UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Fixed T dynamic panel data estimators with multifactor errors

Juodis, A.; Sarafidis, V.

DOI
10.1080/00927872.2016.1178875
Publication date
2018
Document Version
Final published version
Published in
Econometric Reviews
License
CC BY-NC-ND

Link to publication

Citation for published version (APA):
Juodis, A., & Sarafidis, V. (2018). Fixed T dynamic panel data estimators with multifactor errors. *Econometric Reviews*, *37*(8), 893-929.
https://doi.org/10.1080/00927872.2016.1178875

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:10 Mar 2023

Econometric Reviews

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

Fixed T dynamic panel data estimators with multifactor errors

Artūras Juodis & Vasilis Sarafidis

To cite this article: Artūras Juodis & Vasilis Sarafidis (2018) Fixed T dynamic panel data estimators with multifactor errors, Econometric Reviews, 37:8, 893-929, DOI: 10.1080/00927872.2016.1178875

To link to this article: https://doi.org/10.1080/00927872.2016.1178875

Published with license by Taylor & Francis Group, LLC© Artūras Juodis and Vasilis Sarafidis

View supplementary material

Published online: 11 Jul 2016.

Submit your article to this journal

Article views: 1750

View related articles

View Crossmark data

Citing articles: 14 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lecr20

ECONOMETRIC REVIEWS
2018, VOL. 37, NO. 8, 893–929
http://dx.doi.org/10.1080/00927872.2016.1178875

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Fixed T dynamic panel data estimators with multifactor errors

Artūras Juodis[a] and Vasilis Sarafidis[b]

[a]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands; [b]Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria, Australia

ABSTRACT

This article analyzes a growing group of fixed *T* dynamic panel data estimators with a multifactor error structure. We use a unified notational approach to describe these estimators and discuss their properties in terms of deviations from an underlying set of basic assumptions. Furthermore, we consider the extendability of these estimators to practical situations that may frequently arise, such as their ability to accommodate unbalanced panels and common observed factors. Using a large-scale simulation exercise, we consider scenarios that remain largely unexplored in the literature, albeit being of great empirical relevance. In particular, we examine (i) the effect of the presence of weakly exogenous covariates, (ii) the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, (iii) the impact of the number of moment conditions on bias and size for GMM estimators, and finally (iv) the effect of sample size. We apply each of these estimators to a crime application using a panel data set of local government authorities in New South Wales, Australia; we find that the results bear substantially different policy implications relative to those potentially derived from standard dynamic panel GMM estimators. Thus, our study may serve as a useful guide to practitioners who wish to allow for multiplicative sources of unobserved heterogeneity in their model.

KEYWORDS
Dynamic panel data; factor model; fixed T consistency; maximum likelihood; Monte Carlo simulation

JEL CLASSIFICATION
C13; C15; C23

1. Introduction

There is a large literature on estimating dynamic panel data models with a two-way error components structure and *T* fixed. Such models have been used in a wide range of economic and financial applications; e.g., *Euler* equations for household consumption, adjustment cost models for firms' factor demand, and empirical models of economic growth. In all these cases, the autoregressive parameter has structural significance and measures state dependence, which is due to the effect of habit formation, technological/regulatory constraints, or imperfect information and uncertainty that often underlie economic behavior and decision making in general.

Recently there has been a surge of interest in developing dynamic panel data estimators that allow for richer error structures—mainly factor residuals. In this case, standard dynamic panel data estimators fail to provide consistent estimates of the parameters; see, e.g., Sarafidis and Robertson (2009), and Sarafidis and Wansbeek (2012) for a recent overview. The multifactor approach is appealing because it allows for multiple sources of multiplicative unobserved heterogeneity, as opposed to the two-way error components structure that represents additive heterogeneity. For example, in an empirical growth model the factor component may reflect country-specific differences in the rate at which countries absorb

CONTACT Artūras Juodis arturas.juodis@economists.lt and a.juodis@rug.nl Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE Groningen, The Netherlands.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lecr.
Supplemental data for the article can be accessed on the publisher's website.

Published with license by Taylor & Francis Group, LLC © Artūras Juodis and Vasilis Sarafidis
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

time-varying technological advances that are potentially available to all of them. In a partial adjustment model of factor input prices, the factor component may capture common shocks that hit all producers, albeit with different intensities. In this study, we provide a review of inference methods for dynamic panel data models with a multifactor error structure.

The majority of estimators developed in this literature is based on the Generalized Method of Moments (GMM) approach. This is presumably because in microeconometric panels endogeneity of the regressors is often an issue of major importance. In particular, Ahn et al. (2013) extend Ahn et al. (2001) to the case of multiple factors, and propose a GMM estimator that relies on quasi-long-differencing to eliminate the common factor component. Nauges and Thomas (2003) utilize the quasi-differencing approach of Holtz-Eakin et al. (1988), which is computationally tractable for the single factor case, and propose similar moment conditions to Ahn et al. (2001) *mutatis mutandis*. Sarafidis et al. (2009) propose using the popular linear first-differenced and System GMM estimators with instruments based solely on strictly exogenous regressors. Robertson and Sarafidis (2015) develop a GMM approach that introduces new parameters representing the unobserved covariances between the factor component of the error and the instruments. Furthermore, they show that given the model's structure there exist restrictions in the nuisance parameters that lead to a more efficient GMM estimator compared to quasi-differencing approaches. Hayakawa (2012) shows that the moment conditions proposed by Ahn et al. (2013) can be linearized at the expense of introducing extra parameters. Finally, Bai (2013b) and Hayakawa (2012) suggest estimators that approximate the factor loadings using a Chamberlain (1982) type projection approach, with a Quasi Maximum Likelihood estimator suggested in the former article and a GMM estimator in the latter one.

The objective of our study is to serve as a useful guide for practitioners who wish to apply methods that allow for multiplicative sources of unobserved heterogeneity in their model. All methods are analyzed using a unified notational approach, to the extent that this is possible of course, and their properties are discussed under deviations from a baseline set of assumptions commonly employed. We pay particular attention to calculating the number of identifiable parameters correctly, which is a requirement for asymptotically valid inferences and consistent model selection procedures. This issue is often overlooked in the literature. Furthermore, we consider the extendability of these estimators to practical situations that may frequently arise, such as their ability to accommodate unbalanced panels, and to estimate models with common observed factors.

Next, we investigate the finite sample performance of the estimators under a number of different designs. In particular, we examine (i) the effect of the presence of weakly exogenous covariates, (ii) the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, (iii) the impact of the number of moment conditions on bias and size for GMM estimators, (iv) the impact of different levels of persistence in the data, and finally (v) the effect of sample size. These are important considerations with high empirical relevance. Notwithstanding, to the best of our knowledge they remain largely unexplored. For example, the simulation study in Robertson and Sarafidis (2015) does not consider the effect of using a different number of instruments on the finite sample properties of their estimator. In Ahn et al. (2013) the design focuses on strictly exogenous regressors (i.e., no dynamics), while in Bai (2013b) the results reported do not include inference. The practical issue of how to choose initial values for the nonlinear algorithms is considered in the Appendix. The results of our simulation study indicate that there are non-negligible differences in the finite sample performance of the estimators, depending on the parametrization considered. Naturally, no estimator dominates the remaining ones universally, although it is fair to say that some estimators are more robust than others.

We apply the aforementioned methodologies to estimate the income elasticity of crime using a panel data set of 153 local government areas in New South Wales (NSW), each one being observed over a period that spans 2006–2012. We note that this is one of the first articles to apply these estimators to a real data set for models with a lagged dependent variable. We find that the results bear substantially different policy implications relative to those potentially derived based on standard dynamic panel GMM estimators, which are widely used and are available in most econometric software packages nowadays.

In particular, the estimated short-run income elasticity of crime obtained from the first-differenced GMM estimator proposed by Arellano and Bond (1991) is roughly twice as large in absolute terms than most GMM estimators that account for a multifactor error structure. In addition, the estimated dynamics of the crime rate process are substantially different across these estimators, with about three periods required on average for 90% of the long run effect to be realized based on first-differenced GMM, and approximately seven periods for other GMM estimators.

The outline of the rest of the article is as follows. The next section introduces the dynamic panel data model with a multifactor error structure and discusses some underlying assumptions that are commonly employed in the literature. Section 3 presents a large range of dynamic panel estimators developed for such model when T is small, and discusses several technical points regarding their properties. Section 4 provides some general remarks on the estimators. Section 5 investigates the finite sample performance of the estimators, and Section 6 applies them to crime dataset from the state of NSW in Australia. A final section concludes. The Appendix analyzes in detail the implementation of all these methods.

In what follows, we briefly introduce our notation. The usual $\text{vec}(\cdot)$ operator denotes the column stacking operator, while $\text{vech}(\cdot)$ is the corresponding operator that stacks only the elements on and below the main diagonal. The elimination matrix $\boldsymbol{B}_a$ is defined such that for any $[a \times a]$ matrix (not necessarily symmetric) $\text{vech}(\cdot) = \boldsymbol{B}_a\text{vec}(\cdot)$. The lag-operator matrix $\boldsymbol{L}_T$ is defined such that for any $[T \times 1]$ vector $\boldsymbol{x} = (x_1, \ldots, x_T)'$, $\boldsymbol{L}_T\boldsymbol{x} = (0, x_1, \ldots, x_{T-1})'$. Shorthand notation $\boldsymbol{x}_{i,s:k}$, $s \leq k$ is used to denote the vectors of the form $\boldsymbol{x}_{i,s:k} = (x_{i,s}, \ldots, x_{i,k})'$. The jth column of the $[x \times x]$ identity matrix is denoted by $\boldsymbol{e}_j$. Finally, $1_{(\cdot)}$ is the usual indicator function. For further details regarding the notation used in this article, see Abadir and Magnus (2002).

2. Theoretical setup

We consider the following dynamic panel data model with a multifactor error structure

$$y_{i,t} = \alpha y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{i,t}^{(k)} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}; \quad i = 1, \ldots, N, t = 1, \ldots, T, \tag{2.1}$$

where the dimension of the unobserved components $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ is $[L \times 1]$.[1] Stacking the observations over time for each individual i yields

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and similarly for $(\boldsymbol{y}_{i,-1}, \boldsymbol{x}_i^{(k)})$, while $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ is of dimension $[T \times L]$.

In what follows, we list some assumptions that are commonly employed in the literature, followed by some preliminary discussion. In Section 3, we provide further discussion with regards to which of these assumptions can be strengthened/relaxed for each estimator analyzed.

Assumption 1. $x_{i,t}^{(k)}$ has finite moments up to fourth order for all k.

Assumption 2. $\varepsilon_{i,t} \sim i.i.d. \left(0, \sigma_\varepsilon^2\right)$ and has finite moments up to fourth order.

Assumption 3. $\boldsymbol{\lambda}_i \sim i.i.d. (\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\lambda}})$ with finite moments up to fourth order, where $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}}$ is positive definite. $\boldsymbol{F}$ is non-stochastic and bounded such that $\|\boldsymbol{F}\| < b < \infty$.

[1]The factor structure is often employed in order to provide a tractable way to model "strong" cross-sectional dependence. When some meaningful concept of "economic distance" is available, the spatial approach is a viable alternative for modelling "weak" cross-sectional dependence. There are strong connections between the two approaches, although it is beyond the scope of this article to analyze these further. The interested reader may refer to Chudik et al. (2011) and Sarafidis and Wansbeek (2012), among others. A recent contribution in the literature of dynamic panel data models with spatial dependence is provided by Sarafidis (2015).

Assumption 4. $\mathrm{E}\left(\varepsilon_{i,t}|\boldsymbol{y}'_{i,0:t-1},\boldsymbol{\lambda}'_i,\boldsymbol{x}^{(k)'}_{i,1:\tau}\right)=0$ for all t and k, where τ is a positive integer that is bounded by T.

Assumption 1 is a standard regularity condition. Assumptions 2 and 3 are employed mainly for simplicity and can be relaxed to some extent, details of which will be documented later.[2]

Assumption 4 can be crucial for identification, depending on the estimation approach, because it characterizes the exogeneity properties of the covariates. In particular, we will refer to covariates that satisfy $\tau = T$ as strictly exogenous with respect to the idiosyncratic error component, whereas covariates that satisfy only $\tau = t$ are weakly exogenous. When $\tau < t$, the covariates are endogenous. The exogeneity properties of the covariates play a major role in the analysis of likelihood-based estimators because the presence of weakly exogenous or endogenous regressors may lead to inconsistent estimates of the structural parameters, α and β_k.

Furthermore, Assumption 4 implies that the idiosyncratic errors are conditionally serially uncorrelated. This can be relaxed in a relatively straightforward way, particularly for GMM estimators; for example, an Moving Average (MA) process of order q can be accommodated by truncating the set of instruments with respect to y based on $\mathrm{E}\left(\varepsilon_{i,t}|\boldsymbol{y}'_{i,0:s},\boldsymbol{\lambda}'_i,\boldsymbol{x}^{(k)'}_{i,1:\tau}\right)=0$, where $s < t-q$. Furthermore, an Autoregressive (AR) structure can be accommodated either by using moment conditions with respect to (lagged values of) $x^{(k)}_{i,\tau}$ only, or based on a Cochrane–Orcutt type procedure.

