section 3 studies the asymptotic properties of the proposed 2SIV estimator. Section 4 analyses a mean-group IV estimator for models with heterogeneous slopes and establishes its properties in large samples. Section 5 provides an asymptotic comparison among 2SIV, IPC, CA and PC. Section 6 studies the finite sample performance of these estimators and Section 7 provides an empirical illustration. Section 8 concludes. Proofs of main theoretical results with necessary lemmas and auxiliary lemmas are relegated to Online Supplement.

Notation: Throughout, we denote the largest eigenvalues of the $N \times N$ matrix $\mathbf{A} = (a_{ij})$ by $\mu_{\max}(\mathbf{A})$, its trace by $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{N} a_{ii}$, its Frobenius norm by $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$. The projection matrix on $\mathbf{A}'$ is $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{M_A} = \mathbf{I} - \mathbf{P_A}$. C is a generic positive constant large enough, C_{min} is a small positive constant sufficiently away from zero, $\delta_{NT}^2 = \min\{N, T\}$. We use $N, T \to \infty$ to denote that N and T pass to infinity jointly.

2. MODEL AND ASSUMPTIONS

We consider the following panel data model:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{u}_i; \quad \mathbf{u}_i = \boldsymbol{\varphi}_i^{0\prime}\mathbf{h}_t^0 + \varepsilon_{it}, \\ \mathbf{x}_{it} &= \boldsymbol{\Gamma}_i^{0\prime}\mathbf{f}_t^0 + \mathbf{v}_{it}; \quad i = 1, ..., N; \quad t = 1, ..., T, \end{aligned} \tag{2.1}$$

where y_{it} denotes the value of the dependent variable for individual i at time t, $\mathbf{x}_{it}$ is a $k \times 1$ vector of regressors and $\boldsymbol{\beta}$ is the corresponding vector of slope coefficients. $\mathbf{u}_i$ follows a factor structure, where $\mathbf{h}_t^0$ is an $r_2 \times 1$ vector of latent factors, $\boldsymbol{\varphi}_i^0$ is the associated factor loading vector, and ε_{it} denotes an idiosyncratic error. The regressors are assumed to be strictly exogenous with respect to ε_{it}, however they are subject to a factor model, where $\mathbf{f}_t^0$ denotes an $r_1 \times 1$ vector of latent factors, $\boldsymbol{\Gamma}_i^0$ is a $r_1 \times k$ matrix of factor loadings, and $\mathbf{v}_{it}$ is an idiosyncratic error of dimension $k \times 1$. We treat r_1 and r_2 as given.[2]

Estimation of the model above has been studied by Pesaran (2006), Bai and Li (2014), Westerlund and Urbain (2015), Juodis and Sarafidis (2020, 2021), Cui et al. (2019) to mention a few. Such model has been employed in a wide variety of fields, including economics and finance.

Remark 2.1. *Permitting different sets of interactive effects in $\mathbf{x}_{it}$ and u_{it} is important not only from the empirical perspective but also from the theoretical perspective. It plays a crucial role when we analytically compare the estimators that eliminate the factors in*

[1] See http://www.kripfganz.de/stata/xtivdfreg.html.

[2] In practice, r_1 can be estimated from the raw data $\{\mathbf{X}_i\}_{i=1}^N$ using methods already available in the literature, such as the information criteria of Bai and Ng (2002) or the eigenvalue-based tests of Kapetanios (2010) and Ahn and Horenstein (2013). r_2 can be estimated in the same way from the residual covariance matrix. An asymptotic justification of such practice is discussed in Bai (2009b, Section C.3). In the Monte Carlo section of the paper we show that these methods provide quite accurate determination of the number of factors.

the error term and the regressors separately (as in our approach), and those estimators that eliminate the factors in the error term only (as in the IPC approach of Bai (2009)). This remark does not apply to estimators that extract factors in $\mathbf{x}_{it}$ and u_{it} jointly, as in the approaches considered by Pesaran (2006) and Westerlund and Urbain (2015)); see Section 5 for more details.

Stacking Eq. (2.1) over t, we have

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i; \quad \mathbf{u}_i = \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i, \\ \mathbf{X}_i &= \mathbf{F}^0\boldsymbol{\Gamma}_i^0 + \mathbf{V}_i, \end{aligned} \tag{2.2}$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT})'$, $\mathbf{F}^0 = (\mathbf{f}_1^0, \cdots, \mathbf{f}_T^0)'$, $\mathbf{H}^0 = (\mathbf{h}_1^0, \cdots, \mathbf{h}_T^0)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$ and $\mathbf{V}_i = (\mathbf{v}_{i1}, \dots, \mathbf{v}_{iT})'$.

Following Norkute et al. (2021), we consider an IV estimation approach that involves two stages. In the first stage, the common factors in $\mathbf{X}_i$ are asymptotically eliminated using principal components analysis. Next, the transformed regressors are used to construct instruments and estimate the model parameters. To illustrate the first-stage IV estimator, suppose that $\mathbf{F}^0$ is observed. Pre-multiplying $\mathbf{X}_i$ by $\mathbf{M}_{\mathbf{F}^0}$ yields

$$\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i = \mathbf{M}_{\mathbf{F}^0}\mathbf{V}_i. \tag{2.3}$$

Assuming $\mathbf{V}_i$ is independent of $\boldsymbol{\varepsilon}_i, \mathbf{H}^0$ and $\boldsymbol{\varphi}_i^0$, it is easily seen that $E[\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i] = E[\mathbf{V}_i'\mathbf{M}_{\mathbf{F}^0}(\mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i)] = \mathbf{0}$. Together with the fact that $\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i$ is correlated with $\mathbf{X}_i$ through $\mathbf{V}_i$, $\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i$ can be regarded as an instrument for $\mathbf{X}_i$.

The first-stage (infeasible) estimator is defined as

$$\widehat{\boldsymbol{\beta}}_{1SIV}^{inf} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{y}_i. \tag{2.4}$$

In the second stage, the space spanned by $\mathbf{H}^0$ is estimated from the residual $\widehat{\mathbf{u}}_i^{inf} = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{inf}$ and then it is projected out. To illustrate, suppose that $\mathbf{H}^0$ is also observed; one can instrument $\mathbf{X}_i$ using $\mathbf{M}_{\mathbf{H}^0}\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i$. Note that $E[\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\mathbf{u}_i] = E[\mathbf{V}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i] = \mathbf{0}$. The (infeasible) second-stage IV (2SIV) estimator of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_{2SIV}^{inf} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\mathbf{y}_i. \tag{2.5}$$

In practice, $\mathbf{F}^0$ and $\mathbf{H}^0$ are typically unobserved. As it will be discussed in detail below, we replace these quantities with estimates obtained using principal components analysis, as advanced in Bai (2003) and Bai (2009).

To obtain our theoretical results it is sufficient to make the following assumptions.

Assumption 2.1. (idiosyncratic error in $\mathbf{y}$) *: We assume that*

1. $\mathbb{E}(\varepsilon_{it}) = 0$ *and* $\mathbb{E}|\varepsilon_{it}|^{8+\delta} \leq C$ *for some* $\delta > 0$*;*
2. *Let* $\sigma_{ij,st} \equiv \mathbb{E}(\varepsilon_{is}\varepsilon_{jt})$*. We assume that there exist* $\bar{\sigma}_{ij}$ *and* $\tilde{\sigma}_{st}$*,* $|\sigma_{ij,st}| \leq \bar{\sigma}_{ij}$ *for all* (s,t)*, and* $|\sigma_{ij,st}| \leq \tilde{\sigma}_{st}$ *for all* (i,j)*, such that* $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\bar{\sigma}_{ij} \leq C$; $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\sigma}_{st} \leq C$; $N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}|\sigma_{ij,st}| \leq C$.
3. *For every* (s,t)*,* $\mathbb{E}\left\|N^{-1/2}\sum_{i=1}^{N}\left[\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}\right]\right\|^4 \leq C$*.*

4 *For each* j, $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}\varepsilon_{jt}-\mathbb{E}\left(\varepsilon_{it}\varepsilon_{jt}\right)\right]\boldsymbol{\varphi}_i^0\right\|^2 \leq C$. *Also, for each* s, $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{is}\varepsilon_{it}-\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\right)\right]\mathbf{g}_t^0\right\|^2 \leq C$, *where* $\mathbf{g}_t^0 = (\mathbf{f}_t^{0\prime}, \mathbf{h}_t^{0\prime})'$.

5 $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left|\text{cov}(\varepsilon_{is_1}\varepsilon_{is_2},\varepsilon_{jt_1}\varepsilon_{jt_2})\right| \leq C$.

ASSUMPTION 2.2. (IDIOSYNCRATIC ERROR IN $\mathbf{x}$) *Let* $\boldsymbol{\Sigma}_{ij,st} \equiv \mathbb{E}\left(\mathbf{v}_{is}\mathbf{v}_{jt}'\right)$. *We assume that*

1 $\mathbf{v}_{it}$ *is group-wise independent from* ε_{it}, $\mathbb{E}\left(\mathbf{v}_{it}\right) = \mathbf{0}$ *and* $\mathbb{E}\left\|\mathbf{v}_{it}\right\|^{8+\delta} \leq C$;

2 *There exist* $\bar{\tau}_{ij}$ *and* $\tilde{\tau}_{st}$, $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \bar{\tau}_{ij}$ *for all* (s,t), *and* $\|\boldsymbol{\Sigma}_{ij,st}\| \leq \tilde{\tau}_{st}$ *for all* (i,j), *such that* $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\bar{\tau}_{ij} \leq C$; $\quad T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\tau}_{st} \leq C$;
$N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\|\boldsymbol{\Sigma}_{ij,st}\| \leq C$. *Additionally, the largest eigenvalue of* $\mathbb{E}\left(\mathbf{V}_i\mathbf{V}_i'\right)$ *is bounded uniformly in* i.

3 *For every* (s,t), $\mathbb{E}\left\|N^{-1/2}\sum_{i=1}^{N}\left[\mathbf{v}_{is}\mathbf{v}_{it}' - \boldsymbol{\Sigma}_{ii,st}\right]\right\|^4 \leq C$.