Assumption 4 also implies that the idiosyncratic error is conditionally uncorrelated with the factor loadings. This is required for identification based on internal instruments in levels. Finally, notice that the set of our assumptions implies that $y_{i,t}$ has finite fourth-order moments, but it does not imply conditional homoskedasticity for the two error components.

Under Assumptions 1–4, the following set of population moment conditions is valid by construction

$$\mathrm{E}[\mathrm{vech}(\boldsymbol{\varepsilon}_i\boldsymbol{y}'_{i,-1})]=\boldsymbol{0}_{T(T+1)/2}. \tag{2.2}$$

In addition, the following sets of moment conditions are valid, depending on whether $\tau = T$ or $\tau = t$ holds true, respectively:

$$\mathrm{E}[\mathrm{vec}(\boldsymbol{\varepsilon}_i\boldsymbol{x}^{(k)'}_i)]=\boldsymbol{0}_{T^2}, \tag{2.3}$$

$$\mathrm{E}[\mathrm{vech}(\boldsymbol{\varepsilon}_i\boldsymbol{x}^{(k)'}_i)]=\boldsymbol{0}_{T(T+1)/2}. \tag{2.4}$$

For all GMM estimators one can easily modify the above moment conditions to allow for endogenous x's. For example, for (say) $\tau = t-1$ in Assumption 4 one may redefine $\boldsymbol{x}^{(k)}_i \equiv (x_{i,0},\ldots,x_{i,T-1})'$ and proceed in exactly the same way as in $\tau = t$.

From now on, we will use the triangular structure of the moment conditions induced by the vech$(\cdot)$ operator to construct the estimating equations for the GMM estimators. To achieve this, we adopt the following matrix notation for the stacked model:

$$\boldsymbol{Y}=\alpha\boldsymbol{Y}_{-1}+\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k+\boldsymbol{\Lambda}\boldsymbol{F}'+\boldsymbol{E};\quad i=1,\ldots,N,$$

where $(\boldsymbol{Y},\boldsymbol{Y}_{-1},\boldsymbol{X}_k,\boldsymbol{E})$ are $[N\times T]$ matrices with typical rows $(\boldsymbol{y}'_i,\boldsymbol{y}'_{i,-1},\boldsymbol{x}^{(k)'}_i,\boldsymbol{\varepsilon}'_i)$, respectively. Similarly, a typical row element of $\boldsymbol{\Lambda}$ is given by $\boldsymbol{\lambda}'_i$.

3. Estimators

Remark 3.1. For notational symmetry, while describing GMM estimators, we assume that $x^{(k)}_{i,0}$ observations are not included in the set of available instruments. Otherwise, additional T or $T-1$ (depending

[2]The zero-mean assumption for $\varepsilon_{i,t}$ is actually implied by Assumption 4.

on the estimator analyzed) moment conditions are available. The same strategy is used in the Monte Carlo section of this article.

3.1. Quasi-differenced (QD) GMM

Replacing the expectations in (2.2) and (2.3) with sample averages yields

$$\mathrm{vech}\left(\frac{1}{N}(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k-\boldsymbol{\Lambda}\boldsymbol{F}')'\boldsymbol{Y}_{-1}\right),$$

$$\mathrm{vech}\left(\frac{1}{N}(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k-\boldsymbol{\Lambda}\boldsymbol{F}')'\boldsymbol{X}_k\right).$$

These moment conditions depend on the unknown matrices $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$. In the simple fixed effects model where $\boldsymbol{F}=\boldsymbol{\iota}_T$, the first-differencing transformation proposed by Anderson and Hsiao (1982) is the most common approach to eliminate the nuisance parameters from the equation of interest. Using a similar idea in the model with a single unobserved time-varying factor, i.e.,

$$y_{i,t}=\alpha y_{i,t-1}+\sum_{k=1}^{K}\beta_k x_{i,t}^{(k)}+\lambda_i f_t+\varepsilon_{i,t},$$

Holtz-Eakin et al. (1988) suggest eliminating the unobserved factor component using the quasi-differencing (QD) transformation

$$y_{i,t}-r_t y_{i,t-1}=\alpha(y_{i,t-1}-r_t y_{i,t-2})+\sum_{k=1}^{K}\beta_k(x_{i,t}^{(k)}-r_t x_{i,t-1}^{(k)})+\varepsilon_{i,t}-r_t\varepsilon_{i,t-1};\quad i=1,\ldots,N,t=2,\ldots,T, \tag{3.1}$$

where $r_t\equiv f_t/f_{t-1}$. By construction, Eq. (3.1) is free from $\lambda_i f_t$ because

$$\lambda_i f_t-r_t\lambda_i f_{t-1}=\lambda_i f_t-\frac{f_t}{f_{t-1}}\lambda_i f_{t-1}=0,\quad \forall t=2,\ldots,T.$$

It is easy to see that the QD approach is well defined only if *all* $f_t\neq 0$. Collecting all parameters involved in QD, we can define the corresponding $[(T-1)\times T]$ QD transformation matrix by

$$\boldsymbol{D}(\boldsymbol{r})=\begin{pmatrix}-r_2 & 1 & 0 & \cdots & 0\\ 0 & -r_3 & & \vdots & 0\\ \vdots & \vdots & \vdots & 1 & \vdots\\ 0 & 0 & \ldots & -r_T & 1\end{pmatrix},$$

where $\boldsymbol{r}=(r_2,\ldots,r_T)'$. The first-differencing (FD) transformation matrix is a special case with $r_2=\ldots=r_T=1$. Premultiplying the terms inside the vech$(\cdot)$ operator in the sample analogue of the population moment conditions above by $\boldsymbol{D}(\boldsymbol{r})$, and noticing that $\boldsymbol{D}(\boldsymbol{r})\boldsymbol{F}=\boldsymbol{0}$, we can rewrite the estimating equations for the QD GMM estimator as

$$\boldsymbol{m}_\alpha=\mathrm{vech}\left(\frac{1}{N}\boldsymbol{D}(\boldsymbol{r})\left(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}(1)'\right),$$

$$\boldsymbol{m}_k=\mathrm{vech}\left(\frac{1}{N}\boldsymbol{D}(\boldsymbol{r})\left(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{X}_k\boldsymbol{J}(1)'\right)\quad \forall k.$$

Here $\boldsymbol{J}(L) = (\boldsymbol{I}_{T-L}, \mathbf{O}_{(T-L)\times L})$ is a selection matrix that appropriately truncates the set of instruments to ensure that the term inside the vech$(\cdot)$ operator is a square matrix. One can easily see that the total number of moment conditions and parameters under the weak exogeneity assumption for all x's is given by

$$\#moments = \frac{(K+1)(T-1)T}{2}; \quad \#parameters = (K+1) + (T-1).$$

The total number of parameters consists of two terms. The first term within the brackets corresponds to $K+1$ parameters of interest (or *structural/model* parameters), while the remaining term corresponds to $T-1$ nuisance parameters, the time-varying factors.

Remark 3.2. If we define $\tilde{r}_t \equiv f_{t-1}/f_t$, we can also consider a QD matrix of the following type:

$$\boldsymbol{D}(\tilde{\boldsymbol{r}}) = \begin{pmatrix} 1 & -\tilde{r}_2 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots & 0 \\ \vdots & \vdots & \vdots & -\tilde{r}_{T-1} & \vdots \\ 0 & 0 & \dots & 1 & -\tilde{r}_T \end{pmatrix}.$$

This transformation approach uses forward differences rather than backward differences. However, similarly to the original transformation matrix of Holtz-Eakin et al. (1988), the estimator based on this transformation requires that all $f_t \neq 0$ for $t = 2, \dots T$. Hence the restrictions imposed by two differencing strategies overlap for $t = 2, \dots, T-1$, but not for $t = 1$ and $t = T$. Finally, one could also consider transformation matrices based on higher order forward or backward differences.

The approach of Holtz-Eakin et al. (1988) as it stands is tailored for models with a single unobserved factor. In principle, it can be extended to multiple factors by removing each factor consecutively based on a $\boldsymbol{D}_{(l)}(\boldsymbol{r}^{(l)})$ matrix, with the final transformation matrix being a product of L such matrices. However, this approach soon becomes computationally very cumbersome as the estimating equations become multiplicative in $\boldsymbol{r}^{(l)}$.

On the other hand, if the model involves some observed factors, the corresponding $\boldsymbol{D}_{(\cdot)}(\cdot)$ matrix is known, leading to a simple estimator that involves equations containing structural parameters and $\boldsymbol{r}$ only. For example, Nauges and Thomas (2003) augment the model of Holtz-Eakin et al. (1988) by allowing for time-invariant individual effects

$$y_{i,t} = \eta_i + \alpha y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{i,t}^{(k)} + \lambda_i f_t + \varepsilon_{i,t}; \quad t = 1, \dots, T,$$

where η_i is eliminated a priori using the FD transformation matrix $\boldsymbol{D}(\boldsymbol{\imath}_{T-1})$, which yields

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \sum_{k=1}^{K} \beta_k \Delta x_{i,t}^{(k)} + \lambda_i \Delta f_t + \Delta\varepsilon_{i,t}; \quad t = 2, \dots, T,$$

followed by the QD transformation, albeit operated based on a $[(T-2) \times (T-1)]$ matrix $\boldsymbol{D}(\boldsymbol{r})$. The resulting number of parameters and moment conditions can be modified accordingly.

Remark 3.3. The FD transformation is by no means the only way to eliminate the fixed effects from the model. Another commonly discussed transformation is *Forward Orthogonal Deviations* (FOD). If one uses FOD instead of FD, the identification of structural parameters would require that all $\dot{f}_t \neq 0$.[3]

[3]Here, $\dot{f}_t \equiv c_t(f_t - (f_{t+1} + \dots + f_T)/(T-t))$ with $c_t^2 = (T-t)/(T-t+1)$.

Depending on the properties of f's, it might be desirable to use FOD even in the absence of η_i since r_t is defined for $f_t \neq 0$ only.

Remark 3.4. Assumption 2 can be easily relaxed. For example, unconditional time-series and cross-sectional heteroskedasticity of the idiosyncratic error component, $\varepsilon_{i,t}$, is allowed in the two-step version of the estimator. Serial correlation can be accommodated by choosing the set of instruments appropriately, as in the discussion provided in Section 2. This is a particularly attractive feature, which is common to all GMM estimators discussed in this article. Unconditional heteroskedasticity in $\boldsymbol{\lambda}_i$ can also be allowed, although this is a less interesting extension for practical purposes since there are no repeated observations over each $\boldsymbol{\lambda}_i$.

Finally, endogeneity of the regressors can be accommodated by selecting appropriate lags of the variables of the model as instruments. The exogeneity property of the covariates can be tested using an overidentifying restrictions test statistic. The same holds for all GMM estimators discussed in this article, which is of course a desirable property from the empirical point of view since the issue of endogeneity in panels with T fixed, e.g., microeconometric panels, may frequently arise.

3.2. Quasi-long-differenced (QLD) GMM

As we have mentioned before, the QD approach in Holtz-Eakin et al. (1988) is difficult to generalize to more than one unobserved factor (or more than one unobserved factor plus observed factors). Rather than eliminating factors using such transformation, Ahn et al. (2013) propose using a quasi-long-differencing (QLD) transformation. The factors can be removed from the model using the QLD transformation matrix $\boldsymbol{D}(\boldsymbol{F}^*)$

$$\boldsymbol{D}(\boldsymbol{F}^*) = (\boldsymbol{I}_{T-L}, \boldsymbol{F}^*) = \boldsymbol{J}(L) + \boldsymbol{F}^*\tilde{\boldsymbol{J}}(L),$$

where $\boldsymbol{F}^*$ is a $[T-L \times L]$ parameter matrix and $\tilde{\boldsymbol{J}}(L) = (\mathbf{O}_{L\times(T-L)}, \boldsymbol{I}_L)$, an $[L \times T]$ selection matrix. Rather than using the lagged observation $y_{i,t-1}$ to remove factors from the model at time t (one-by-one), the QLD approach uses long-differences based on the last observations $\boldsymbol{y}_{i,T-L+1:T}$ to remove all L factors at once.

To see this, partition $\boldsymbol{F} = (\boldsymbol{F}_A', -\boldsymbol{F}_B')'$ where $\boldsymbol{F}_A$ and $\boldsymbol{F}_B$ are of dimensions $[(T-L)\times L]$ and $[L \times L]$, respectively. Then assuming that $\boldsymbol{F}_B$ is invertible, one can redefine (or normalize) the factors and factor loadings as

$$\boldsymbol{F}\boldsymbol{\lambda}_i = \begin{pmatrix} \boldsymbol{F}^* \\ -\boldsymbol{I}_L \end{pmatrix} \boldsymbol{\lambda}_i^*; \quad \boldsymbol{F}^* \equiv \boldsymbol{F}_A\boldsymbol{F}_B^{-1}; \quad \boldsymbol{\lambda}_i^* \equiv \boldsymbol{F}_B\boldsymbol{\lambda}_i.$$

Using fairly straightforward matrix algebra, it then follows

$$\boldsymbol{D}(\boldsymbol{F}^*)\boldsymbol{F}\boldsymbol{\lambda}_i = (\boldsymbol{I}_{T-L}, \boldsymbol{F}^*) \begin{pmatrix} \boldsymbol{F}^* \\ -\boldsymbol{I}_L \end{pmatrix} \boldsymbol{\lambda}_i^* = \mathbf{0}_{T-L}.$$

One can express all available moment conditions for this estimator as

$$\boldsymbol{m}_\alpha = \text{vech}\left(\boldsymbol{D}(\boldsymbol{F}^*)\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)' \boldsymbol{Y}_{-1}\boldsymbol{J}(L)' \right),$$

$$\boldsymbol{m}_k = \text{vech}\left(\boldsymbol{D}(\boldsymbol{F}^*)\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)' \boldsymbol{X}_k\boldsymbol{J}(L)' \right) \quad \forall k.$$

Counting the number of moment conditions and resulting parameters, we have

$$\#moments = \frac{(K+1)(T-L)(T-L+1)}{2}; \quad \#parameters = K + 1 + (T-L)L.$$

However, we will further argue that the number of identifiable parameters is smaller than $K+1+(T-L)L$. To explain the reason for this, let $K=1$, and rewrite the transformed equation for $y_{i,1}$ as

$$y_{i,1}+\sum_{l=1}^{L} f_1^{*(l)} y_{i,T-l}=\alpha\left(y_{i,0}+\sum_{l=1}^{L} f_1^{*(l)} y_{i,T-l-1}\right)+\beta\left(x_{i,1}+\sum_{l=1}^{L} f_1^{*(l)} x_{i,T-l}\right)+\left(\varepsilon_{i,1}+\sum_{l=1}^{L} f_1^{*(l)} \varepsilon_{i,T-l}\right). \tag{3.2}$$

This equation has $2+L$ unknown parameters in total, while the number of moment conditions is 2 (constructed based on $y_{i,0}$ and $x_{i,1}$). Thus, L "nuisance parameters" are identified only up to a linear combination, unless $L \leq 2$ (or $L \leq K+1$ for the general model), which implies that the total number of identifiable parameters is

$$\#parameters = K+1+(T-L)L-1_{(L\geq K+1)}\frac{(L-K-1)(L-K)}{2}.$$

Notice that for $L=1$ the number of moment conditions and the number of identifiable parameters is exactly the same as in the QD transformation. Thus, one expects that the corresponding GMM estimators are asymptotically equivalent.[4]

Remark 3.4 regarding Assumptions 2–4, as discussed in Section 3.1, applies identically here as well. Ahn et al. (2013) show that under conditional homoskedasticity in $\varepsilon_{i,t}$ the estimation procedure simplifies considerably because it can be performed through iterations. Furthermore, for the case where the regressors are *strictly exogenous*, the resulting estimator is invariant to the chosen normalization scheme; see their Appendix A.

Remark 3.5. Note that for any $T-L$ dimensional invertible matrix $\boldsymbol{A}$, one can consider a rotated QLD transformation matrix $\boldsymbol{A}\boldsymbol{D}(\boldsymbol{F}^{*})$ (for which it obviously holds that $\boldsymbol{A}\boldsymbol{D}(\boldsymbol{F}^{*})\boldsymbol{F}=\mathbf{O}_{T-L}$). The same observation is also applicable to the estimation techniques in Section 3.1.

Remark 3.6. One can view the quasi long-differencing transformation matrix as the limiting case (in terms of the longest difference) of the forward differencing transformation matrix in Remark 3.2.

3.3. Factor IV

3.3.1. Unrestricted factor IV estimator (FIVU)

Rather than eliminating the incidental parameters $\boldsymbol{\lambda}_i$, Robertson and Sarafidis (2015) propose a GMM estimator that reduces these parameters onto a finite set of estimable coefficients. Their approach makes use of centered moment conditions of the form

$$\boldsymbol{m}_{\alpha}=\operatorname{vech}\left(\frac{1}{N}\left(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{Y}_{-1}-\boldsymbol{F}\boldsymbol{G}'\right),$$

$$\boldsymbol{m}_{k}=\operatorname{vech}\left(\frac{1}{N}\left(\boldsymbol{Y}-\alpha\boldsymbol{Y}_{-1}-\sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{X}_k-\boldsymbol{F}\boldsymbol{G}_k'\right) \quad \forall k,$$

[4] Although Eq. (3.2) does not appear to be in "differences" at first glance, identification of the factors is up to a column wise sign change. Thus, one could equivalently define

$$\boldsymbol{F}\boldsymbol{\lambda}_i=\begin{pmatrix}-\boldsymbol{F}^{*}\\ \boldsymbol{I}_L\end{pmatrix}(-\boldsymbol{\lambda}_i^{*});\quad \boldsymbol{D}(\boldsymbol{F}^{*})=(\boldsymbol{I}_{T-L},-\boldsymbol{F}^{*}),$$

and obtain an expression in differences.

where $(\boldsymbol{G}, \boldsymbol{G}_k)$ are defined as

$$\boldsymbol{G} = \mathrm{E}[\boldsymbol{y}_{i,-1}\boldsymbol{\lambda}_i']; \quad \boldsymbol{G}_k = \mathrm{E}[\boldsymbol{x}_i^{(k)}\boldsymbol{\lambda}_i'],$$

with typical row elements $\boldsymbol{g}_t'$ and $\boldsymbol{g}_t^{(k)'}$, respectively. The $(\boldsymbol{G}, \boldsymbol{G}_k)$ matrices represent the unobserved covariances between the instruments and the factor loadings in the error term. This approach adopts essentially a (correlated) random effects treatment of the factor loadings, which is natural because the asymptotics apply for N large and T fixed, and there are no repeated observations over each $\boldsymbol{\lambda}_i$. This is in the spirit of Chamberlain's projection approach. Different sensitivities to the factors (i.e., differences in the factor loadings) can be generated by different values of the variance of the cross-sectional distribution of $\boldsymbol{\lambda}_i$. Notice that as in Holtz-Eakin et al. (1988) and Ahn et al. (2013), factors corresponding to loadings that are uncorrelated with the regressors can be accommodated through the variance-covariance matrix of the idiosyncratic error component, $\varepsilon_{i,t}$, i.e., $\mathrm{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\right)$, since the latter can be left unrestricted.

For this estimator, the total number of moment conditions is given by

$$\#moments = \tfrac{(K+1)T(T+1)}{2}.$$

As the model stands right now, $\boldsymbol{G}_k$ (all $K+1$) and $\boldsymbol{F}$ are not separately identifiable because

$$\boldsymbol{F}\boldsymbol{G}' = \boldsymbol{F}\boldsymbol{U}\boldsymbol{U}^{-1}\boldsymbol{G}'$$

for any invertible $[L \times L]$ matrix $\boldsymbol{U}$. This rotational indeterminacy can be eliminated in the standard factor literature by imposing L^2 restrictions on an $[L \times L]$ submatrix of $\boldsymbol{F}$ (e.g., it could be restricted to the identity matrix).[5] These restrictions correspond to the L^2 term in the equation below. However, in the present case, $L > 1$ additional normalizations are required due to the fact that the moment conditions are of triangular vech$(\cdot)$ type. In particular, the number of identifiable parameters is

$$\#parameters = (K+1)(1+TL) + TL - L^2 - (K+1)\frac{L(L-1)}{2} - 1_{(L \geq K+1)}\frac{(L-K-1)(L-K)}{2}.$$

The $(K+1)L(L-1)/2$ term corresponds to the unobserved "last" $\boldsymbol{g}$, while the last term involving the indicator function corresponds to the unobserved "first" $\boldsymbol{f}$ and is identical to the right-hand side term in the corresponding expression for the number of identifiable parameters in the approach by Ahn et al. (2013).

Notwithstanding, as shown in Robertson and Sarafidis (2015) if one is only interested in the structural parameters, α and β_k, it is not essential to impose any identifying normalizations on $\boldsymbol{G}$ and $\boldsymbol{F}$; the resulting unrestricted estimator for structural parameters is consistent and asymptotically normal, while the variance-covariance matrix can be consistently estimated using the corresponding subblock of the generalized inverse of the unrestricted variance-covariance matrix.[6] Avoiding imposing normalization restrictions can be particularly attractive. For instance, in the case where all right-hand side variables are strictly exogenous, this means that all is required for identification of the structural parameters is that some $[L \times L]$ submatrix of $\boldsymbol{F}$ is invertible, but not necessarily the submatrix on the south east corner of $\boldsymbol{F}$, as it is the case with, e.g., QLD GMM.

Remark 3.7. Compared with the QLD estimator of Ahn et al. (2013) this estimator utilizes $L(K+1)(T-(L-1)/2)$ extra moment conditions, at the expense of estimating exactly the same number of additional parameters. Hence these estimators are asymptotically equivalent. Although in unrestricted factor IV estimator (FIVU) estimation one does not have to impose any restrictions on $\boldsymbol{F}$, for asymptotic identification in the weak exogeneity case the true value of $\boldsymbol{F}_B$ (as defined for QLD estimator) should

[5]Robertson and Sarafidis (2015) discuss which submatrix of $\boldsymbol{F}$ has to be be invertible in order for the estimator with weakly exogenous regressors to be consistent.
[6]For further details, see Theorem 3 in the corresponding article.

still satisfy the full rank condition. Notwithstanding, according to the simulation results that follow, it is worth noting that FIVU without normalizations appears to be more robust than QLD to this issue.

Finally, the FIVU estimator remains consistent even if the independent and identically distributed (i.i.d.) assumption on $\boldsymbol{\lambda}_i$ is replaced by independent and heteroskedastically distributed (i.h.d.). However, in that situation, a consistent estimation of the variance-covariance matrix is not possible. Ahn (2015) also discusses this issue. Note that all other estimators that do not difference away $\boldsymbol{\lambda}_i$ are also subject to this issue.

3.3.2. Restricted factor IV estimator (FIVR)

The autoregressive nature of the model suggests that individual rows of the $\boldsymbol{G}$ matrix have also an autoregressive structure, i.e.,

$$\boldsymbol{g}_t = \alpha \boldsymbol{g}_{t-1} + \sum_{k=1}^{k} \beta_k \boldsymbol{g}_t^{(k)} + \boldsymbol{\Sigma}_{\lambda} \boldsymbol{f}_t.$$

For identification, one may impose $L(L+1)/2$ restrictions so that without loss of generality $\boldsymbol{\Sigma}_{\lambda} = \boldsymbol{I}_L$. Thus, one can express $\boldsymbol{F}$ in terms of other parameters as follows:

$$\boldsymbol{F} = \left(\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T\right) \boldsymbol{G} + \boldsymbol{e}_T \boldsymbol{g}_T' - \sum_{k=1}^{k} \beta_k \boldsymbol{G}_k.$$

Here $\boldsymbol{L}_T$ is the usual lag matrix, while the additional parameter $\boldsymbol{g}_T$ is introduced to take into account the fact that in the original set of moment conditions $\boldsymbol{g}_T = \mathrm{E}[\boldsymbol{\lambda}_i y_{i,T}]$ does not appear as a parameter.

Robertson and Sarafidis (2015) show that restricted factor IV estimator (FIVR) is asymptotically more efficient than FIVU and consequently more efficient than procedures involving some form of differencing. Furthermore, the restrictions imposed on a subset of the nuisance parameters appear to provide substantial efficiency gains in finite samples. Notably, the autoregressive structure of the model implies a reduced form for $\boldsymbol{F}$, and as such the vector of structural parameters is identified even if the true value of $\boldsymbol{F}_B$ (as defined for the QLD estimator) is rank deficient.

Counting the total number of moment conditions and parameters, we have

$$\#moments = \frac{(K+1)T(T+1)}{2}; \quad \#parameters = (K+1)(1+TL) + L - (K+1)\frac{L(L-1)}{2}.$$

Remark 3.8. Note that in the model without any regressors (or if regressors are strictly exogenous), the $(K+1)L(L-1)/2$ term reduces to $L(L-1)/2$. Together with $L(L+1)/2$ restrictions imposed on $\boldsymbol{\Sigma}_{\lambda}$, one then has in total L^2 restrictions (which is a standard number of restrictions usually imposed for factor models).

Remark 3.9. In principle, we have T additional moment conditions (by the zero mean assumption of $\varepsilon_{i,t}$ for each time period t), given by

$$\boldsymbol{m}_{\iota} = \mathrm{vec}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k\right)' \boldsymbol{\iota}_N - \boldsymbol{F}\boldsymbol{g}_{\iota}\right).$$

Here $\boldsymbol{g}_{\iota}$ represents the mean of $\boldsymbol{\lambda}_i$. The same is exactly true for Ahn et al. (2013), although there exist $(T-L)$ moment conditions in that case.