4 *For each* j, $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\varphi}_i^0\otimes\left[\mathbf{v}_{it}\mathbf{v}_{jt}'-\mathbb{E}\left(\mathbf{v}_{it}\mathbf{v}_{jt}'\right)\right]\right\|^2 \leq C$. *Additionally, for each* s, $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{v}_{is}'\mathbf{v}_{it}-\mathbb{E}\left(\mathbf{v}_{is}'\mathbf{v}_{it}\right)\right]\mathbf{g}_t^0\right\|^2 \leq C$.

5 $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left|\text{cov}(\mathbf{v}_{is_1}'\mathbf{v}_{is_2},\mathbf{v}_{jt_1}'\mathbf{v}_{jt_2})\right| \leq C$.

ASSUMPTION 2.3. (FACTORS) $\mathbb{E}\|\mathbf{f}_t^0\|^4 \leq C$, $T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0 \xrightarrow{p} \boldsymbol{\Sigma}_F^0$ *as* $T\to\infty$ *for some non-random positive definite matrix* $\boldsymbol{\Sigma}_F^0$. $\mathbb{E}\|\mathbf{h}_t^0\|^4 \leq C$, $T^{-1}\mathbf{H}^{0\prime}\mathbf{H}^0 \xrightarrow{p} \boldsymbol{\Sigma}_H^0$ *as* $T\to\infty$ *for some non-random positive definite matrix* $\boldsymbol{\Sigma}_H^0$.

ASSUMPTION 2.4. (LOADINGS) $\mathbb{E}\left\|\boldsymbol{\Gamma}_i^0\right\|^4 \leq C$, $\boldsymbol{\Upsilon}^0 = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\Gamma}_i^{0\prime} \xrightarrow{p} \bar{\boldsymbol{\Upsilon}}^0$ *as* $N\to\infty$, *and* $\mathbb{E}\left\|\boldsymbol{\varphi}_i^0\right\|^4 \leq C$, $\boldsymbol{\Upsilon}_\varphi^0 = N^{-1}\sum_{i=1}^{N}\boldsymbol{\varphi}_i^0\boldsymbol{\varphi}_i^{0\prime} > 0 \xrightarrow{p} \bar{\boldsymbol{\Upsilon}}_\varphi^0$ *as* $N\to\infty$ *for some non-random positive definite matrices* $\bar{\boldsymbol{\Upsilon}}^0$ *and* $\bar{\boldsymbol{\Upsilon}}_\varphi^0$. *In addition,* $\boldsymbol{\Gamma}_i^0$ *and* $\boldsymbol{\varphi}_i^0$ *are independent groups from* ε_{it}, $\mathbf{v}_{it}$, $\mathbf{f}_t^0$ *and* $\mathbf{h}_t^0$.

ASSUMPTION 2.5. (IDENTIFICATION) *The matrix* $T^{-1}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i$ *has full column rank and* $\mathbb{E}\left\|T^{-1}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i\right\|^{2+2\delta} \leq C$ *for all* i.

Unlike Norkute et al. (2021), Assumptions 2.1 and 2.2 permit weak cross-sectional and serial dependence in ε_{it} and $\mathbf{v}_{it}$, in a similar manner to Bai (2009). Assumptions 2.3 and 2.4 on the moments and the limit variance of factors and factor loadings are standard and in line with Bai (2009). Note that these assumptions permit that $T^{-1}\mathbf{G}^{0\prime}\mathbf{G}^0 \xrightarrow{p} \boldsymbol{\Sigma}_G^0$, a positive semi-definite matrix, where $\mathbf{G}^0 = (\mathbf{F}^0, \mathbf{H}^0)$. Assumption 2.5 is sufficient for identification of heterogeneous slope coefficients.

3. ESTIMATION OF MODELS WITH HOMOGENEOUS SLOPES

In line with Norkute et al. (2021), we propose the following two-stage IV procedure:

1 Estimate the span of $\mathbf{F}^0$ by $\widehat{\mathbf{F}}$, defined as $\sqrt{T}$ times the eigenvectors corresponding to the r_1 largest eigenvalues of the $T\times T$ matrix $N^{-1}T^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i'$. Then estimate $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}_{1SIV} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i. \tag{3.6}$$

2 Let $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{1SIV}$. Define $\widehat{\mathbf{H}}$ to be $\sqrt{T}$ times the eigenvectors corresponding to the r_2 largest eigenvalues of the $T \times T$ matrix $(NT)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$. The second-stage estimator of $\boldsymbol{\beta}$ is defined as follows:[3]

$$\widehat{\boldsymbol{\beta}}_{2SIV} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{y}_i. \tag{3.7}$$

In order to establish the asymptotic properties of these estimators, we first expand (3.6) as follows:

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{1SIV} - \boldsymbol{\beta}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i. \tag{3.8}$$

The following Proposition shows the $\sqrt{NT}$-consistency of the first-stage estimator, $\widehat{\boldsymbol{\beta}}_{1SIV}$:

Proposition 3.1. *Under Assumptions 2.1-2.5, we have*

$$N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{u}_i = N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i + \mathbf{b}_{0F} + \mathbf{b}_{1F} + \mathbf{b}_{2F} + O_p(\sqrt{NT}\delta_{NT}^{-3})$$

with

$$\mathbf{b}_{0F} = -N^{-1/2}T^{-1/2}\sum_{i=1}^{N}N^{-1}\sum_{\ell=1}^{N}\boldsymbol{\Gamma}_i^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_\ell^0\mathbf{V}_\ell'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i;$$

$$\mathbf{b}_{1F} = -\sqrt{\frac{T}{N}}\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{h=1}^{N}\mathbb{E}(\mathbf{V}_i'\mathbf{V}_h)\boldsymbol{\Gamma}_h^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}(T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{H}^0\boldsymbol{\varphi}_i^0$$
$$+\sqrt{\frac{T}{N}}\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{\ell=1}^{N}\boldsymbol{\Gamma}_i^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}\boldsymbol{\Gamma}_\ell^0\mathbb{E}(\mathbf{V}_\ell'\mathbf{V}_h)\boldsymbol{\Gamma}_h^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}(T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{H}^0\boldsymbol{\varphi}_i^0;$$

$$\mathbf{b}_{2F} = -\sqrt{\frac{N}{T}}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i^{0\prime}(\boldsymbol{\Upsilon}^0)^{-1}(T^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\Sigma}\mathbf{M}_{\mathbf{F}^0}\mathbf{H}^0\boldsymbol{\varphi}_i^0,$$

where $\boldsymbol{\Upsilon}^0 = \sum_{i=1}^{N}\boldsymbol{\Gamma}_i^0\boldsymbol{\Gamma}_i^{0\prime}/N$, $\boldsymbol{\Sigma} = N^{-1}\sum_{i=1}^{N}\mathbb{E}(\mathbf{V}_i'\mathbf{V}_i)$, *and* $N^{-1/2}T^{-1/2}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i$, $\mathbf{b}_{0F}$, $\mathbf{b}_{1F}$ *and* $\mathbf{b}_{2F}$ *are* $O_p(1)$ *when* $N/T \longrightarrow C$. *Consequently,*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{1SIV} - \boldsymbol{\beta}) = O_p(1).$$

Proposition 3.1 implies that $\widehat{\boldsymbol{\beta}}_{1SIV}$ is consistent but asymptotically biased. Rather than bias-correcting this estimator, we show that the second-stage IV estimator is free from asymptotic bias. To begin with, we make use of the following expansion:

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{2SIV} - \boldsymbol{\beta}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i. \tag{3.9}$$

The next proposition provides an asymptotic representation of $\widehat{\boldsymbol{\beta}}_{2SIV}$.

[3] An alternative estimator, $(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}(\mathbf{y}_i - \mathbf{P}_{\widehat{\mathbf{H}}}\widehat{\mathbf{u}}_i)$, was considered but not included as the finite sample performance was worse than that of $\widehat{\boldsymbol{\beta}}_{2SIV}$.

PROPOSITION 3.2. *Under Assumptions 2.1-2.5, as $N,T\to\infty$, $N/T\to C$, we have*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{2SIV}-\boldsymbol{\beta}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\mathbf{X}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i + O_p(\sqrt{NT}\delta_{NT}^{-3})$$
$$= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{V}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i + O_p(\sqrt{NT}\delta_{NT}^{-3}).$$

Proposition 3.2 shows that the effects of estimating $\mathbf{F}^0$ from $\mathbf{X}_i$ and $\mathbf{H}^0$ from $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{1SIV}$ are asymptotically negligible. Moreover, $\widehat{\boldsymbol{\beta}}_{2SIV}$ is asymptotically equivalent to a least-squares estimator obtained by regressing $(\mathbf{y}_i - \mathbf{H}^0\boldsymbol{\varphi}_i^0)$ on $(\mathbf{X}_i - \mathbf{F}^0\boldsymbol{\Gamma}_i^0)$.

To establish asymptotic normality under weak cross-sectional and serial error dependence, we place additional assumption, which is in line with Assumption E in Bai (2009).

ASSUMPTION 3.1. $\text{plim}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{V}_j/T = \mathbf{B}$, *and* $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i \xrightarrow{d} N(\mathbf{0},\mathbf{B})$, *for some non-random positive definite matrix* $\mathbf{B}$.

Using Proposition 3.2 and Assumption 3.1, it is straightforward to establish the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{2SIV}$:

THEOREM 3.1. *Under Assumptions 2.1-3.1, as $N,T\to\infty$, $N/T\to C$, we have*

$$\sqrt{NT}(\widehat{\boldsymbol{\beta}}_{2SIV}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Psi})$$

where $\boldsymbol{\Psi} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$.

Note that despite the fact that our assumptions permit serial correlation and heteroskedasticity in $\mathbf{v}_{it}$ and ε_{it}, $\widehat{\boldsymbol{\beta}}_{2SIV}$ is not subject to any asymptotic bias. We discuss this property in more detail in Section 5.

As discussed in Bai (2009) and Norkute et al. (2021), in general consistent estimation of $\boldsymbol{\Psi}$ is not feasible when the idiosyncratic errors are both cross-section and time-series dependent. Following Norkute et al. (2021) and Cui et al. (2019), we propose using the following estimator:

$$\widehat{\boldsymbol{\Psi}} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1} \tag{3.10}$$

with $\widehat{\mathbf{A}} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i$ and $\widehat{\mathbf{B}} = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\dot{\mathbf{u}}_i\dot{\mathbf{u}}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i$, where $\dot{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{2SIV}$. In line with the discussion in Hansen (2007), it can be shown that when $\{\mathbf{v}_{it}', \varepsilon_{it}\}$ follows a certain strong mixing process over t and is independent over i, $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} \xrightarrow{p} \mathbf{0}$ as $N,T\to\infty$, $N/T\to C$.

4. MODELS WITH HETEROGENEOUS SLOPES

We now turn our focus on models with heterogeneous coefficients:

$$\begin{aligned}\mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i\,,\\ \mathbf{X}_i &= \mathbf{F}^0\boldsymbol{\Gamma}_i^0 + \mathbf{V}_i.\end{aligned} \tag{4.11}$$

We first consider the following individual-specific estimator

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{y}_i.$$

PROPOSITION 4.1. *Under Assumptions 2.1-2.5, for each i we have*

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = \left(T^{-1}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i\right)^{-1} \times T^{-1/2}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i + O_p\left(\delta_{NT}^{-1}\right) + O_p\left(T^{1/2}\delta_{NT}^{-2}\right)$$

and

$$T^{-1/2}\mathbf{X}_i'\mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_i)$$

where $\boldsymbol{\Omega}_i = T^{-1}\mathrm{plim}_{T\to\infty}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{u}_{is}\tilde{u}_{it}\mathbb{E}(\mathbf{v}_{is}\mathbf{v}_{it}')$ *and* $\tilde{\mathbf{u}}_i = \mathbf{M}_{\mathbf{F}^0}\mathbf{u}_i \equiv (\tilde{u}_{i1}, \cdots, \tilde{u}_{iT})'$.

We also consider inference on the mean of $\boldsymbol{\beta}_i$. We make the following assumptions.

ASSUMPTION 4.1. (RANDOM COEFFICIENTS) $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{e}_i$, *where* $\mathbf{e}_i$ *is independently and identically distributed over i with mean zero and variance* $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$. *Furthermore,* $\mathbf{e}_i$ *is independent with* $\boldsymbol{\Gamma}_j^0$, $\boldsymbol{\varphi}_j^0$, ε_{jt}, $\mathbf{v}_{jt}$, $\mathbf{f}_t^0$ *and* $\mathbf{h}_t^0$ *for all* i, j, t.

ASSUMPTION 4.2. (MOMENTS) *For each i,* $\mathbb{E}\left\|T^{-1/2}\mathbf{V}_i'\mathbf{F}^0\right\|^4 \le C$, $\left\|T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{\Sigma}\mathbf{F}^0\right\|^4 \le C$, $\mathbb{E}\|\frac{1}{\sqrt{NT}}\sum_{\ell=1}^{N}\boldsymbol{\varepsilon}_i'\mathbf{V}_\ell\boldsymbol{\Gamma}_\ell^{0\prime}\|^4 \le C$, $\mathbb{E}\left\|T^{-1/2}\sum_{t=1}^{T}\left[\mathbf{V}_i'\mathbf{V}_i - \boldsymbol{\Sigma}\right]\right\|^4 \le C$, $\mathbb{E}\left\|N^{-1/2}T^{-1/2}\sum_{\ell=1}^{N}\left(\mathbf{V}_i'\mathbf{V}_\ell - \mathbb{E}\left(\mathbf{V}_i'\mathbf{V}_\ell\right)\right)\boldsymbol{\Gamma}_\ell^{0\prime}\right\|^4 \le C$, *and* $0 < C_{min} \le \|\boldsymbol{\Sigma}\| \le C$.

In line with Norkute et al. (2021), we propose the following mean-group IV (MGIV) estimator:

$$\widehat{\boldsymbol{\beta}}_{MGIV} = N^{-1}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}. \tag{4.12}$$

THEOREM 4.1. *Under Assumptions 2.1-2.5 and 4.1-4.2, we have*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{MGIV} - \boldsymbol{\beta}) = N^{-1/2}\sum_{i=1}^{N}\mathbf{e}_i + O_p(N^{3/4}T^{-1}) + O_p(NT^{-3/2}) + O_p(N^{1/2}\delta_{NT}^{-2}),$$

such that for $N^3/T^4 \to 0$ *as* $N, T \to \infty$, *we obtain*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{MGIV} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}).$$

Furthermore, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} - \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{0}$, *where*

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{MGIV})(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_{MGIV})'. \tag{4.13}$$

5. ASYMPTOTIC COMPARISON OF $\widehat{\boldsymbol{\beta}}_{2SIV}$ WITH EXISTING ESTIMATORS

This section investigates asymptotic bias properties and relative efficiency of the 2SIV, IPC, PC and CA estimators for the models with homogeneous slopes. For this purpose, let $\mathbf{G}^0 = (\mathbf{F}^0, \mathbf{H}^0)$ denote a $T \times r$ matrix, where $r = r_1 + r_2$. We shall assume that $\mathbf{G}^{0\prime}\mathbf{G}^0/T \xrightarrow{p} \boldsymbol{\Sigma}_G^0 > 0$, a positive definite matrix. Note that, together with Assumption 2.3, this implies that $\mathbf{F}^0$ and $\mathbf{H}^0$ are linearly independent of each other (and can be correlated), which is slightly stronger than Assumption 2.3.

5.1. 2SIV estimator

Recall that $\mathbf{X}_i = \mathbf{F}^0\mathbf{\Gamma}_i^0 + \mathbf{V}_i$ and $\mathbf{u}_i = \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i$. Proposition 3.2 in Appendix B demonstrates that under Assumptions 2.1-2.5 $\left(N^{-1}T^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}_{2SIV} - \boldsymbol{\beta}\right)$ can be expanded as follows:

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i + \mathbf{b}_{0FH} + \mathbf{b}_{1FH} + \mathbf{b}_{2FH} + O_p\left(\sqrt{NT}\delta_{NT}^{-3}\right), \tag{5.14}$$

where

$$\mathbf{b}_{0FH} = -\frac{1}{N^{1/2}}\frac{1}{NT^{1/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{\Gamma}_i^{0\prime}(\mathbf{\Upsilon}^0)^{-1}\mathbf{\Gamma}_j^0 + \boldsymbol{\varphi}_j^{0\prime}(\mathbf{\Upsilon}_\varphi^0)^{-1}\boldsymbol{\varphi}_i^0)\mathbf{V}_j'\boldsymbol{\varepsilon}_i;$$

$$\mathbf{b}_{1FH} = -\frac{1}{N^{1/2}}\frac{1}{N^2T^{1/2}}\sum_{i=1}^{N}\sum_{\ell=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_i^{0\prime}(\mathbf{\Upsilon}^0)^{-1}\mathbf{\Gamma}_\ell^0\left(\mathbf{V}_\ell'\boldsymbol{\varepsilon}_j\right)\boldsymbol{\varphi}_j^{0\prime}(\mathbf{\Upsilon}_\varphi^0)^{-1}\boldsymbol{\varphi}_i^0;$$

$$\mathbf{b}_{2FH} = -\frac{1}{T^{1/2}}\frac{1}{N^{3/2}T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_i^{0\prime}(\mathbf{\Upsilon}^0)^{-1}\mathbf{\Gamma}_j^0\mathbf{V}_j'\mathbf{\Sigma}_\varepsilon\mathbf{H}^0\left(\frac{\mathbf{H}^{0\prime}\mathbf{H}^0}{T}\right)^{-1}(\mathbf{\Upsilon}_\varphi^0)^{-1}\boldsymbol{\varphi}_i^0,$$

with $\mathbf{\Sigma}_\varepsilon = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'\right)$. It is easily seen that (see proof of Proposition 3.2) $\mathbf{b}_{0FH} = O_p\left(N^{-1/2}\right)$, $\mathbf{b}_{1FH} = O_p\left(N^{-1/2}\right)$ and $\mathbf{b}_{2FH} = O_p\left(T^{-1/2}\right)$. Hence, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{F}}}\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\boldsymbol{\varepsilon}_i + o_p(1).$$

5.2. Asymptotic bias of Bai's (2009a) IPC-type estimator

It is instructive to consider a PC estimator that is asymptotically equivalent to Bai (2009) but avoids iterations:

$$\widetilde{\boldsymbol{\beta}}_{2SIV} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{y}_i.$$

Observe that this estimator projects out $\widehat{\mathbf{H}}$ from $(\mathbf{X}_i, \mathbf{y}_i)$, but it does not eliminate $\widehat{\mathbf{F}}$ from $\mathbf{X}_i$. $\widehat{\mathbf{H}}$ is estimated using the residuals of the first-stage IV estimator, $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_{1SIV}$.

Using similar derivations as in Section 5.1, Proposition 5.1 below shows that $\left(N^{-1}T^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{X}_i\right)\times \sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}_{2SIV} - \boldsymbol{\beta}\right)$ has the following asymptotic expansion:

Proposition 5.1. *Under Assumptions 2.1-2.5, we have*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{H}}}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i + \mathbf{b}_{0H} + \mathbf{b}_{1H} + \mathbf{b}_{2H} + O_p\left(\sqrt{NT}\delta_{NT}^{-3}\right) \tag{5.15}$$

with

$$\mathbf{b}_{0H} = -\frac{1}{N^{3/2}T^{1/2}}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\mathbf{X}_j'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\varepsilon}_i;$$

$$\mathbf{b}_{1H} = -\sqrt{\frac{T}{N}}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\mathcal{X}}_i'\mathbf{H}^0\left(\frac{\mathbf{H}^{0\prime}\mathbf{H}^0}{T}\right)^{-1}(\boldsymbol{\Upsilon}_{\varphi}^0)^{-1}\boldsymbol{\varphi}_j^0\mathbb{E}\left(\boldsymbol{\varepsilon}_j'\boldsymbol{\varepsilon}_i/T\right);$$

$$\mathbf{b}_{2H} = -\sqrt{\frac{N}{T}}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\mathbf{H}^0}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{H}^0\left(\frac{\mathbf{H}^{0\prime}\mathbf{H}^0}{T}\right)^{-1}(\boldsymbol{\Upsilon}_{\varphi}^0)^{-1}\boldsymbol{\varphi}_i^0,$$

where $a_{ij} = \boldsymbol{\varphi}_j^{0\prime}(\boldsymbol{\Upsilon}_{\varphi}^0)^{-1}\boldsymbol{\varphi}_i^0$, $\boldsymbol{\mathcal{X}}_i = \mathbf{X}_i - N^{-1}\sum_{\ell=1}^{N} a_{i\ell}\mathbf{X}_\ell$ *and* $\boldsymbol{\Sigma}_{\varepsilon} = \frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'\right)$.