3.4. Linearized QLD GMM

Hayakawa (2012) proposes a linearized GMM version of the QLD model in Ahn et al. (2013) under strict exogeneity, at the expense of introducing extra parameters. The moment conditions can be written as follows:

$$\boldsymbol{m}_\alpha = \text{vech}\left(\frac{1}{N}\left(\boldsymbol{Y}(\boldsymbol{J}+\boldsymbol{F}^*\tilde{\boldsymbol{J}}(L))' - \boldsymbol{Y}_{-1}(\alpha\boldsymbol{J}+\boldsymbol{F}^*_\alpha\tilde{\boldsymbol{J}}(L))' - \sum_{k=1}^{K}\boldsymbol{X}_k\left(\beta_k\boldsymbol{J}+\boldsymbol{F}^*_{\beta_k}\tilde{\boldsymbol{J}}(L)\right)'\right)'\boldsymbol{Y}_{-1}\boldsymbol{J}'\right),$$

$$\boldsymbol{m}_k = \text{vec}\left(\frac{1}{N}\left(\boldsymbol{Y}(\boldsymbol{J}+\boldsymbol{F}^*\tilde{\boldsymbol{J}}(L))' - \boldsymbol{Y}_{-1}(\alpha\boldsymbol{J}+\boldsymbol{F}^*_\alpha\tilde{\boldsymbol{J}}(L))' - \sum_{k=1}^{K}\boldsymbol{X}_k\left(\beta_k\boldsymbol{J}+\boldsymbol{F}^*_{\beta_k}\tilde{\boldsymbol{J}}(L)\right)'\right)'\boldsymbol{X}_k\right) \quad \forall k.$$

The parameters $\boldsymbol{F}^*_\alpha$, $\boldsymbol{F}^*_{\beta_k}$ do not appear in the estimator of Ahn et al. (2013). That estimator can be obtained directly by noting that

$$\boldsymbol{F}^*_\alpha = \alpha\boldsymbol{F}^*; \quad \boldsymbol{F}^*_{\beta_k} = \beta_k\boldsymbol{F}^*.$$

The linearized estimator is linear in parameters, and thereby, it is computationally easy to implement. On the other hand, this simplicity is not without price, as this estimator is not as efficient as the estimator in Ahn et al. (2013). In total, under strict exogeneity of all $x^{(k)}_{i,t}$, we have

$$\#moments = \frac{(T-L)(T-L+1)}{2} + KT(T-L),$$

$$\#parameters = \underbrace{K+1+(T-L)L}_{ALS} + \underbrace{(T-L)L(K+1)}_{Linearization} - \frac{L(L-1)}{2}.$$

Notice that the last term in the equation for the total number of parameters is not present in the original study of Hayakawa (2012). To explain the necessity of this term, consider the $(T-L)$th equation (for ease of exposition, we set $L=2$) *without* exogenous regressors

$$y_{i,T-2} - f^{(1)}_{T-2}y_{i,T} - f^{(2)}_{T-2}y_{i,T-1} = \alpha y_{i,T-3} + f^{(1)}_{\alpha T-2}y_{i,T-1} + f^{(2)}_{\alpha T-2}y_{i,T-2} + \varepsilon_{T-2,t} - f^{(1)}_{T-2}\varepsilon_{i,T} - f^{(2)}_{T-2}\varepsilon_{i,T-1}.$$

Clearly, only $f^{(2)}_{T-2} + f^{(1)}_{\alpha T-2}$ can be identified but not the individual terms separately. As a result $L(L-1)/2$ normalizations need to be imposed. Furthermore, as it can be easily seen, this term is unaltered if additional regressors are present in the model so long as they do not contain other lags of $y_{i,t}$ or lags of exogenous regressors.

Remark 3.10. Although not discussed in Hayakawa (2012), the same linearization strategy for the QD estimator of Holtz-Eakin et al. (1988) is also feasible.

In what follows, we consider more specifically the case where the covariates are weakly exogenous. To facilitate exposition, assume there exists a single weakly exogenous covariate. Observe that we can rewrite the first equation of the transformed model as

$$y_{i,1} + \sum_{l=1}^{L} f^{(l)}_1 y_{i,T-l} = \alpha y_{i,0} + \beta x_{i,1} + \sum_{l=1}^{L} f^{(l)}_{\alpha_1} y_{i,T-l-1} + \sum_{l=1}^{L} f^{(l)}_{\beta_1} x_{i,T-l} + \cdots. \tag{3.3}$$

This equation contains $2+3L$ unknown parameters, with only two available moment conditions (assuming $x_{i,0}$ is not observed, otherwise 3). Hence the full set of parameters in this equation cannot be identified without further normalizations. It then follows that the minimum value of T required in order to identify the structural parameters of interest is such that (for simplicity assume $L=1$)

$$2(T-1) = 2+3 \Longrightarrow \quad \min\{T\} = 1 + \lceil 2.5 \rceil = 4,$$

where $\lceil x \rceil$ is the smallest integer not less than x ("ceiling" function). For more general models with $K>1$, the condition $\min\{T\}=4$ continues to hold as

$$(K+1)(T-1) \geq (K+2)+(K+1) \Longrightarrow \quad \min\{T\} = 1 + \left\lceil \frac{2K+3}{K+1} \right\rceil = 4.$$

Notice that for the nonlinear estimator $\min\{T\} = 3$ in the single-factor case. As a result, for $L = 1$ under weak exogeneity, the number of identifiable parameters and moment conditions is given by

$$\#moments = (K+1)\frac{(T-L)(T-L+1)}{2} - (K+1),$$
$$\#parameters = \underbrace{K+1+(T-L)L}_{ALS} + \underbrace{(T-L)L(K+1)}_{Linearization} - \frac{L(L-1)}{2} - (K+2),$$

where $-(K+1)$ and $-(K+2)$ adjustments are made to take into account the fact that for $t = 1$ there are $(K+2)$ nuisance parameters to be estimated with $(K+1)$ available moment conditions. Both expressions can be similarly modified for $L > 1$.

3.5. Projection GMM

Following Bai (2013b),[7] Hayakawa (2012) suggests approximating $\boldsymbol{\lambda}_i$ using a Mundlak (1978)–Chamberlain (1982) type projection of the form

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi} \boldsymbol{z}_i + \boldsymbol{v}_i,$$

where $\boldsymbol{z}_i = (1, \boldsymbol{x}_i^{(1)\prime}, \ldots, \boldsymbol{x}_i^{(K)\prime}, y_{i,0})'$. Notice that by definition of the projection, $\mathrm{E}[\boldsymbol{v}_i \boldsymbol{z}_i'] = \mathbf{O}_{L\times(TK+2)}$. As a result, the stacked model for individual i can be written as

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{z}_i + \boldsymbol{F}\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i. \tag{3.4}$$

While Bai (2013b) proposes maximum likelihood estimation of the above model, Hayakawa (2012) advocates a GMM estimator; in our standard notation, the total set of moment conditions used by Hayakawa (2012) is given by

$$\boldsymbol{m}_\alpha = \frac{1}{N}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right)' \boldsymbol{Y}_{-1}\boldsymbol{e}_1,$$
$$\boldsymbol{m}_\iota = \frac{1}{N}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right)' \boldsymbol{\iota}_N,$$
$$\boldsymbol{m}_k = \mathrm{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right)' \boldsymbol{X}_k\right), \quad \forall k.$$

Assuming weak exogeneity of the covariates, one has

$$\#moments = 2T + \frac{KT(T+1)}{2},$$
$$\#parameters = \underbrace{(K+1)+(T-L)L}_{ALS} + \underbrace{L(TK+2)}_{Projection}.$$

Similarly to the FIVU estimator of Robertson and Sarafidis (2015), the number of identifiable parameters is smaller than the nominal one and depends on the projected variables $\boldsymbol{z}_i$.

3.5.1. Equivalence with FIVU

Following Bond and Windmeijer (2002), we consider a more general projection specification of the form

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi} \boldsymbol{z}_i + \boldsymbol{v}_i,$$

[7] Note that the first version of this article dates back to 2009.

where $\boldsymbol{z}_i = (\boldsymbol{x}_i^{(1)'}, \ldots, \boldsymbol{x}_i^{(K)'}, \boldsymbol{y}_{i,-1}')'$. The true value of $\boldsymbol{\Phi}$ has the usual expression for the projection estimator

$$\boldsymbol{\Phi}_0 \equiv \mathrm{E}\left[\boldsymbol{\lambda}_i \boldsymbol{z}_i'\right] \mathrm{E}\left[\boldsymbol{z}_i \boldsymbol{z}_i'\right]^{-1}.$$

The first term in the notation of Robertson and Sarafidis (2015) is simply

$$\mathrm{E}\left[\boldsymbol{\lambda}_i \boldsymbol{z}_i'\right] = \left(\boldsymbol{G}_1', \ldots, \boldsymbol{G}_K', \boldsymbol{G}'\right). \tag{3.5}$$

This estimator coincides asymptotically with the FIVU estimator of Robertson and Sarafidis (2015), as well as with the QLD GMM estimator of Ahn et al. (2013) and QD estimator of Holtz-Eakin et al. (1988) (for $L = 1$) if all $T(T+1)(K+1)/2$ moment conditions are used. A proof for the equivalence between FIVU, QLD, and QD GMM estimators is given in Robertson and Sarafidis (2015).

3.6. Linear GMM

In their discussion of the test for cross-sectional dependence, Sarafidis et al. (2009) observe that if one can assume

$$\boldsymbol{x}_{i,t} = \boldsymbol{\Pi}(\boldsymbol{x}_{i,t-1}, \ldots, \boldsymbol{x}_{i,0}) + \boldsymbol{\Gamma}_{xi}\boldsymbol{f}_t + \boldsymbol{\pi}(\varepsilon_{i,t-1}, \ldots, \varepsilon_{i,0}) + \boldsymbol{\varepsilon}_{i,t}^x \tag{3.6}$$

where $\boldsymbol{\Pi}(\cdot)$ and $\boldsymbol{\pi}(\cdot)$ are measurable functions, and the stochastic components are such that

$$\mathrm{E}[\boldsymbol{\varepsilon}_{i,s}^x \varepsilon_{i,l}] = \boldsymbol{0}_K, \forall s, l,$$
$$\mathrm{E}[\mathrm{vec}(\boldsymbol{\Gamma}_{xi})\boldsymbol{\lambda}_i'] = \boldsymbol{O}_{KL\times L},$$

then the following moment conditions are valid even in the presence of unobserved factors in both equations for $y_{i,t}$ and $\boldsymbol{x}_{i,t}$:

$$\mathrm{E}[(y_{i,t} - \alpha y_{i,t-1} - \boldsymbol{\beta}'\boldsymbol{x}_{i,t})\Delta \boldsymbol{x}_{i,s}] = 0, \forall s \le t,$$
$$\mathrm{E}[(\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \boldsymbol{\beta}'\Delta \boldsymbol{x}_{i,t})\boldsymbol{x}_{i,s}] = 0, \forall s \le t-1.$$

The total number of valid (*nonredundant*) moment conditions is given by

$$\#moments = K\left(\frac{(T-1)T}{2} + (T-1)\right),$$

if one does not include $\boldsymbol{x}_{i,0}$ and $\Delta \boldsymbol{x}_{i,1}$ among the instruments. Under mean stationarity, additional moment conditions become available in the equations in levels, giving rise to a system GMM estimator.

Identification of the structural parameters crucially depends on the condition that no lagged values of $y_{i,t}$ are present in (3.6) as well as on the assumption that the factor loadings of the y and x processes are uncorrelated. However, it is important to stress that all exogenous regressors are allowed to be weakly exogenous due to the possible nonzero $\boldsymbol{\pi}(\cdot)$ function, or even endogenous provided that $\varepsilon_{i,t}$ is serially uncorrelated.

3.7. Conditional quasi maximum likelihood (QML) estimator

To control for the correlation between the strictly exogenous regressors and the initial condition with factor loadings $\boldsymbol{\lambda}_i$, Bai (2013b), similarly to the GMM estimator proposed in Hayakawa (2012), considers a linear projection of the following form:

$$\boldsymbol{\lambda}_i = \boldsymbol{\Phi}\boldsymbol{z}_i + \boldsymbol{\nu}_i, \quad \mathrm{E}[\boldsymbol{\nu}_i \boldsymbol{\nu}_i'] = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}.$$

However, instead of relying on covariances as in the GMM framework, the quasi maximum likelihood (ML) approach makes use of the second moment estimator

$$\boldsymbol{S}(\boldsymbol{\theta}) = \frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right)'\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k \boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right),$$

where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}', \text{vec}\boldsymbol{F}', \text{vec}\boldsymbol{\Phi}')'$. Evaluated at the true values of the parameters, the expected value of $\boldsymbol{S}(\boldsymbol{\theta}_0)$ is

$$\mathrm{E}[\boldsymbol{S}(\boldsymbol{\theta}_0)] = \boldsymbol{\Sigma} = \boldsymbol{I}_T\sigma^2 + \boldsymbol{F}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\boldsymbol{F}'.$$

To solve the rotational indeterminacy problem, one can normalize $\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \boldsymbol{I}_L$ and redefine $\boldsymbol{F} \equiv \boldsymbol{F}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{1/2}$ and $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}^{-1/2}$, similarly to the FIVR estimator of Robertson and Sarafidis (2015). To evaluate the distance between $\boldsymbol{S}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}$, Bai (2013b)[8] suggests maximizing the following QML objective function to obtain consistent estimates of the underlying parameters

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\left(\log|\boldsymbol{\Sigma}| + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big)\right).$$

Under standard regularity conditions for M-estimators, the estimator obtained as the maximizer of the objective function $\ell(\boldsymbol{\theta})$ is consistent and asymptotically normal for fixed T, with asymptotic variance-covariance matrix of "sandwich" form irrespective of the distributional assumptions imposed on the combined error term $\varepsilon_{i,t} + \boldsymbol{\nu}_i'\boldsymbol{f}_t$. If one can replace the projection assumption by the assumption of conditional expectations, the resulting estimator can be seen as a QML estimator conditional on exogenous regressors $\boldsymbol{X}_k$ and the initial observation $y_{i,0}$.