The above asymptotic bias terms are identical to those of the IPC estimator of Bai (2009). As a result, it suffices to compare $\widehat{\boldsymbol{\beta}}_{2SIV}$ with $\widetilde{\boldsymbol{\beta}}_{2SIV}$. Incidentally, as shown in Bai (2009), the term $\mathbf{b}_{0H}$ tends to a normal random vector, which necessitates the transformation of the regressor matrix to $\boldsymbol{\mathcal{X}}_i$; see equation (5.16) below.

The terms $\mathbf{b}_{0H}$, $\mathbf{b}_{1H}$ and $\mathbf{b}_{2H}$ in (5.15) are comparable to the terms $\mathbf{b}_{0FH}$, $\mathbf{b}_{1FH}$ and $\mathbf{b}_{2FH}$, respectively, in (5.14). One striking result is that $\mathbf{b}_{0H}$, $\mathbf{b}_{1H}$ and $\mathbf{b}_{2H}$ are not asymptotically ignorable, whereas $\mathbf{b}_{0FH}$, $\mathbf{b}_{1FH}$ and $\mathbf{b}_{2FH}$ are. This difference stems solely from the fact that $\widehat{\boldsymbol{\beta}}_{2SIV}$ asymptotically projects out $\mathbf{F}^0\boldsymbol{\Gamma}_i^0$ from $\mathbf{X}_i$ and $\mathbf{H}^0\boldsymbol{\varphi}_i^0$ from $\mathbf{u}_i$ separately, whereas $\widetilde{\boldsymbol{\beta}}_{2SIV}$ projects out $\mathbf{H}^0\boldsymbol{\varphi}_i^0$ from $\mathbf{u}_i$ *only*. Therefore, the asymptotic bias terms of $\widetilde{\boldsymbol{\beta}}_{2SIV}$, $\mathbf{b}_{0H}$, $\mathbf{b}_{1H}$ and $\mathbf{b}_{2H}$, contain correlations between the regressors $\mathbf{X}_i$ and the disturbance $\mathbf{u}_i(= \mathbf{H}^0\boldsymbol{\varphi}_i^0 + \boldsymbol{\varepsilon}_i)$ since the estimation error of $\widehat{\mathbf{H}}$ contains $\mathbf{u}_i$. Recalling that $\mathbf{X}_i = \mathbf{F}^0\boldsymbol{\Gamma}_i^0 + \mathbf{V}_i$, such correlations may not be asymptotically negligible because $\mathbf{H}^{0\prime}\mathbf{F}^0/T = O_p(1)$ and $\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{0\prime}\text{vec}\left(\boldsymbol{\Gamma}_i^0\right)/N = O_p(1)$.

On the other hand, $\widehat{\boldsymbol{\beta}}_{2SIV}$ asymptotically projects out $\mathbf{F}^0\boldsymbol{\Gamma}_i^0$ from $\mathbf{X}_i$ as well as $\mathbf{H}^0\boldsymbol{\varphi}_i^0$ from $\mathbf{u}_i$. Therefore, $\mathbf{b}_{0FH}$, $\mathbf{b}_{1FH}$ and $\mathbf{b}_{2FH}$ contain correlations between $\mathbf{M}_{\mathbf{F}^0}\mathbf{X}_i = \mathbf{M}_{\mathbf{F}^0}\mathbf{V}_i$ and $\mathbf{u}_i$. Since $\mathbf{V}_i$, $\mathbf{H}^0\boldsymbol{\varphi}_i^0$ and $\boldsymbol{\varepsilon}_i$ are independent of each other, such correlations are asymptotically negligible. As a result, our estimator $\widehat{\boldsymbol{\beta}}_{2SIV}$ does not suffer from asymptotic bias.

Using similar reasoning, it turns out that in some special cases, some of the bias terms of $\widetilde{\boldsymbol{\beta}}_{2SIV}$ may disappear as well. For instance, when $\mathbf{F}^0 \subseteq \mathbf{H}^0$, we have $\mathbf{M}_{\mathbf{H}^0}\mathbf{X}_j = \mathbf{M}_{\mathbf{H}^0}\mathbf{V}_j$ because $\mathbf{M}_{\mathbf{H}^0}\mathbf{F}^0 = \mathbf{0}$. Thus, $\mathbf{b}_{0H} = O_p\left(N^{-1/2}\right)$ and $\mathbf{b}_{2H} = O_p\left(T^{-1/2}\right)$ although $\mathbf{b}_{1H}$ remains $O_p(1)$. Note that under our assumptions all three bias terms, $\mathbf{b}_{0H}$, $\mathbf{b}_{1H}$ and $\mathbf{b}_{2H}$, are asymptotically negligible only if $\mathbf{H}^0 = \mathbf{F}^0$, which can be a highly restrictive condition in practice.[4]

5.3. Asymptotic bias of PC and CA estimators

Pesaran (2006) and Westerlund and Urbain (2015) put forward pooled estimators in which the whole set of factors in $\mathbf{X}_i$ and $\mathbf{u}_i$ are estimated *jointly*, rather than *separately*. This feature makes these estimators asymptotically biased. To show this, we rewrite the

[4]When $\varepsilon_{it} \sim i.i.d.(0,\sigma^2)$, $\mathbf{b}_{0H}$ remains $O_p(1)$ whilst $\mathbf{b}_{1H}$ and $\mathbf{b}_{2H}$ become asymptotically negligible. See Corollary 1 in Bai (2009).

model as

$$\mathbf{Z}_i = (\mathbf{y}_i, \mathbf{X}_i) = \mathbf{G}^0 \boldsymbol{\Lambda}_i^0 + \mathbf{U}_i,$$

where

$$\boldsymbol{\Lambda}_i^0 = \begin{pmatrix} \boldsymbol{\Gamma}_i^0 \boldsymbol{\beta} & \boldsymbol{\Gamma}_i^0 \\ \boldsymbol{\varphi}_i^0 & \mathbf{0} \end{pmatrix}, \ \mathbf{U}_i = (\mathbf{V}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \mathbf{V}_i).$$

We also define $\boldsymbol{\Upsilon}_\Lambda^0 = N^{-1} \sum_{i=1}^N \boldsymbol{\Lambda}_i^0 \boldsymbol{\Lambda}_i^{0\prime}$ which is assumed to be positive definite.

In the PC approach of Westerlund and Urbain (2015), a span of $\mathbf{G}^0$ is estimated as $\sqrt{T}$ times the eigenvectors corresponding to the first r largest eigenvalues of $\sum_{i=1}^N \mathbf{Z}_i \mathbf{Z}_i'/N$, which is denoted by $\widehat{\mathbf{G}}_z$. The resulting PC estimator is defined as

$$\widehat{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{G}}_z} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{G}}_z} \mathbf{y}_i.$$

In line with Pesaran (2006), the CA estimator of Westerlund and Urbain (2015) approximates a span of $\mathbf{G}^0$ by a linear combination of $\bar{\mathbf{Z}} = N^{-1} \sum_{i=1}^N \mathbf{Z}_i$. The associated CA estimator is given by

$$\widehat{\boldsymbol{\beta}}_{CA} = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\bar{\mathbf{Z}}} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\bar{\mathbf{Z}}} \mathbf{y}_i.$$

As discussed in Westerlund and Urbain (2015), both PC and CA are asymptotically biased due to the correlation between the estimation error of $\widehat{\mathbf{G}}_z$ and $\{\mathbf{X}_i, \mathbf{u}_i\}$. The estimation error of $\widehat{\mathbf{G}}_z$ contains the error term of the system equation $\mathbf{U}_i$, which is a function of both $\mathbf{V}_i$ and $\boldsymbol{\varepsilon}_i$. Therefore, the estimation error of $\widehat{\mathbf{G}}_z$ is correlated with $\mathbf{M}_\mathbf{G} \mathbf{X}_i$ and $\mathbf{M}_\mathbf{G} \mathbf{u}_i$, which causes the asymptotic bias. In what follows, we shall focus on the PC estimator as the bias analysis for the CA estimator is very similar.

Following Westerlund and Urbain (2015), we expand $\left(N^{-1} T^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{G}}_z} \mathbf{X}_i \right) \sqrt{NT} \left(\widehat{\boldsymbol{\beta}}_{PC} - \boldsymbol{\beta} \right)$ as follows:

Proposition 5.2. *Under Assumptions 2.1-2.5*

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{G}}_z} \mathbf{u}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{V}_i' \boldsymbol{\varepsilon}_i + \mathbf{b}_{1G} + \mathbf{b}_{2G} + \mathbf{b}_{3G} + O_p\left(\sqrt{NT} \delta_{NT}^{-3} \right),$$

with

$$\mathbf{b}_{1G} = -\sqrt{\frac{T}{N}} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \left(\boldsymbol{\Gamma}_i^{0\prime}, \mathbf{0}' \right) (\boldsymbol{\Upsilon}_\Lambda^0)^{-1} \boldsymbol{\Lambda}_j^0 \mathbb{E} \left(\mathbf{U}_j' \boldsymbol{\varepsilon}_i / T \right);$$

$$\mathbf{b}_{2G} = -\sqrt{\frac{T}{N}} \frac{1}{N^2} \sum_{i=1}^N \sum_{\ell=1}^N \sum_{j=1}^N \left(\boldsymbol{\Gamma}_i^{0\prime}, \mathbf{0}' \right) (\boldsymbol{\Upsilon}_\Lambda^0)^{-1} \boldsymbol{\Lambda}_\ell^0 \mathbb{E} \left(\mathbf{U}_\ell' \mathbf{U}_j / T \right) \boldsymbol{\Lambda}_j^{0\prime} (\boldsymbol{\Upsilon}_\Lambda^0)^{-1} \left(\frac{\mathbf{G}^{0\prime} \mathbf{G}^0}{T} \right)^{-1} \frac{\mathbf{G}^{0\prime} \mathbf{H}^0}{T} \boldsymbol{\varphi}_i^0;$$

$$\mathbf{b}_{3G} = -\sqrt{\frac{T}{N}} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \frac{\mathbf{V}_i' \mathbf{U}_j}{T} \boldsymbol{\Lambda}_j^{0\prime} (\boldsymbol{\Upsilon}_\Lambda^0)^{-1} \left(\frac{\mathbf{G}^{0\prime} \mathbf{G}^0}{T} \right)^{-1} \frac{\mathbf{G}^{0\prime} \mathbf{H}^0}{T} \boldsymbol{\varphi}_i^0.$$

It is easily seen that $\mathbf{b}_{1G}$, $\mathbf{b}_{2G}$ and $\mathbf{b}_{3G}$ are all $O_p(1)$. Note that the asymptotic bias terms are functions of $\boldsymbol{\Lambda}_\ell^0$ and $\boldsymbol{\Upsilon}_\Lambda^0$, which depend on the slope coefficient vector $\boldsymbol{\beta}$.