The theoretical and finite sample properties of this estimator without factors are discussed in Alvarez and Arellano (2003), Kruiniger (2013), and Bun et al. (2016) among others, while Westerlund and Norkutė (2014) discuss the properties of this estimator for possibly nonstationary data with large T.

The above version of the estimator requires time series homoskedasticity in $\varepsilon_{i,t}$ for consistency. If this condition holds true and all covariates are strictly exogenous, the estimator provides efficiency gains over the GMM estimators analyzed before since the latter do not make use of moment conditions that exploit homoskedasticity (see, e.g., Ahn et al., 2001). The estimator can be modified in a straightforward manner under time series heterosedasticity to estimate all σ_t^2. On the other hand, cross-sectional heteroskedasticity cannot be allowed without additional restrictions.

Furthermore, the estimator generally requires $\tau = T$ in Assumption 4, i.e., strict exogeneity of the regressors. An exception to this is discussed in the following remark.

Remark 3.11. If it is plausible to assume that all covariates have the dynamic specification

$$x_{i,t}^{(k)} = \beta_x x_{i,t-1}^{(k)} + \alpha_x y_{i,t-1} + \boldsymbol{f}_t'\boldsymbol{\lambda}_i^{x(k)} + \varepsilon_{i,t}^x, \tag{3.7}$$

so that $x_{i,t}^{(k)}$ is possibly weakly exogenous, then according to Bai (2013b) it is sufficient to project on $(1, x_{i,0}^{(1)}, \ldots, x_{i,0}^{(K)}, y_{i,0})$ only, resulting in a more efficient estimator. A necessary condition for this approach to be valid is that the factor loadings $(\boldsymbol{\lambda}_i^{x(k)}, \boldsymbol{\lambda}_i)$ are independent, once conditioned on the initial observations $(1, x_{i,0}^{(1)}, \ldots, x_{i,0}^{(K)}, y_{i,0})$.

4. Some general remarks on the estimators

4.1. (Non)invariance to λ_i

In the situations where the model contains fixed effects only, i.e., $\boldsymbol{\lambda}_i'\boldsymbol{f}_t = \lambda_i$, some of the classical panel data estimators can be invariant to individual effects. For example, under mean stationarity of the initial condition the GMM estimators of Anderson and Hsiao (1982) (with instruments in first differences), Hayakawa (2009), or the Transformed ML estimators as in Hsiao et al. (2002), Kruiniger (2013), and Juodis (2016a) are invariant to the distribution of the fixed effects λ_i. In general, irrespective of the

[8]Strictly speaking in the aforementioned article the author solely describes the approach in terms of the likelihood function, while in Bai (2013a) the author describes a QML objective function as just one possibility.

properties of $y_{i,0}$, none of the estimators present in this article are invariant to $\boldsymbol{\lambda}_i'\boldsymbol{f}_t$ for fixed T. For GMM estimators, invariance would require knowledge of the whole history $\{\boldsymbol{f}_t\}_{t=-\infty}^{T}$ in order to construct instruments that are invariant to $\boldsymbol{\lambda}_i$. This conclusion is true both for estimators that involve some sort of differencing (QD, QLD) and projection (FIVU, Projection GMM).

4.2. Unbalanced samples

As it is mentioned in, e.g., Judois (2016b), the quasi-long-differencing transformation of Ahn et al. (2013) requires that for *all* individuals at least L common time indices observations are available to the researcher. In the model with weakly exogenous regressors this requirement is even more specific as the last L observations should be observed for all individuals. Otherwise, the $\boldsymbol{D}(\boldsymbol{F}^*)$ transformation matrix might become group-specific, if one can group observations based on availability.

To see this in more detail, consider Eq. (3.2). As it stands, the quasi-long-differencing transformation that removes the incidental parameters from the error is feasible for individual i only if the last L periods are available. Otherwise, these individuals may either be dropped out altogether, or be grouped such that it becomes possible to normalize on different $T - L$ periods. Either way, the estimator may suffer from a substantial loss in efficiency, as a result of removing observations, or splitting the sample. On the other hand, if it is plausible to assume that the model contains only strictly exogenous regressors, then it is sufficient that there exist L common time indices $t^{(1)}, \ldots, t^{(L)}$ where observations for all individuals are available.

The extension of FIVU and FIVR to unbalanced samples follows trivially by simply introducing indicators, depending on whether a particular moment condition is available for individual i or not (as for the standard fixed effects estimator).

The QD GMM estimator of Nauges and Thomas (2003) can be trivially modified as well, as in the standard Arellano and Bond (1991) procedure. However, similarly to that procedure, this transformation might result in dropping quite a lot of observations.

The projection estimator of Hayakawa (2012) requires further modification in order to take into account that projection variables $\boldsymbol{z}_i$ are not fully observed for each individual. We conjecture that the modification could be performed in a similar way as in the model without a factor structure, as discussed by Abrevaya (2013). For ML-based estimators, such extendability appears to be a more challenging task.

Remark 4.1. The above discussion relies on the fact that there exists a large enough number of consecutive time periods for each individual in the sample. For example, FIVU requires at least two consecutive periods and quasi-differencing type procedures require at least three. Under these circumstances, we note that estimators in their existing form may not be fully efficient. For example, if one observes *only* $y_{i,T}$ and $y_{i,T-2}$ for a substantial group of individuals, assuming exogenous covariates are available at all time periods, then one could use backward substitution and consider moment conditions within the FIVU framework, which are quadratic in the autoregressive parameter and result in efficiency gains. For projection-type methodologies, however, such substantial unbalancedness may affect the consistency of the estimators as one cannot substitute unobserved quantities for zeros in the projection term. This issue is discussed in detail by Abrevaya (2013).

4.3. Observed factors

In some situations one might wish to estimate models with both observed and unobserved factors at the same time. Taking the structure of observed factors into account may improve the efficiency of the estimators, although one can still consistently estimate the model by treating the observed factors as unobserved. One such possibility has been already discussed in Nauges and Thomas (2003) for models with an individual-specific, time-invariant effect. In this section we will briefly summarize implementability issues for all estimators when observed factors are present in the model alongside their unobserved counterparts.[9]

For the GMM estimators that involve some form of differencing, e.g. Holtz-Eakin et al. (1988) and Ahn et al. (2013), one can deal with observed factors using a similar procedure as in Nauges and Thomas (2003), that is, by removing the observed factors first (one-by-one) and then proceeding to remove the unobserved factors from the model. The first step can be most easily implemented using a quasi-differencing matrix $\boldsymbol{D}(\boldsymbol{r})$ with known weights.

For the GMM estimators of Robertson and Sarafidis (2015) (FIVU) and Hayakawa (2012), since the unobserved factors are not removed from the model, the treatment of the observed factors is somewhat easier. One merely needs to split the $\boldsymbol{F}\boldsymbol{G}'$ terms into two parts, observed and unobserved factors, and then proceed as in the case of unobserved factors. In this case the number of identifiable parameters will be smaller than in the case where one treats the observed factors as unobserved. As a result, one gains in efficiency, at the expense, however, of robustness.

For FIVR one needs to take care when solving for $\boldsymbol{F}$ in terms of the remaining parameters, because in the model with observed factors one estimates the variance-covariance matrix of the factor loadings for the observed factors, while for those which are unobserved their variance-covariance matrix is normalized.

The extension of the likelihood estimator of Bai (2013b) to observed factors can be implemented in a similar way to the projection GMM estimator. As in FIVR, one would have to estimate the variance-covariance matrix of the factor loadings for the observed factors, while the covariances of unobserved factors can be w.l.o.g. normalized as before.

5. Finite sample performance

This section investigates the finite sample performance of the estimators analyzed above using simulated data. Our focus lies on examining the effect of the presence of weakly exogenous covariates, the effect of changing the magnitude of the correlation between the factor loadings of the dependent variable and those of the covariates, as well as the impact of changing the number of moment conditions on bias and size for GMM estimators. We also investigate the effect of changing the level of persistence in the data, as well as the sample size in terms of both N and T.

5.1. Monte Carlo design

We consider model (2.1) with $K = 1$, i.e.,

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}; \quad u_{i,t} = \sum_{\ell=1}^{L} \lambda_{\ell,i} f_{\ell,t} + \varepsilon^{y}_{i,t}.$$

The process for $x_{i,t}$ and for f_t is given, respectively, by

$$x_{i,t} = \delta y_{i,t-1} + \alpha_x x_{i,t-1} + \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,t} + \varepsilon^{x}_{i,t},$$

$$f_{\ell,t} = \alpha_f f_{\ell,t-1} + \sqrt{1-\alpha_f^2}\,\varepsilon^{f}_{\ell,t}; \quad \varepsilon^{f}_{\ell,t} \sim \mathcal{N}(0,1), \quad \forall \ell.$$

The factor loadings are generated by $\lambda_{\ell,i} \sim \mathcal{N}(0,1)$ and

$$\gamma_{\ell,i} = \rho\lambda_{\ell,i} + \sqrt{1-\rho^2}\,\upsilon^{f}_{\ell,i}; \quad \upsilon^{f}_{\ell,i} \sim \mathcal{N}(0,1)\forall\ell,$$

[9] Under the assumption that appropriate regularity conditions hold, which prohibit asymptotic collinearity between the observed and unobserved factors.

where ρ denotes the correlation between the factor loadings of the y and x processes. Furthermore, the idiosyncratic errors are generated as[10]

$$\varepsilon^{y}_{i,t} \sim \mathcal{N}(0,1); \quad \varepsilon^{x}_{i,t} \sim \mathcal{N}\left(0,\sigma^2_x\right).$$

The starting period for the model is $t = -S$, and the initial observations are generated as

$$y_{i,-S} = \sum_{\ell=1}^{L} \lambda_{\ell,i} f_{\ell,-S} + \varepsilon^{y}_{i,-S}; \quad x_{i,-S} = \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,-S} + \varepsilon^{x}_{i,-S}; \quad f_{-S} \sim \mathcal{N}(0,1).$$

The signal-to-noise ratio (SNR) of the model is defined as follows:

$$SNR \equiv \frac{1}{T}\sum_{t=1}^{T} \frac{\mathrm{var}\left(y_{i,t}|\lambda_{\ell,i},\gamma_{\ell,i},\{f_{\ell,s}\}^{t}_{s=-S}\right)}{\mathrm{var}\varepsilon^{y}_{i,t}} - 1.$$

σ^2_x is set such that the SNR is equal 5 in all designs.[11] This particular value of SNR is chosen so that it is possible to control this measure across all designs. Lower values of SNR (e.g., 3 as in Bun and Kiviet, 2006) would require $\sigma^2_x < 0$ *ceteris paribus* in order to satisfy the desired equality for all designs.

We set $\beta = 1 - \alpha$ such that the long run parameter is equal to 1, $\alpha_x = 0.6$, $\alpha_f = 0.5$, and $L = 1$.[12] We consider $N = \{200; 800\}$ and $T = \{4; 8\}$. Furthermore, $\alpha = \{0.4; 0.8\}$, $\rho = \{0; 0.6\}$, and $\delta = \{0; 0.3\}$. The minimum number of replications performed equals 2,000 for each design, and the factors are drawn in each replication. The choice of the initial values of the parameters for the nonlinear algorithms is discussed in 7. When at least one of the estimators fails to converge in a particular replication, that replication is discarded.[13]

Note that for the QML estimator we use standard errors based on a "sandwich" variance-covariance matrix, as opposed to the simple inverse of the Hessian variance matrix. First-order conditions as well as Hessian matrices for likelihood estimators are obtained using analytical derivatives to speed up the computations.[14]

Although feasible, in this article we do not implement the linearized GMM estimator of Hayakawa (2012) adapted to weakly exogenous regressors. This is mainly due to the fact that this estimator merely provides an easy way to obtain starting values for the remaining estimators, which involve nonlinear optimization algorithms.

Motivated from our theoretical discussion regarding the estimators considered in this article, some implications can be discussed *a priori*, based on our Monte Carlo design.

1. When $\delta \neq 0$, likelihood based estimators are inconsistent because $x_{i,t}$ is not strictly exogenous, with the exception of the modified estimator of Bai (2013b) conditional on $(y_{i,0}, x_{i,0})$.
2. For $\rho \neq 0$, the likelihood estimator conditional on $(y_{i,0}, x_{i,0})$ is inconsistent because the conditional independence assumption is violated.

[10] We have also explored the effect of non-normal errors based on the chi-squared distribution (centered and normalized). The results were almost identical and therefore, to save space, we refrain from reporting them.

[11] To ensure this, we also set $S = 5$.

[12] Similar results have been obtained for $L = 2$. To avoid repeating similar conclusions, we refrain from reporting these results. We note that the number of factors can be estimated for all GMM estimators based on the model information criteria developed by Ahn et al. (2013). The performance of these procedures appears to be more than satisfactory; the interested reader may refer to the aforementioned article, as well as to the Monte Carlo study in Robertson et al. (2014). The size of L is treated as known in this article because there is currently no equivalent methodology proposed for testing the number of factors within the likelihood framework.

[13] For the numerical maximization, we used the BFGS method as implemented in the OxMetrics statistical software. Convergence is achieved when the difference in the value of the given objective function between two consecutive iterations is less than 10^{-4}. Other values of this criterion were considered in the preliminary study with similar qualitative conclusions, although the number of times particular estimators fail to converge varies. For further details on OxMetrics, see Doornik (2009). All algorithms are available upon request.

[14] In the preliminary study, results based on analytical and numerical derivatives were compared. Since the results were quantitatively and qualitatively almost identical (for designs where estimators were consistent), we prefer the use of analytical derivatives solely for practical reasons.

3. For $\rho = 0$ and $\delta = 0$, the projection GMM estimator might suffer from weak instruments, particularly when $\alpha = 0.8$, because $y_{i,0}$ remains the only relevant instrument and this might be weakly correlated with the regressors when the difference apart in time between $y_{i,0}$ and $y_{i,t}$ increases, i.e., as $t \to T$.

5.2. MC results

The results are reported in the Appendix in terms of median bias and root median square error (RMSE), which is defined as

$$RMedSE = \sqrt{\text{med}\left[(\widehat{\alpha}_r - \alpha)^2\right]},$$

where $\widehat{\alpha}_r$ denotes the value of α obtained in the rth replication using a particular estimator (and similarly for β). As an additional measure of dispersion, we report the radius of the interval centered on the median containing 80% of the observations, divided by *1.28*. This statistic, which we shall refer to as "quasi-standard deviation" (denoted qStd) provides an estimate of the population standard deviation if the distribution were normal, with the advantage that it is more robust to the occurrence of outliers compared to the usual expression for the standard deviation. The reason we report this statistic is that, on the one hand, the root mean square error is extremely sensitive to outliers, and on the other hand, it is fair to say that the root median square error does not depend on outliers pretty much at all. Therefore, the former could be unduly misleading given that in principle, for any given data set, one could estimate the model using a large set of different initial values in an attempt to avoid local minima, or lack of convergence in some cases (which we deal with in our experiments by discarding those particular replications). In a large-scale simulation experiment as ours, however, the set of initial values naturally needs to be restricted in some sensible/feasible way. The quasi-standard deviation lies in-between because while it provides a measure of dispersion that is less sensitive to outliers compared to the root mean square error, it is still more informative about the variability of the estimators relative to the root median square error. Finally, we report size, where nominal size is set at 5%.[15] For the GMM estimators, we also report size of the overidentifying restrictions (J) test statistic.[16]

Initially, we discuss results for the OLS estimator and the GMM estimator proposed by Sarafidis et al. (2009)[17] as well as the linearized GMM estimator of Hayakawa (2012) (see Table B.1); these estimators have been used to obtain initial values for the parameters for the nonlinear estimators, among other (random) choices. In many circumstances, the OLS estimator exhibits large median bias, while the size of the estimator is most often not far from unity. On the other hand, the linear GMM estimator proposed by Sarafidis et al. (2009) does fairly well both in terms of bias and RMedSE when $\delta = 0$ and $\rho = 0$, i.e., when the covariate is strictly exogenous with respect to the total error term, $u_{i,t}$. The size of the estimator appears to be somewhat upwardly distorted, especially for T large, but one expects that this would substantially improve if one made use of the finite-sample correction proposed by Windmeijer (2005). On the other hand, the estimator is not consistent for the remaining parameterizations of our design, and this is well reflected in its finite sample performance. Notably, the J statistic appears to have high power to detect violations of the null, even if N is small. In the online appendix of this article, we present results for GMM estimators when only a subset of moment conditions is used for estimation.

With regards to the linearized GMM estimator of Hayakawa (2012), both median bias and RMedSE are reasonably small, even for $N = 200$, so long as $\delta = 0$, i.e., under strict exogeneity of x with respect

[15]In actual fact, the results on size also partially reflect extreme tail performance of the estimators. Following the suggestion by a referee, an online appendix of the article (see http://arturas.economist.lt/JS_online.pdf) reports results in terms of root mean square error (RMSE) and standard deviation. We will comment on these results at the end of this section.

[16]To calculate the J statistic, we use the uncentered weighting matrix evaluated based on the first step estimators. Alternatively, one can use a centered weighting matrix. However, simulation (and theoretical) evidence in the dynamic panel data context in Bun and Poldermans (2015) and Hayakawa (2016) suggest that such procedure can have worse size properties (oversized) with similar size-adjusted power. In our preliminary study using the FIVU estimator, a similar behavior was observed for the factor model, which confirms the aforementioned findings.

[17]See Table 1 in the supplementary material (http://arturas.economist.lt/JS_online.pdf).

to the idiosyncratic error. However, the estimator appears to be quite sensitive to high values of α, both in terms of bias and qStd, an outcome that may be partially related to the fact that the value of β is small in this case, which implies that a many-weak instruments' type problem might arise. Naturally, the performance of the estimator deteriorates for $\delta = 0.3$ as the moment conditions are invalidated in this case. While the size of the J statistic appears to be distorted upwards when the estimator is consistent, it has in general quite large power to detect violations of strict exogeneity, and for high values of α this holds true even with a relatively small size of N.

Table B.2 report results for the quasi-long-differenced GMM estimator proposed by Ahn et al. (2013). The estimator appears to have small median bias under all designs. This is expected given that the estimator is consistent. The qStd results indicate that the estimator has large dispersion in some designs, especially when T is small. We have explored further the underlying reason for this result. We found that this is often the case when the value of the factor at the last time period, i.e., f_T, is relatively close to zero. Thus, the estimator appears to be potentially sensitive to this issue, because the normalization scheme sets $f_T = 1$.[18] The two-step version improves on these results. On the other hand, inferences based on one-step estimates seem to be relatively more reliable. This outcome may be attributed to the standard argument provided for linear GMM estimators, which is that two-step estimators rely on an estimate of the variance-covariance matrix of the moment conditions, which, in samples where N is small, can lead to conservative standard errors. Truncating the moment conditions for $T = 8$ seems to have a negligible effect on the size properties of the one-step estimator but does improve size for the two-step estimator quite substantially (see Table 2 in the online appendix). This result seems to apply for all overidentified GMM estimators actually. The J statistic exhibits small size distortions upwards.

Simulation results with regards to the QD GMM estimator by Holtz-Eakin et al. (1988) are reported in Table B.3. As we can see, qualitative similar conclusions apply as above, except that the dispersion of the estimator in terms of RMedSE and qStd is substantially larger than that of QLD GMM. As explained in Subsection 3.1, this may be attributed to the fact that the QD transformation involves $r_t = f_t/f_{t-1}$, which requires that f_t, $t = 1, \dots T - 1$, lie sufficiently far from zero; otherwise, the estimator may face convergence problems.

Tables B.4 and B.5 report results for FIVU and FIVR based on full sets of moment conditions, proposed by Robertson and Sarafidis (2015). Similarly to Ahn et al. (2013), both estimators have very small median bias in all circumstances. Furthermore, they perform well in terms of qStd. Especially the two-step versions have small dispersion regardless of the design. Naturally, the dispersion decreases further with high values of T because the degree of overidentification of the model increases. As expected, Root Median Squared Error (RMedSE) appears to go down roughly at the rate of $\sqrt{N}$. FIVR dominates FIVU, which is not surprising given that the former imposes overidentifying restrictions arising from the structure of the model and thus it estimates a smaller number of parameters. The size of one-step FIVU and FIVR estimators is close to its nominal value in all circumstances. On the other hand, the two-step versions appear to be size distorted when T is large, especially when $N = 200$. The distortion decreases when only a subset of the moment conditions is used; see Tables 4 and 5 in the online appendix. Thus, one may conclude that using the full set of moment conditions and relying on inferences based on first-step estimates is a sensible strategy. From the empirical point of view, this is appealing because it simplifies matters regarding how many instruments to be used; an important question that often arises in two-way error component models estimated using linear GMM estimators. Finally, the size of the J statistic is often slightly distorted when N is small, but improves rapidly as N increases.

The projection GMM estimator proposed by Hayakawa (2012) (Table B.6) has small bias and performs well in general in terms of qStd unless α is close to unity, in which case outliers seem to occur relatively more frequently. One could suspect that this design is the worst case scenario for the estimator because only $y_{i,0}$ is included in the set of instruments, while lagged values of $x_{i,t}$ are only weakly

[18]It turns out that this problem has already been known in the literature; see, e.g., Kruiniger (2008, p. 16). Notice that normalizing the factor value at a different time period would result in losing moment conditions, as explained in the main text; for example, normalizing $f_{T-1} = 1$ ($f_{T-2} = 1$) results in dropping T ($2T - 1$) moment conditions.

correlated with $y_{i,t-1}$. Inferences based on the first-step estimator are reasonably accurate, certainly more so compared to the two-step version, although the latter improves for the truncated set of moment conditions (Table 6 in the online appendix). The J statistic seems to be size-distorted downwards but it slowly improves for larger values of N.

Remark 5.1. Monte Carlo evidence in Juodis and Sarafidis (2015) suggest that the standard error correction as in Windmeijer (2005) can substantially improve the empirical size of the two-step FIVU estimator. We suspect that the same is also applicable to the estimators of Ahn et al. (2013) and Hayakawa (2012). However, extensive analysis of this issue is beyond the scope of this article.

Finally, Table B.7 reports results for the conditional maximum likelihood estimator proposed by Bai (2013b). The left panel corresponds to the estimator that treats $x_{i,t}$ as strictly exogenous with respect to the idiosyncratic error, while the panel on the right-hand side corresponds to the estimator that is consistent under weak exogeneity of a first-order form, which is satisfied in our design when $\rho = 0$. Interestingly, the former appears to exhibit negligible median bias in all cases, even when both δ and ρ take nonzero values. The dispersion of the estimator is small as well, unless $T = 4$ and $\delta = 0.3$. Likewise, for $\delta = 0.3$ the size of the estimator is distorted upwards, and it gets worse with higher values of N, which is natural given that the estimator is not consistent in this case. However, for cases where this estimator is consistent ($\delta = 0$ and $\rho = 0$), it may serve as a benchmark because it has negligible bias and excellent size. This can be expected given the asymptotic optimality of this estimator under conditional homoskedasticity of $\varepsilon_{i,t}$. This conclusion is pretty much invariant to different values of N, T or ρ. The second estimator, in designs with $\rho = 0.6$ (where it is not consistent) tends to have substantial bias for both α and β. On the other hand, when it is supposed to be consistent ($\delta = 0.3, \rho = 0.0$) it is more size distorted than the first estimator that is inconsistent. This is a somewhat puzzling finding.

Figure 1 provides a snapshot illustration of our discussion regarding the size properties of the estimators for the autoregressive coefficient when $N = 200, T = 4$, while Figure 2 illustrates root mean

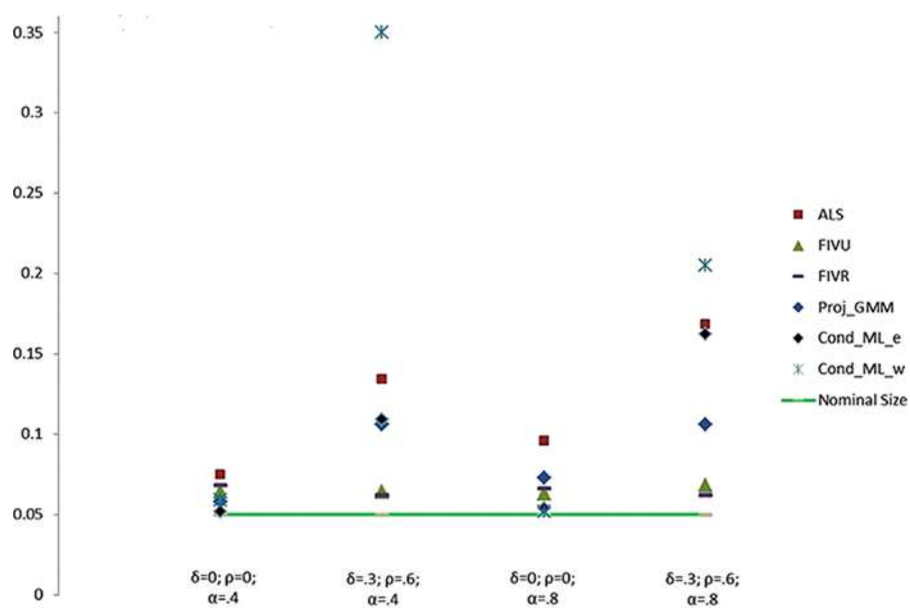

Figure 1. Empirical rejection frequencies; $N = 200, T = 4$.