5.4. Relative asymptotic efficiency of 2SIV, IPC, PC and CA estimators

Finally, we compare the asymptotic efficiency of the estimators. To make the problem tractable and as succinct as possible, we shall assume that ε_{it} is i.i.d. over i and t with $\mathbb{E}(\varepsilon_{it}) = 0$ and $\mathbb{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2$. In this case, it is easily seen that the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{2SIV}$ is

$$\boldsymbol{\Psi} = \sigma_\varepsilon^2 \left(\text{plim} N^{-1} T^{-1} \sum_{i=1}^{N} \mathbf{V}_i' \mathbf{V}_i \right)^{-1}.$$

Next, using Proposition 5.2, consider the bias-corrected PC estimator

$$\widehat{\boldsymbol{\beta}}_{PC}^{*} = \widehat{\boldsymbol{\beta}}_{PC} - N^{1/2} T^{1/2} \left(\sum_{i=1}^{N} \mathbf{V}_i' \mathbf{V}_i \right)^{-1} (\mathbf{b}_{1G} + \mathbf{b}_{2G} + \mathbf{b}_{3G}).$$

We can see that the asymptotic variance of the bias-corrected PC estimator is identical to $\boldsymbol{\Psi}$. Therefore, the 2SIV and the bias-corrected PC estimators are asymptotically equivalent.

Consider now $\widetilde{\boldsymbol{\beta}}_{2SIV}$. Noting that $\mathbf{b}_{0H}$ tends to a normal distribution, and following Bai (2009), the bias-corrected estimator with transformed regressors can be written as:

$$\widetilde{\boldsymbol{\beta}}_{2SIV}^{*} = \widetilde{\boldsymbol{\beta}}_{2SIV}^{+} - N^{1/2} T^{1/2} \left(\sum_{i=1}^{N} \boldsymbol{\mathcal{X}}_i' \mathbf{M}_{\mathbf{H}^0} \boldsymbol{\mathcal{X}}_i \right)^{-1} (\mathbf{b}_{1H} + \mathbf{b}_{2H}),$$

where

$$\widetilde{\boldsymbol{\beta}}_{2SIV}^{+} = \left(\sum_{i=1}^{N} \boldsymbol{\mathcal{X}}_i' \mathbf{M}_{\mathbf{H}^0} \boldsymbol{\mathcal{X}}_i \right)^{-1} \sum_{i=1}^{N} \boldsymbol{\mathcal{X}}_i' \mathbf{M}_{\mathbf{H}^0} \mathbf{y}_i. \tag{5.16}$$

The asymptotic variance of this bias-corrected estimator is given by

$$\widetilde{\boldsymbol{\Psi}} = \sigma_\varepsilon^2 \left(\text{plim} N^{-1} T^{-1} \sum_{i=1}^{N} \boldsymbol{\mathcal{X}}_i' \mathbf{M}_{\mathbf{H}^0} \boldsymbol{\mathcal{X}}_i \right)^{-1}.$$

There exist two differences compared to $\boldsymbol{\Psi}$. First, in general $\mathbf{M}_{\mathbf{H}^0} \mathbf{X}_i \neq \mathbf{M}_{\mathbf{H}^0} \mathbf{V}_i$ as the factors in $\mathbf{X}_i$ may not be identical to the factors in $\mathbf{u}_i$. Second, regressors are to be transformed as $\boldsymbol{\mathcal{X}}_i = \mathbf{X}_i - N^{-1} \sum_{\ell=1}^{N} a_{i\ell} \mathbf{X}_\ell$. Therefore, $\boldsymbol{\Psi} - \widetilde{\boldsymbol{\Psi}}$ can be positive semi-definite or negative-semi-definite. Thus, the asymptotic efficiency of the bias-corrected IPC estimator of Bai (2009) relative to 2SIV and the bias-corrected PC/CA estimators, is indeterminate. However, in the special case where $\mathbf{F}^0 \subseteq \mathbf{H}^0$, we have $\mathbf{M}_{\mathbf{H}^0} \boldsymbol{\mathcal{X}}_i = \mathbf{M}_{\mathbf{H}^0} \boldsymbol{\mathcal{V}}_i$, with $\boldsymbol{\mathcal{V}}_i = \mathbf{V}_i - N^{-1} \sum_{\ell=1}^{N} a_{i\ell} \mathbf{V}_\ell$. The second term of $\boldsymbol{\mathcal{V}}_i$ is $O_p(N^{-1/2})$ because $\mathbf{V}_\ell$ and $a_{i\ell}$ are independent. Hence, in this case $\widetilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi}$, and the bias-corrected IPC estimator is asymptotically as efficient as the bias-corrected PC/CA estimator and 2SIV.

6. MONTE CARLO SIMULATIONS

We conduct a Monte Carlo simulation exercise in order to assess the finite sample behaviour of the statistics discussed above in terms of bias, standard deviation (s.d.), root mean squared error (RMSE), empirical size and power of the t-test. More specifically, we investigate the performance of 2SIV, defined in (3.7), and MGIV defined in (4.12). For the purposes of comparison, we also consider the (bias-corrected) IPC of Bai (2009) and the PC estimator, labeled as (BC-)IPC and (BC-)PC respectively, the CA estimator, as well as the mean-group versions of PC and CA (denoted as MGPC and MGCA), which were put forward by Pesaran (2006), Westerlund and Urbain (2015) and Reese and Westerlund (2018). The t-statistics for 2SIV and MGIV are computed using the variance estimators defined by (3.10) and (4.13), respectively. The t-statistics for IPC, PC and CA estimators and their MG versions (if any) employ analogous variance estimators.

6.1. Design

We consider the following panel data model:

$$y_{it} = \alpha_i + \sum_{\ell=1}^{k} \beta_{\ell i} x_{\ell it} + u_{it}; \quad u_{it} = \sum_{s=1}^{r_2} \gamma_{si}^0 f_{s,t}^0 + \varepsilon_{it}, \tag{6.17}$$

$i = 1, ..., N$, $t = -49, ..., T$, where the process for the covariates is given by

$$x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{r_1} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell it}; \quad i = 1, 2, ..., N; \quad t = -49, -48, ..., T. \tag{6.18}$$

We set $k = 2$, $r_2 = 2$ and $r_1 = 3$. This implies that the first two factors in u_{it}, f_{1t}^0 and f_{2t}^0, are also in the DGP of $x_{\ell it}$ for $\ell = 1, 2$, while f_{3t}^0 is included in $x_{\ell it}$ only. Observe that, using notation of earlier sections, $\mathbf{h}_t^0 = (f_{1t}^0, f_{2t}^0)'$ and $\mathbf{f}_t^0 = (f_{1t}^0, f_{2t}^0, f_{3t}^0)'$. [5]

The factors $f_{s,t}^0$ are generated using the following AR(1) process $f_{s,t}^0 = \rho_{fs} f_{s,t-1}^0 + (1 - \rho_{fs}^2)^{1/2} \zeta_{s,t}$ where $\rho_{fs} = 0.5$ and $\zeta_{s,t} \sim i.i.d.N(0, 1)$ for $s = 1, ..., 3$. The idiosyncratic error of y_{it}, ε_{it}, is non-normal and heteroskedastic across both i and t, such that $\varepsilon_{it} = \varsigma_\varepsilon \sigma_{it}(\epsilon_{it} - 1)/\sqrt{2}$, $\epsilon_{it} \sim i.i.d.\chi_1^2$, with $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \sim i.i.d.\chi_2^2/2$, and $\varphi_t = t/T$ for $t = 0, 1, ..., T$ and unity otherwise. We define $\pi_u := \varsigma_\varepsilon^2 / \left(r_2 + \varsigma_\varepsilon^2\right)$ which is the proportion of the average variance of u_{it} due to ε_{it}. This implies $\varsigma_\varepsilon^2 = \pi_u r_2 \left(1 - \pi_u\right)^{-1}$. We set ς_ε^2 such that $\pi_u \in \{1/4, 3/4\}$.

The idiosyncratic errors of the covariates follow an AR(1) process $v_{\ell it} = \rho_{\upsilon,\ell} v_{\ell it-1} + (1 - \rho_{\upsilon,\ell}^2)^{1/2} \varpi_{\ell it}$; $\varpi_{\ell it} \sim i.i.d.N(0, \varsigma_\upsilon^2)$ for $\ell = 1, 2$. We set $\rho_{\upsilon,\ell} = 0.5$ for all ℓ. We define the signal-to-noise ratio (SNR) as $SNR := \left(\beta_1^2 + \beta_2^2\right) \varsigma_\upsilon^2 \varsigma_\varepsilon^{-2}$ where $\rho_\upsilon = \rho_{\upsilon,\ell}$ for $\ell = 1, 2$. Solving for ς_υ^2 gives $\varsigma_\upsilon^2 = \varsigma_\varepsilon^2 SNR \left(\beta_1^2 + \beta_2^2\right)^{-1}$. We set $SNR = 4$, which lies within the values considered by Bun and Kiviet (2006) and Juodis and Sarafidis (2018, 2020).

The individual-specific effects are generated by drawing initially *mean-zero* random variables as $\mu_{\ell i}^* = \rho_{\mu,\ell} \alpha_i^* + (1 - \rho_{\mu,\ell}^2)^{1/2} \omega_{\ell i}$, where $\alpha_i^* \sim i.i.d.N(0, 1)$, $\omega_{\ell i} \sim i.i.d.N(0, 1)$, for $\ell = 1, 2$. We set $\rho_{\mu,\ell} = 0.5$ for $\ell = 1, 2$. Subsequently, we set $\alpha_i = \alpha + \alpha_i^*$ and $\mu_{\ell i} = \mu_\ell + \mu_{\ell i}^*$ where $\alpha = 1/2$, $\mu_1 = 1$, $\mu_2 = -1/2$, for $\ell = 1, 2$.

Similarly, the factor loadings in u_{it} are generated at first instance as *mean-zero* random variables such that $\gamma_{si}^{0*} \sim i.i.d.N(0, 1)$ for $s = 1, ..., r_2 = 2$, $\ell = 1, 2$; the factor loadings in x_{1it} and x_{2it} are generated as

$$\gamma_{\ell si}^{0*} = \rho_{\gamma,\ell s} \gamma_{si}^{0*} + (1 - \rho_{\gamma,\ell s}^2)^{1/2} \xi_{\ell si}; \quad \xi_{\ell si} \sim i.i.d.N(0, 1) \text{ for } s = 1, 2; \tag{6.19}$$

$$\gamma_{13i}^{0*} = \rho_{\gamma,13} \gamma_{1i}^{0*} + (1 - \rho_{\gamma,13}^2)^{1/2} \xi_{13i}; \quad \xi_{13i} \sim i.i.d.N(0, 1); \tag{6.20}$$

[5] Tables E1-E3 in Appendix E present results for a different specification, where $r_2 = 3$ and $r_1 = 2$. To save space, we do not discuss these results here but it suffices to say that the conclusions are similar to those in Section 6.2.

$$\gamma_{23i}^{0*} = \rho_{\gamma,23}\gamma_{2i}^{0*} + (1-\rho_{\gamma,23}^2)^{1/2}\xi_{23i}; \quad \xi_{23i} \sim i.i.d.N(0,1). \tag{6.21}$$

The process (6.19) allows the factor loadings to $f_{1,t}^0$ and $f_{2,t}^0$ in x_{1it} and x_{2it} to be correlated with the factor loadings to $f_{1,t}^0$ and $f_{2,t}^0$ in u_{it}. On the other hand, (6.20) and (6.21) ensure that the factor loadings to $f_{3,t}^0$ in x_{1it} and x_{2it} can be correlated with the factor loadings to $f_{1,t}^0$ and $f_{2,t}^0$ in u_{it}. We consider $\rho_{\gamma,11} = \rho_{\gamma,12} = \rho_{\gamma,21} = \rho_{\gamma,22} = \rho_{\gamma,13} = \rho_{\gamma,23} = 0.5$. The factor loadings that enter into the model are then generated as

$$\mathbf{\Gamma}_i^0 = \mathbf{\Gamma}^0 + \mathbf{\Gamma}_i^{0*} \tag{6.22}$$

where

$$\mathbf{\Gamma}_i^0 = \begin{pmatrix} \gamma_{1i}^0 & \gamma_{11i}^0 & \gamma_{21i}^0 \\ \gamma_{2i}^0 & \gamma_{12i}^0 & \gamma_{22i}^0 \\ 0 & \gamma_{13i}^0 & \gamma_{23i}^0 \end{pmatrix} \text{ and } \mathbf{\Gamma}_i^{0*} = \begin{pmatrix} \gamma_{1i}^{0*} & \gamma_{11i}^{0*} & \gamma_{21i}^{0*} \\ \gamma_{2i}^{0*} & \gamma_{12i}^{0*} & \gamma_{22i}^{0*} \\ 0 & \gamma_{13i}^{0*} & \gamma_{23i}^{0*} \end{pmatrix}.$$

Observe that, using notation of earlier sections, $\boldsymbol{\gamma}_{yi}^0 = (\gamma_{1i}^0, \gamma_{2i}^0)'$ and $\mathbf{\Gamma}_{x,i}^0 = (\boldsymbol{\gamma}_{1i}^0, \boldsymbol{\gamma}_{2i}^0, \boldsymbol{\gamma}_{3i}^0)'$ with $\boldsymbol{\gamma}_{\ell i}^0 = (\gamma_{\ell 1i}^0, \gamma_{\ell 2i}^0, \gamma_{\ell 3i}^0)'$ for $\ell = 1, 2$. It is easily seen that the average of the factor loadings is $E\left(\mathbf{\Gamma}_i^0\right) = \mathbf{\Gamma}^0$. We set

$$\mathbf{\Gamma}^0 = \begin{pmatrix} \gamma_1^0 & \gamma_{11}^0 & \gamma_{21}^0 \\ \gamma_2^0 & \gamma_{12}^0 & \gamma_{22}^0 \\ 0 & \gamma_{13}^0 & \gamma_{23}^0 \end{pmatrix} = \begin{pmatrix} 1/4 & 1/4 & -1 \\ 1/2 & -1 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}. \tag{6.23}$$

The slope coefficients in (6.17) are generated as

$$\beta_{1i} = \beta_1 + \eta_{\beta_1 i}; \quad \beta_{2i} = \beta_2 + \eta_{\beta_2 i}, \tag{6.24}$$

such that $\beta_1 = 3$ and $\beta_2 = 1$. In the case of homogeneous slopes, we impose $\rho_i = \rho$, $\beta_{1i} = \beta_1$ and $\beta_{2i} = \beta_2$, whereas in the case of heterogeneous slopes, we specify $\eta_{\rho i} \sim i.i.d.U\left[-c, +c\right]$, and $\eta_{\beta_\ell i} = \left[(2c)^2/12\right]^{1/2} \rho_\beta \xi_{\beta_\ell i} + \left(1-\rho_\beta^2\right)^{1/2} \eta_{\rho i}$, where $\xi_{\beta_\ell i}$ is the standardised squared idiosyncratic errors in $x_{\ell it}$, computed as $\xi_{\beta_\ell i} = \left(\overline{v_{\ell i}^2} - \overline{v_\ell^2}\right)\left[N^{-1}\sum_{i=1}^N \left(\overline{v_{\ell i}^2} - \overline{v_\ell^2}\right)^2\right]^{-1/2}$, with $\overline{v_{\ell i}^2} = T^{-1}\sum_{t=1}^T v_{\ell it}^2$, $\overline{v_\ell^2} = N^{-1}\sum_{i=1}^N \overline{v_{\ell i}^2}$, for $\ell = 1, 2$. We set $c = 1/5$, $\rho_\beta = 0.4$ for $\ell = 1, 2$.

We consider various combinations of (T, N), i.e. $T \in \{25, 50, 100, 200\}$ and $N \in \{25, 50, 100, 200\}$. The results are obtained based on 2,000 replications, and all tests are conducted at the 5% significance level. For the size of the "t-test", $H_0 : \beta_\ell = \beta_\ell^0$ for $\ell = 1, 2$, where β_1^0 and β_2^0 are the true parameter values. For the power of the test, $H_0 : \beta_\ell = \beta_\ell^0 + 0.1$ for $\ell = 1, 2$ against two sided alternatives are considered.

Prior to computing the estimators except for CA and MGCA , the data are demeaned using the within transformation in order to eliminate individual-specific effects. For the CA and MGCA estimators, the untransformed data are used, but a $T \times 1$ vector of ones is included along with the cross-sectional averages. The number of factors r_1 and r_2 are estimated in each replication using the eigenvalue ratio (ER) statistic proposed by Ahn and Horenstein (2013).

6.2. Results

Tables 1–2 report results for β_1 in terms of bias, standard deviation, RMSE, empirical size and power for the model in (6.17).[6]

Table 1 focuses on the case where $N = T = 200$ and π_u alternates between $\{1/4, 3/4\}$. Consider first the homogeneous model with $\pi_u = 3/4$. As we can see, the bias ($\times 100$) for 2SIV and MGIV is very close to zero and takes the smallest value compared to the remaining estimators. The bias of BC-IPC is larger in absolute value than that of IPC but of opposite sign. This may suggest that bias-correction over-corrects in this case. MGPC and PC perform similarly and exhibit larger bias than IPC. Last, both CA and MGCA are subject to substantial bias, which is not surprising as these estimators may require bias-correction in the present DGP.

In regards to the dispersion of the estimators, the standard deviation of 2SIV and PC is very similar, which is in line with our theoretical results. For this specific design, IPC takes the smallest s.d. value among the estimators under consideration. On the other hand, when it comes to the bias-corrected estimators, bias-correction appears to inflate dispersion and thus the standard deviation of BC-IPC and BC-PC is relatively large (equal to 0.805 and 0.885, respectively). As a result, 2SIV outperforms BC-IPC and BC-PC, with a s.d. value equal to 0.586.

In terms of RMSE, IPC appears to perform best, although this estimator is not recommended in practice due to its asymptotic bias. 2SIV takes the second smallest RMSE value, followed by MGIV. CA and MGCA exhibit the largest RMSE values, an outcome that reflects the large bias of these estimators.

Next, we turn our attention to the model with heterogeneous slopes and $\pi_u = 3/4$. In comparison to the homogeneous model, all estimators suffer a substantial increase in bias; the only exception is MGIV, which has the smallest bias. MGPC and MGCA are severely biased, both in absolute magnitude as well as relative to the remaining inconsistent estimators. The s.d. values of MGIV and MGPC are very similar and relatively small compared to the other estimators. The smallest RMSE value is that of MGIV.