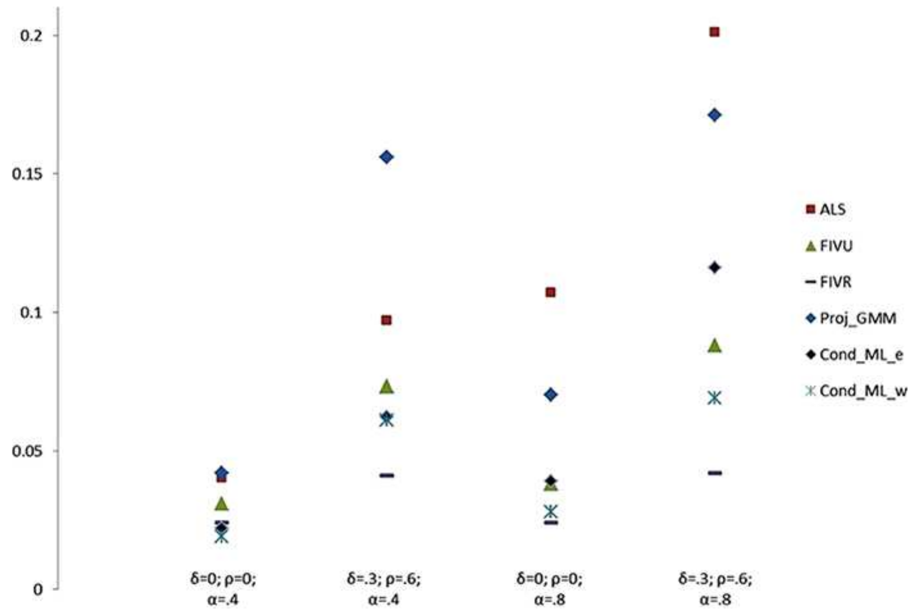

Figure 2. RMSE of estimators; $N = 200, T = 4$.

squared error performance of the estimators for the case where $N = 200, T = 4$.[19] The size of the estimators improves as N increases and, for the two-step GMM estimators, it deteriorates as T increases.

In order to provide a broader picture of the performance of the estimators, the online appendix of our article presents alternative results in terms of mean bias and RMSE. The conclusions are qualitatively similar to what we have already alluded, in that the ranking of the estimators in terms of their performance is clearly preserved. For most estimators mean bias and median bias are of similar magnitude, which implies that differences between root mean and RMSEs are mainly attributed to the dispersion of the estimators and, ultimately, outliers. As discussed previously, the frequency of such outliers can possibly be reduced to some extent by enlarging the set of initial values sufficiently. However, in a large-scale and sophisticated simulation experiment, as it is ours, the set of initial values naturally needed to be restricted in some sensible and feasible way.

6. Empirical illustration: Income elasticity of crime

There is a well-established literature on the effect of crime rate on economic activity. Essentially, crime may act on the entire economy like a tax, which increases uncertainty, reduces economic competitiveness and discourages foreign direct investment (see, e.g., Anderson, 1999). At the same time, there is relatively little empirical evidence emanating from econometric analysis, which provides some quantification for the opposite cause-and-effect relationship — that is, the impact of income on crime rate.

Income may affect crime rate through several channels. For instance, during prolonged periods of economic hardship a higher proportion of the population may become unemployed, possibly leading to more property crime and robberies, as criminals steal coveted items they cannot afford.

In addition, the consequences of being arrested and found guilty of a criminal offence include not just the punishment meted out by the criminal justice system, but also the indirect sanctions imposed by society; a convicted individual may no longer enjoy the same opportunities in the labor market, and so the opportunity cost of lost income increases with higher existing income levels. With this line of

[19] We do not include results for the QD and Linearized QLD estimators in order to maintain a small enough scale on the vertical axis, such that the difference in the performance of the remaining estimators is clearly visible.

reasoning, the effect of income on certain types of crime, such as property crime, is expected to be negative.[20]

A negative average effect of income on property crime is also in accordance with economic theory. In particular, according to the seminal articles by Becker (1968) and Ehrlich (1975), individuals engage in criminal activity because the subjective expected benefit exceeds the expected cost of doing so. Criminals, therefore, do not differ from the rest of society in their basic motivation but in their appraisal of benefits and costs. The idea of a rational criminal suggests that since the opportunity cost with respect to crime increases with higher levels of income, the latter may exert a negative influence on crime.

In this section, we examine the effect of income on crime using a panel data set of 153 Local Government Areas (LGA) in NSW, each observed over a period of 7 years. The Australian Standard Geographic Classification defines the LGA as the lowest level of aggregation, following the census Collection District (CD) and Statistical Local Area. Thus, the LGA represents a low level of aggregation compared to standard practice in the literature, where regressions using city-, state-, and country-level data are common.[21] The time interval of the sample spans 2006–2012, and includes a period where the Australian economy did slow down quite significantly as a result of the global financial crisis, although it did not fall into recession.[22]

We consider the following model

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}; \quad u_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{i,t}; \quad i = 1, \ldots, 153, t = 1, \ldots, 6, \tag{6.1}$$

where $y_{i,t}$ denotes (the log of) property crime rate, defined as property crime incidents divided by population, in LAG i at time t and $x_{i,t}$ denotes (the log of) average disposable income in real prices. The error term is composite and contains an unknown number of factors, plus a purely idiosyncratic component. The factors may represent common shocks that hit all LGAs, albeit with different intensities. Failing to take into account of such error structure may lead to inconsistent estimates of the long-run coefficient, and misleading inferences, since factors are correlated with the lagged dependent variable by construction, and most likely with income; see, e.g., Sarafidis and Robertson (2009).

We have fitted models with $L = 1, 2, 3$ factors. The number of factors has been estimated using the criterion

$$BIC_\ell = J_\ell - f(N, T) \times g(c, df(\ell)), \tag{6.2}$$

where J_ℓ denotes the value of the GMM objective function obtained by fitting ℓ factors into the model, while $f(N, T) = \log(N)/T^{0.3}$ and $g(c, df(\ell)) = 0.75 \times df(\ell)$, while $df(\ell)$ denotes the number of degrees of freedom associated with ℓ factors.[23] For all GMM estimators, the value of L that minimizes BIC equals unity, and therefore we set $\widehat{L} = 1$. To save space, in what follows we report results for this value only.

The results are reported in Table 1. "FD GMM" refers to the first-differenced GMM estimator of Arellano and Bond (1991), and "System GMM" is the system GMM estimator put forward by Blundell and Bond (1998); both are used as a benchmark, since they are available in popular econometric software packages, and therefore, they are widely used by empirical practitioners.[24] "Linear GMM" and "Linearized GMM" refer to the estimators discussed in Sections 3.6 and 3.4, respectively. The remaining estimators are self-explanatory, with "1" or "2" referring to the one- and two-step versions of the GMM estimators, while "s" denotes whether a subset of the total number moment conditions is used. In this

[20]The effect of income on other types of crime may not be as clear-cut. For example, better economic times might also translate into a higher demand for drugs.

[21]Each CD contains on average about 225 households (2001 Census). There are about 37,000 CDs throughout Australia. The boundaries of an SLA are designed to be typically coterminous with Local Government Areas unless the LGA does not fit entirely into a Statistical Subdivision, or is not of a comparative nature to other LGAs. There are 193 SLAs in NSW.

[22]See, for example, McDonald and Morling (2011).

[23]See BIC1 in Ahn et al. (2013, p. 8). The performance of the criterion in the context of a dynamic panel is investigated in Robertson and Sarafidis (2015).

[24]Common time effects have been included for both estimators, since this is currently common practice in estimating dynamic panel data models. As discussed by Sarafidis and Robertson (2009), the inclusion of common time effects is one way to reduce the effect of factor residuals on estimation.

Table 1. Estimation results, $\widehat{L} = 1$.

	$\widehat{\alpha}$	st.error	z-stat.	$\widehat{\beta}$	st.error	z-stat.	LR	J	df
FD GMM	0.553	0.173	30.19	−0.139	0.218	−0.640	−0.310	22.3*	14
System GMM	0.722	0.118	6.12	−0.094	0.113	−0.830	−0.336	24.3	18
Linear GMM	0.621	0.043	14.5	−0.101	0.012	−8.64	−0.265	38.2*	18
Linearized GMM	0.034	0.019	1.79	0.043	0.018	2.34	0.045	18.1	28
QLD GMM 1 s	0.867	0.172	5.03	−0.032	0.033	−0.948	−0.265	—	—
QLD GMM 2 s	0.712	0.075	9.41	−0.054	0.022	−2.53	−0.189	21.2	21
QLD GMM 1	0.866	0.173	4.99	−0.032	0.034	−0.934	−0.238	—	—
QLD GMM 2	0.834	0.046	18.1	−0.039	0.012	−3.30	−0.236	27.4	23
FIVU 1 s	0.764	0.113	6.74	−0.052	0.029	−1.76	−0.219	—	—
FIVU 2 s	0.895	0.030	30.3	−0.030	0.008	−3.79	−0.284	16.5	17
FIVU 1	0.856	0.153	5.57	−0.032	0.032	−1.01	−0.219	—	—
FIVU 2	0.736	0.062	11.9	−0.057	0.017	−3.36	−0.218	23.0	23
FIVR 1 s	0.802	0.128	6.27	−0.042	0.031	−1.35	−0.212	—	—
FIVR 2 s	0.890	0.026	34.6	−0.031	0.007	−4.36	−0.278	20.0	21
FIVR 1	0.848	0.148	5.73	−0.035	0.032	−1.10	−0.229	—	—
FIVR 2	0.730	0.058	12.4	−0.061	0.016	−3.90	−0.222	25.5	27
Proj. GMM 1 s	0.812	0.098	8.30	−0.046	0.024	−1.92	−0.242	—	—
Proj. GMM 2 s	0.741	0.053	14.1	−0.061	0.015	−4.22	−0.238	14.9	15
Proj. GMM 1	0.813	0.098	8.27	−0.045	0.024	−1.91	−0.242	—	—
Proj. GMM 2	0.741	0.055	13.8	−0.061	0.015	−4.08	−0.236	15.8	18
ML strict. exog.	0.587	0.151	3.88	−0.084	0.033	−2.53	−0.203	—	—
ML weak. exog.	0.592	0.064	9.20	−0.085	0.016	−5.38	−0.207	—	—

case, the estimators make use of the four most recent instruments available with respect to the lagged dependent variable. Finally, "Proj. GMM" refers to the projection GMM estimator discussed in Section 3.5.[25]

"LR" denotes the long-run estimated average effect of income on crime, "J" denotes the test statistic for overidentifying restrictions, and finally "df" denotes the number of degrees of freedom in the model, i.e., the number of moment conditions minus the number of identifiable parameters. Both "J" and "df" are applicable only for GMM estimators, and the overidentifying restrictions test statistic is asymptotically valid for two-step GMM estimators only, and hence we just report these results. "*" denotes statistical significance at the 5% level.

Starting values for the nonlinear estimators have been obtained in a way similar to the Monte Carlo study, described in 7, except that the set of random initializations is much larger, due to the fact that using a large set of starting values for a single data set is not as time-consuming as it would have been in the MC section. In particular, for estimators that require starting values only for the structural (α, β) parameters, such as QLD GMM, we have used additionally 200 $\mathcal{U}[-1; 1]$ random variables. For the FIVU and Projection GMM estimators 200 random values of (α, β) formed the basis to extract principal components, which were subsequently used as starting values for the iterations. For other estimators (FIVR, QML) that also require specifying starting values for the nuisance parameters (e.g., $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{\Phi}$ matrices), we used 200 sets of $\mathcal{U}[-5; 5]$ random variables. The uniform interval of random draws has been expanded compared to the simulations because the computational burden is far smaller, and we wish to eliminate local minima/maxima of the criterion function to the extent that is possible. Furthermore, unlike structural parameters (α, β), nuisance parameters do not have an intuitive interval of plausible population values, thus it is important not to consider a very tight interval used for starting values of these parameters.

As we can see, in most cases both α and β are statistically significant at the 5% level of significance. For most estimators, the long-run estimated coefficient ranges between −0.2 and −0.28. Notable exceptions are FD GMM and System GMM, which appear to overestimate the average effect of income on crime in absolute terms. The J statistic for FD GMM is statistically significant at the 5% level, which indicates that

[25] We do not report results with respect to QD GMM, given its similarities with QLD GMM.

the model is misspecified. This is expected because this estimator is not consistent under a multifactor error structure. Ironically, this appears not to be the case for System GMM, which however should yield asymptotically the same conclusion, as it makes use of the same moment conditions as FD GMM plus instruments with respect to equations in levels. We are tempted to conjecture that this may be attributed to lack of power of this type of *J* test in finite samples.

We note that even if the differences between the values of the *LR* coefficient obtained from FD GMM/System GMM and the remaining estimators might not appear to be very substantial, the policy implications derived from these results are significantly different. For instance, the estimate of β is about -0.139 for FD GMM, -0.061 for Proj. GMM 2, and -0.061 for FIVR 2. This means that the estimated short-run income elasticity of crime (i.e., the average sensitivity of crime to changes in income within the same time period) is more than double for DIF GMM. Moreover, these estimates are statistically different at the 5% level. In addition, the estimated autoregressive coefficient is about 0.552 for FD GMM, 0.741 for Proj. GMM 2, and 0.730 for FIVR 2. Thus, given these results, it takes approximately 3 time periods on average for 90% of the long run effect to be realized under FD GMM and 7 time periods for Proj. GMM 2 and FIVR 2. Therefore, it is clear that the results bear distinctive policy implications, since the estimated short-run effect of income, as well as the dynamics of the crime rate process, are substantially different across the estimators.