We now discuss the results in the lower panel of Table 1, which correspond to $\pi_u = 1/4$. The relative performance of the estimators is similar to the case where $\pi_u = 3/4$, except for a noticeable improvement in the performance of BC-IPC. Thus, the results for BC-IPC and IPC are quite comparable, suggesting that the bias-correction term is close to zero and so over-correction is avoided. The results for 2SIV are very similar to those for $\pi_u = 3/4$, which indicates that the estimator is robust to different values of the variance ratio. The conclusions with heterogeneous slopes for $\pi_u = 1/4$ are similar to those for $\pi_u = 3/4$.

In regards to inference, the size of the t-test associated with 2SIV and MGIV is close to the nominal value of 5% under the setting of homogeneous slopes. The same appears to hold true for BC-IPC when $\pi_u = 1/4$, although there are substantial distortions when $\pi_u = 3/4$. The t-test associated with BC-PC is oversized when $\pi_u = 3/4$ and the distortion becomes more severe with $\pi_u = 1/4$. CA and MGCA have the largest size distortions. In the case of heterogeneous slopes, MGIV performs well and size is close to 5%. MGPC and MGCA have substantial size distortions regardless of the value of π_u.

Table 2 presents results for the case where $(N, T) = (200, 25)$ (i.e. N is large relative

[6] The results for β_2 are qualitatively similar and so we do not report them to save space. These results are available upon request.

to T) and $(N, T) = (25, 200)$ (N is small relative to T) for $\pi_u = 3/4$. In the former case, 2SIV performs best in terms of bias. IPC has the smallest RMSE, followed by 2SIV. CA has the largest bias and RMSE. In the case of heterogeneous slopes, MGIV has smaller absolute bias than MGPC and MGCA. Therefore, MGIV is superior among mean-group type estimators, which are the only consistent estimators in this design. In the case where T is large relative to N, 2SIV and MGIV again outperform BC-IPC, BC-PC and CA in terms of bias, standard deviation and RMSE. As for the properties of the t-test, 2SIV and MGIV have the smallest size distortions relative to the other estimators, and inference based on 2SIV and MGIV remains credible even for small values of N or T. Moreover, 2SIV and MGIV exhibit good power properties, whereas MGPC has the lowest power when N is small relative to T.

7. ILLUSTRATION

In this section we apply our methodology to study the effect of climate shocks on economic growth using an unbalanced panel of 125 countries over the period 1961-2003. The data set is taken from Dell et al (2012).

In line with existing literature (e.g. Dell et al 2014), we consider the following benchmark static panel data model:

$$g_{it} = \beta_1 temp_{it} + \gamma_1 prec_{it} + \beta_2 D_i temp_{it} + \gamma_2 D_i prec_{it} + \eta_i + \tau_t + u_{it}, \tag{7.25}$$

where g_{it} denotes the growth rate of per-capita output for country i at year t, while $temp_{it}$ and $prec_{it}$ denote the level of temperature (in degrees Celcius) and precipitation (in units of 100 mm) for country i at year t, $i = 1, ..., 126$, $t = 1, ..., T_i$, where $\min\{T_i\} = 21$, $\max\{T_i\} = 43$ and $\overline{T} = N^{-1}\sum_{i=1}^{N} T_i \approx 40$. D_i denotes a binary variable that equals one if the country i is characterised as "developing" and zero if it is characterised as "developed". Thus, β_1 and γ_1 reflect the effect of temperature and precipitation, respectively, on economic growth rate for developed economies, whereas β_2 and γ_2 capture the corresponding differential effects between developing and developed economies. The main reason for such a distinction is that developing economies are often reliant on agriculture or outdoor activities, and therefore they are vulnerable to climate shocks. Following Dell et al (2012), in the present application a country is defined as developing if it has below-median PPP-adjusted per capita GDP in the first year the country enters the dataset, otherwise it is defined as developed.

As it is common practice in the literature (e.g. Colacito et al, 2018), we include country effects effects, η_i and year effects, τ_t. In addition to these additive effects, we also allow for unobserved interactive effects. This offers wider scope for controlling for omitted variables, including situations where there is cross-sectional dependence. In particular, u_{it} is given by

$$u_{it} = \boldsymbol{\varphi}_i^{0\prime}\mathbf{h}_t^0 + \varepsilon_{it},$$

where $\mathbf{h}_t^0$ is a $r_2 \times 1$ vector of year-specific unobserved common shocks with corresponding country-specific loadings given by $\boldsymbol{\varphi}_i^0$, whereas ε_{it} is a purely idiosyncratic error.

We employ four estimators; namely, the two-stage IV (2SIV) estimator analysed in this paper, a fixed effects (FE) estimator that allows for two-way clustering per country and region-year[7], as in Cameron et al (2011), the pooled common correlated effects (CA)

[7]See footnote 12 in Del et al. (2012; p.74) for the definition of geographical regions.

estimator of Pesaran (2006), and the iterative principal components (IPC) estimator of Bai (2009).

The FE estimator imposes $u_{it} = \varepsilon_{it}$ by construction, i.e. it assumes there exists no factor structure. However, in order to try and neutralise the effect of common shocks, we follow Dell et al (2012) and include year fixed effects interacted with region dummies, as well as year fixed effects interacted with the developing country dummy.[8] For 2SIV, the number of factors in regressors and u_{it}, r_1 and r_2 respectively, is estimated using the eigenvalue ratio test of Ahn and Horenstein (2013). In order to carry out a specification test of our model (the well-known overidentifying restrictions J-test), we make use of present and lagged values of all defactored regressors as instruments. Thus, the total number of instruments equals 8. Since temperature and precipitation are measured in rather different units, we defactor these variables separately. The CA estimator is implemented using year-specific cross-sectional averages of all regressors. For the IPC estimator of Bai (2009), the number of factors in u_{it} is estimated using the Bai and Ng (2002) model information criteria.

The results are presented in Table 3. For all estimators, we run two different models. Column (A) corresponds to a specification that imposes the restriction $\beta_2 = \gamma_2 = 0$; that is, developing and developed economies are pooled together. The estimates of the coefficients of temperature and precipitation are negative across all four estimators. However, the temperature effect is statistically significant only for 2SIV and CA, both at the 10% level. Moreover, the precipitation coefficient is statistically significant only for 2SIV (at the 5% level). The J-test statistic of 2SIV rejects the specification of the model, which implies that developing and developed economies may be affected in a different manner.

Column (B) corresponds to the specification in (7.25), which relaxes the pooling restriction. In this case, FE replicates the main panel results reported in Table 3, Column (3), by Dell et al (2012). As we can see, the effect of temperature on growth appears to be positive for developed economies and highly negative for developing ones, indicating substantial heterogeneity between the two groups. However, the estimate of β_1 is statistically significant only for 2SIV and IPC but not FE or CA. Thus, the implications are substantially different. In particular, the results obtained by 2SIV indicate that a 1^oC rise in temperature appears to increase (decrease) growth rates for developed (developing) economies by .530 (1.42) percentage points (hereafter, p.p.), all other things being equal.[9] The specification of the model is not rejected by the J-test statistic. The estimated effect of temperature obtained by IPC is somewhat smaller in absolute magnitude than that of 2SIV, both in terms of developing and developed economies. Last, for FE (CA) the estimated coefficients indicate that a 1^oC rise in temperature decreases the growth rate for developing economies by 1.61 (1.76) p.p., whereas it does not exert a statistically significant impact on developed economies.

In regards to precipitation, the results obtained by 2SIV indicate that an extra 100 mm of annual rainfall is expected to decrease growth rates for both developed and developing economies by approximately .08 p.p. all other things being equal. On the other hand, the estimated effect of precipitation obtained from IPC and CA is not statistically significant for either group of economies. Finally, for FE the estimated precipitation effect for developed economies is very close to that obtained by 2SIV. However, the estimated precipitation effect is significantly positive for developing economies.

[8] Thus, τ_t cannot be separately identified per se.

[9] The latter estimate is obtained by adding $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

8. CONCLUSIONS

We have analysed the IV estimation approach put forward by Norkute et al. (2021), in the context of a static, linear panel data model with interactive effects in the error term and regressors. For models with homogeneous slopes, we studied a two-stage IV estimator (2SIV), and established $\sqrt{NT}$-consistency and asymptotic normality, under weak cross-section and time-series dependence in the idiosyncratic errors. For models with heterogeneous slopes, we studied a mean-group IV estimator (MGIV) and established $\sqrt{N}$-consistency and asymptotic normality.

We have compared and contrasted the asymptotic expressions of our 2SIV estimator, IPC of Bai (2009), PC and CA of Westerlund and Urbain (2015) and Pesaran (2006), for models with homogeneous slopes. Under conditions similar to those in Bai (2009), we showed that 2SIV is free from asymptotic bias, whereas the remaining estimators suffer from asymptotic bias. In addition, it is revealed that 2SIV is asymptotically as efficient as the bias-corrected versions of PC and CA, while the relative efficiency of the bias-corrected IPC estimator is generally indeterminate. The theoretical results are corroborated in a Monte Carlo simulation exercise, which shows that 2SIV and MGIV perform competently and can outperform existing estimators.

ACKNOWLEDGEMENTS

Sarafidis thanks the Australian Research Council (Grant Number DP-170103135) and Yamagata thanks the Japan Society for the Promotion of Science (KAKENHI JP18K01545) for financial support. We are grateful to Kazuhiko Hayakawa, Joakim Westerlund, and the participants in various seminars and conferences, in which we presented, for helpful comments. The views expressed in this paper are those of the authors and do not necessarily represent the official views of the Bank of Lithuania or the Eurosystem.

Table 1: Bias, root mean squared error (RMSE) of the estimators of β_1, and size and power of the associated t-tests when $\pi_u = \{1/4, 3/4\}$ and $N = T = 200$.

Estimator	Homogeneous Slopes					Heterogenous Slopes				
	Bias (×100)	S.D. (×100)	RMSE (×100)	Size	Power	Bias (×100)	S.D. (×100)	RMSE (×100)	Size	Power
					$\pi_u = 3/4$					
2SIV	0.003	0.586	0.586	5.5	100.0	0.583	0.960	1.122	7.9	100.0
BC-IPC	−0.149	0.805	0.818	21.9	100.0	0.238	1.246	1.268	10.0	100.0
IPC	0.020	0.528	0.528	6.1	100.0	0.408	1.061	1.137	6.4	100.0
BC-PC	0.306	0.885	0.937	19.7	100.0	0.891	1.181	1.479	17.9	100.0
PC	−0.638	0.589	0.868	21.2	100.0	−0.081	0.969	0.973	4.5	100.0
CA	1.859	0.806	2.026	80.1	100.0	2.469	1.131	2.716	64.3	100.0
MGIV	0.000	0.593	0.592	5.1	100.0	0.014	0.958	0.958	4.2	100.0
MGPC	−0.650	0.595	0.882	21.5	100.0	−0.636	0.963	1.154	8.7	100.0
MGCA	1.623	0.722	1.776	72.4	100.0	1.693	1.064	1.999	38.3	100.0
					$\pi_u = 1/4$					
2SIV	−0.002	0.573	0.572	6.0	100.0	0.559	0.992	1.138	9.0	100.0
BC-IPC	−0.073	0.438	0.444	6.1	100.0	0.100	1.645	1.648	8.7	100.0
IPC	−0.073	0.437	0.443	6.3	100.0	0.107	1.645	1.648	8.8	100.0
BC-PC	2.786	2.520	3.756	72.4	100.0	3.446	2.785	4.430	65.8	100.0
PC	−0.638	0.576	0.859	20.2	100.0	−0.097	0.993	0.998	4.7	100.0
CA	2.083	0.920	2.278	84.4	100.0	2.645	1.229	2.916	69.0	100.0
MGIV	−0.002	0.582	0.582	5.4	100.0	−0.008	0.980	0.979	4.5	100.0
MGPC	−0.646	0.586	0.872	20.3	100.0	−0.649	0.983	1.177	9.5	100.0
MGCA	1.789	0.788	1.955	76.5	100.0	1.827	1.111	2.138	42.4	100.0

Table 2: Bias, root mean squared error (RMSE) of the estimators of β_1, and size and power of the associated t-tests when $\pi_u = 3/4$, $N = 200$, $T = 25$ and $N = 25$, $T = 200$.

Estimator	Homogeneous Slopes					Heterogeneous Slopes				
	Bias ($\times100$)	S.D. ($\times100$)	RMSE ($\times100$)	Size	Power	Bias ($\times100$)	S.D. ($\times100$)	RMSE ($\times100$)	Size	Power
					$N = 200$, $T = 25$					
2SIV	0.126	1.941	1.944	6.7	99.8	1.519	2.156	2.637	12.4	100.0
BC-IPC	−1.180	2.610	2.864	23.6	97.6	−0.070	2.911	2.911	17.1	98.1
IPC	0.374	1.870	1.906	8.7	99.9	1.301	2.234	2.585	12.9	100.0
BC-PC	0.825	2.746	2.867	12.7	99.8	2.185	2.842	3.584	20.9	100.0
PC	−0.211	2.756	2.763	11.6	99.7	1.145	2.842	3.063	12.6	99.8
CA	2.084	2.000	2.888	21.4	100.0	3.404	2.218	4.062	37.8	100.0
MGIV	0.482	2.534	2.578	9.9	99.4	0.606	2.687	2.754	10.8	99.6
MGPC	−0.414	2.554	2.587	9.0	99.0	−0.279	2.737	2.751	9.9	98.0
MGCA	1.850	2.127	2.819	15.9	100.0	1.914	2.334	3.018	14.8	100.0
					$N = 25$, $T = 200$					
2SIV	0.016	1.715	1.715	9.2	99.9	0.480	2.736	2.777	8.7	97.7
BC-IPC	−2.552	9.303	9.644	65.0	79.1	−2.679	10.032	10.381	51.4	69.5
IPC	0.639	2.883	2.953	14.8	98.2	0.939	3.885	3.996	13.2	91.1
BC-PC	2.547	5.525	6.083	29.5	95.7	2.910	6.102	6.759	24.5	87.7
PC	−5.703	2.103	6.078	82.5	57.8	−5.413	3.011	6.194	42.6	33.2
CA	5.971	3.267	6.805	64.3	100.0	6.277	4.086	7.489	39.9	99.7
MGIV	0.038	1.742	1.742	6.6	99.9	0.036	2.725	2.725	5.6	94.7
MGPC	−6.047	2.179	6.427	83.6	48.3	−5.997	3.018	6.713	48.3	26.5
MGCA	4.705	2.610	5.380	54.6	100.0	4.689	3.416	5.801	32.0	99.5

Table 3: Climate shocks and economic growth

	2SIV		FE		CA		IPC	
	(A)	(B)	(A)	(B)	(A)	(B)	(A)	(B)
$\widehat{\beta}_1$	-.427*	.530*	-.328	.262	-.412*	.194	-.093	.392*
	(.241)	(.315)	(.285)	(.311)	(.230)	(.232)	(.176)	(.202)
$\widehat{\beta}_2$	–	-1.946***	–	-1.610***	–	-1.764***	–	-1.368***
		(.534)		(.485)		(.528)		(.306)
$\widehat{\gamma}_1$	-.089**	-.079*	-.008	-.083*	-.009	-.083	-.012	-.052
	(.041)	(.046)	(.044)	(.050)	(.052)	(.083)	(.034)	(.047)
$\widehat{\gamma}_2$	–	-.016	–	.153*	–	.068	–	.064
		(.088)		(.078)		(.102)		(.069)
$\widehat{\beta}_1 + \widehat{\gamma}_1$		-1.417***		-1.348***		-1.570***		-.976***
		(.429)		(.408)		(.474)		(.272)
$\widehat{\beta}_2 + \widehat{\gamma}_2$	–	-.033	–	.070*	–	-.015	–	.012
		(.056)		(.042)		(.060)		(.046)
$\widehat{r}_2$	1	3	–	–	–	–	2	2
$\widehat{r}_1$	3	3	–	–	–	–	–	–
J	13.21	4.42	–	–	–	–	–	–
	[.040]	[.352]						

Notes: Standard errors in parentheses and p-values in square brackets. * p<0.10, ** p<0.05, *** p<0.01. In column (B), the effect on developing economies is obtained as $\widehat{\beta}_1 + \widehat{\gamma}_1$ (or $\widehat{\beta}_2 + \widehat{\gamma}_2$).

REFERENCES

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

Bai, J., (2003). Inferential theory for factor models of large dimensions. *Econometrica, 71*, 135–73.

Bai, J., (2009a). Panel data models with interactive fixed effects. *Econometrica, 77*, 1229–1279.
Bai, J., (2009b). Supplement to panel data models with interactive fixed effects: technical details and proofs. *Econometrica Supplemental Material,*.
Bai, J., Li, K., (2014). Theory and methods of panel data models with interactive effects. *Annals of Statistics, 42*, 142–170.
Bai, J., Ng, S., (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica, 70*, 191–221.
Bun, M., Kiviet, J., (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics, 132*, 409–444.
Cameron, A. C., Gelbach, J. B., Miller, D. L., (2011). Robust Inference with Multiway Clustering. *Journal of Business and Economic Statistics, 29*, 238–49.
Colacito, R., Hoffmann, B., Phan, T., (2018). Temperature and Growth: A Panel Analysis of the United States. *Federal Reserve Bank of Richmond Working Paper 18/09*.
Cui, G., Hayakawa, K., Nagata, S., Yamagata, Y., (2019). A robust approach to heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models with interactive effects. *mimeo,*.
Dell, M., Jones, B. F., Olken, B. A., (2012). Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics, 4*, 66-95.
Dell, M., Jones, B. F., Olken, B. A., (2014). What Do We Learn From the Weather? The New Climate Economy Literature. *Journal of Economic Literature, 52*, 740-798.
Hansen, C. B., (2007). Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T is Large. *Journal of Econometrics, 141*, 597–620.
Jiang, B., Yang, Y., Gao, J., and Hsiao, C., (2017). Recursive estimation in large panel data models: theory and practice. *Monash University, Working Paper,* No. 05/17.
Juodis, A., Karabiyik, H., and Westerlund, J. (2021). On the robustness of the pooled CCE estimator. *Journal of Econometrics, 220*, 325–348.
Juodis, A., Sarafidis, V., (2018). Fixed T Dynamic Panel Data Estimators with Multi-Factor Errors. *Econometric Reviews, 37*, 893–929.
Juodis, A., Sarafidis, V., (2020). A Linear Estimator for Factor-Augmented Fixed-T Panels With Endogenous Regressors. *Journal of Business & Economic Statistics,* 1–15.
Juodis, A., Sarafidis, V., (2021). An Incidental Parameters Free Inference Approach for Panels with Common Shocks. *Journal of Econometrics,* forthcoming.
Kapetanios, G., (2010). A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets. *Journal of Business & Economic Statistics, 28*, 397–409.
Kapetanios, G., Pesaran, H. M., (2005). Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns. *CESIFO Working Paper,* No. 1416.
Kripfganz, S., Sarafidis, V., (2021). Instrumental Variable Estimation of Large Panel Data Models with Common Factors. *Stata Journal,* forthcoming. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3668588`.
Moon, H. R., Weidner, M., (2015). Linear regression for panel with unknown number of factors as interactive effects. *Econometrica, 83*, 1543–1579.
Sarafidis, V., Weber, N., (2015). A Partially Heterogeneous Framework for Analyzing Panel Data. *Oxford Bulletin of Economics and Statistics, 77*, 274-296.
Moon, H. R., Weidner, M., (2017). Dynamic linear panel regression models with interactive effects. *Econometric Theory, 33*, 158–195.
Norkute, M., Sarafidis, V., Yamagata, T., and Cui, G., (2021). Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors and a Multifactor Error Structure. *Journal of Econometrics, 33*, 1–31.
Pesaran, M. H., (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica, 74*, 967–1012.
Reese, S., and Westerlund, J., (2018). Estimation of factor-augmented panel regressions with weakly influential factors. *Econometric Reviews, 37*, 401–465.
Sarafidis, V., Wansbeek, T., (2020). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics,* 1–12.
Westerlund, J., Urbain, J. P., (2015). Cross-sectional averages versus principal components. *Journal of Econometrics, 185*, 372–377.