The value of the *J* statistic for the linear GMM estimator of Sarafidis et al. (2009) is statistically significant, which indicates that the factor component is correlated with income. In terms of the parametrization design in the MC section, this means that $\delta \neq 0$ and/or $\rho \neq 0$. As shown in the Monte Carlo section of this paper, this type of *J* test has very good power to detect such violations from the null hypotheses in finite samples. The estimated coefficients appear to be biased in the direction of FD GMM/System GMM.

The linearized GMM estimator yields a value of the autoregressive coefficient that is close to zero, and a positive value for the slope coefficient, which is counterintuitive. This may be due to the fact that the estimator relies on the assumption that the regressor is strictly exogenous. Since a substantial body of the literature argues that crime *does* have an effect on economic activity, and thereby on average disposable income, this assumption is likely to be violated. That is, the parameter δ introduced in the MC section is unlikely to equal zero.[26]

Proj. GMM and FIVR appear to provide similar results. This is consistent with the hypothesis that neither $\delta = 0$, nor $\rho = 0$, i.e., strict exogeneity of income is violated and the factor loadings are likely to be correlated. Moreover, QLD GMM and FIVU yield fairly similar results to Proj. GMM and FIVR as well. In most cases, the estimated coefficients are not statistically different. For example, the upper bound of the 95% confidence interval for α obtained using FIVU 2 is approximately 0.860, which exceeds the value of α obtained using QLD GMM 2.

The maximum likelihood estimator that imposes strict exogeneity of income appears to be biased towards the direction of FD GMM. This would confirm that the variable income is only weakly exogenous, i.e., $\delta \neq 0$ in terms of the parametrization employed in the MC section. Interestingly, similar results were obtained for the version of the maximum likelihood estimator that imposes weak exogeneity of income. It might be useful to note here that this estimator was prone to substantial numerical instabilities for $L = 2$. Given the very good performance of ML in the Monte Carlo section, we conjecture that these results might be driven from a possible violation of the assumption that the factor loadings of the y and x processes are independent, conditional on the initial observations, i.e., a violation of the restriction $\rho = 0$. Future research might shed more light onto this issue.

In summary, it appears that for this particular model and data set, GMM-based estimators may be better suited for estimation and inference.

[26] All other GMM estimators employed in this application treat income as weakly exogenous. Based on the results of the J statistic obtained from these estimators, which show that it is not significant in all cases, one can infer that weak exogeneity of income, versus endogeneity, is supported by the data.

7. Conclusion

In this article we have analyzed a group of fixed T dynamic panel data estimators with a multifactor error structure. All currently available estimators have been presented using a unified notational approach. Both their theoretical properties as well as possible limitations are discussed. We have considered a model with a lagged dependent variable and additional regressors, possibly weakly exogenous or endogenous. We found that the number of identifiable parameters for the GMM estimators can be smaller than what can be found in the literature. This result is of major importance for practitioners when performing model selection based on overidentifying test statistics. Theoretical discussions in this article were complemented by a finite sample study based on Monte Carlo simulation.

We designed our Monte Carlo exercise to shed some light on the relative merits of the various estimation approaches. It was found that the likelihood estimator of Bai (2013b), when consistent, can serve as a benchmark in that it has negligible bias and good size control, irrespective of the sample size. Under such circumstances, the FIVR estimator proposed by Robertson and Sarafidis (2015) performs closely as well. However, FIVR is more robust to violations from strict exogeneity, as well as from the no conditional correlation condition between the factor loadings. The latter applies to other GMM estimators as well, at least provided that the cross-sectional dimension is large enough.

This article assumes that the time-series dimension is fixed. Bai (2013b) shows that the presence of factors does not result in an incidental parameters problem for the conditional maximum likelihood estimator as far as the structural parameters are concerned. A natural question to ask is whether GMM estimators in models where the number of parameters and number of moment conditions grows with T suffer from an incidental parameters problem. We leave this issue for future research.

Appendices

Appendix A. Implementation

Appendix A.1. Starting values for non-linear estimators

This appendix discusses the choice of starting values used for the nonlinear optimization algorithms.

Ahn et al. (2013). This estimator can be implemented through an iterative procedure. Iterations start given some set of initial values for the structural parameters, α, β. For this purpose, we use both the one- and two-step linearized GMM estimator as proposed by Hayakawa (2012), as well as the OLS estimator. The two-step estimator is implemented in exactly the same way except that the set of initial values for the structural parameters includes the one-step estimator. Once final estimates of $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\boldsymbol{F}}$ are obtained, these are used as initial values in the nonlinear optimization algorithm, which optimizes all parameters at once. This is implemented in order to make sure that we indeed find the global minimum of the objective function.

QD. Starting values for the QD estimator have been obtained in the same way as with the estimator by Ahn et al. (2013)

FIVU. Similarly to the previous estimator, FIVU can also be implemented in steps. Iterations start given a set of starting values for the factors $\boldsymbol{F}$. This set is obtained using the linearized GMM estimator, estimates of the principal components extracted from OLS residuals, and one set of uniform random variables on $[-1; 1]$. Unlike for Ahn et al. (2013), joint nonlinear optimization is not used as a final step in order to save computational time.

FIVR. For this estimator, the main source of starting values is obtained from FIVU with the starting value of $\boldsymbol{g}_T$ implied in terms of other parameters. Other starting values include those based on the OLS estimator and the one- and two-step linearized GMM estimator. In this case, starting values for the nuisance parameters $\boldsymbol{G}$ are simply drawn from uniform $[-1; 1]$.

Projection GMM. This estimator is implemented in exactly the same way as Ahn et al. (2013), i.e., firstly an iterative procedure is used, followed by a nonlinear one. Starting values for the factors are obtained using the principal components extracted from OLS residuals, the estimate of $\boldsymbol{f}$ obtained from the linearized GMM estimator, and two sets of uniform random variables on $[-1; 1]$. In order to uniquely identify all parameters up to rotation, we impose $f_T = 1$ in estimation. We suspect that, similarly to FIVU, one can estimate the model without normalizations and perform a degrees of freedom correction at the end. We leave this question open for future research.

QML. Starting values for the structural parameters are obtained using the linearized GMM estimator, OLS, and two sets of uniform random variables on $[-1; 1]$. The remaining parameters (including $\log(\sigma^2)$) are drawn as uniform random variables on $[0; 1]$. In the preliminary study we also tried $[-1; 1]$; however, the results were identical. Alternatively, one could also use the principal component estimates of $\boldsymbol{F}$ obtained from OLS residuals, as suggested by Bai (2013b).

Subset GMM Estimators. For $T = 8$ when both the subset and full-set GMM estimators are available, we estimate the subset estimators first using the algorithms as described above and then use the subset estimator as starting values for the estimators that make use of the full set of moment conditions.

Appendix A.2. Specifics: Ahn et al. (2013)

To describe the procedure assume for simplicity that there are no $x's$, such that the only available moment conditions are

$$\boldsymbol{m}_l = \frac{1}{N}\text{vech}\left(\boldsymbol{J}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1}\right)' \boldsymbol{Y}_{-1}\boldsymbol{J}' + \boldsymbol{F}^*\tilde{\boldsymbol{J}}(L)\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1}\right)' \boldsymbol{Y}_{-1}\boldsymbol{J}'\right).$$

The objective function for this estimator is simply given by

$$f(\alpha, \text{vec}(\boldsymbol{F}^*)) = \boldsymbol{m}_l' \boldsymbol{W}_N \boldsymbol{m}_l.$$

For any given value of α, the moment conditions are linear vec$(\boldsymbol{F}^*)$. That is,

$$\boldsymbol{m}_l = \text{vech}(\boldsymbol{Z}) + \boldsymbol{B}_{(T-L)}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{T-L})\text{vec}(\boldsymbol{F}^*) = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}.$$

Here $\boldsymbol{Z}$ and $\boldsymbol{Q}$ are given by

$$\begin{aligned}
\boldsymbol{Z} &= \frac{1}{N}\boldsymbol{J}\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1}\right)' \boldsymbol{Y}_{-1}\boldsymbol{J}', \\
\boldsymbol{Q} &= \frac{1}{N}\tilde{\boldsymbol{J}}(L)\left(\boldsymbol{Y} - \alpha \boldsymbol{Y}_{-1}\right)' \boldsymbol{Y}_{-1}\boldsymbol{J}', \\
\boldsymbol{y} &= \text{vech}(\boldsymbol{Z}), \\
\boldsymbol{X} &= \boldsymbol{B}_{(T-L)}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{T-L}), \\
\boldsymbol{\beta} &= -\text{vec}(\boldsymbol{F}^*).
\end{aligned}$$

Hence the usual formula for the OLS estimator implies that

$$-\text{vec}(\boldsymbol{F}^*) = \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{W}_N\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}_N\boldsymbol{y}.$$

If, on the other hand, $\boldsymbol{F}^*$ is known, then α is obtained in exactly the same way with $\boldsymbol{\beta} = \alpha$, while

$$\begin{aligned}
\boldsymbol{y} &= \frac{1}{N}\text{vech}(\boldsymbol{D}(\boldsymbol{\Phi}^*)\boldsymbol{Y}'\boldsymbol{Y}_{-1}\boldsymbol{J}'), \\
\boldsymbol{X} &= \frac{1}{N}\text{vech}(\boldsymbol{D}(\boldsymbol{\Phi}^*)\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\boldsymbol{J}').
\end{aligned}$$

Appendix A.3. Specifics: Restricted estimator of Robertson and Sarafidis (2015)

The moment conditions are given by

$$\begin{aligned}
\boldsymbol{m}_l &= \text{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{Y}_{-1} - \boldsymbol{F}\boldsymbol{G}'\right), \\
\boldsymbol{m}_k &= \text{vech}\left(\frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{X}_k - \boldsymbol{F}\boldsymbol{G}_k'\right) \quad \forall k.
\end{aligned}$$

$\boldsymbol{F}$ obeys the following restriction:

$$\boldsymbol{F} = \left(\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T\right) \boldsymbol{G} + \boldsymbol{e}_T \boldsymbol{g}_T' - \sum_{k=1}^{k} \beta_k \boldsymbol{G}_k.$$

The differential of vec$\boldsymbol{F}$ is simply given by

$$\begin{aligned}\mathrm{dvec}\boldsymbol{F} = &-\mathrm{vec}(\boldsymbol{G})\mathrm{d}\alpha + \left(\boldsymbol{I}_L \otimes \left(\boldsymbol{L}_T' - \alpha \boldsymbol{I}_T\right)\right) \mathrm{dvec}\boldsymbol{G} \\ &- \sum_{k=1}^{K} \mathrm{vec}(\boldsymbol{G}_k)\mathrm{d}\beta_k - (\boldsymbol{I}_L \otimes \boldsymbol{I}_T) \sum_{k=1}^{K} \beta_k \mathrm{dvec}\boldsymbol{G}_k \\ &+ (\boldsymbol{I}_L \otimes \boldsymbol{e}_T)\, \mathrm{d}\boldsymbol{g}_T.\end{aligned}$$

By the chain rule for differentials, we have

$$\begin{aligned}\mathrm{d}\boldsymbol{m}_l = &-\frac{1}{N}\mathrm{vech}\left(\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\mathrm{d}\alpha - \sum_{k=1}^{K} \frac{1}{N}\mathrm{vech}\left(\boldsymbol{X}_k'\boldsymbol{Y}_{-1}\right)\mathrm{d}\beta_k \\ &- \boldsymbol{B}_T\left(\mathbf{K}_{T,T}(\boldsymbol{F} \otimes \boldsymbol{I}_T)\mathrm{d}(\mathrm{vec}\boldsymbol{G}) + (\boldsymbol{G} \otimes \boldsymbol{I}_T)\mathrm{d}(\mathrm{vec}\boldsymbol{F})\right).\end{aligned}$$

Here the commutation matrix $\mathbf{K}_{a,b}$ is defined such that for any $[a \times b]$ matrix $\boldsymbol{A}$, $\mathrm{vec}(\boldsymbol{A}') = \mathbf{K}_{a,b}\mathrm{vec}(\boldsymbol{A})$. The result for $\mathrm{d}\boldsymbol{m}_k$ follows analogously.

Appendix A.4. Specifics: Bai (2013b)

Some specific results for this estimator can be written as follows:

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{\Sigma}_\tau + \boldsymbol{F}\boldsymbol{F}', \\ \boldsymbol{\Sigma}_\tau &= \sigma^2 \boldsymbol{I}_T, \\ \boldsymbol{v}_i &= \boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\gamma} - \boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{z}_i.\end{aligned}$$

The corresponding differentials are

$$\begin{aligned}\mathrm{d}\boldsymbol{\Sigma} &= \boldsymbol{I}_T\mathrm{d}\sigma^2 + \boldsymbol{F}(\mathrm{d}\boldsymbol{F})' + (\mathrm{d}\boldsymbol{F})\boldsymbol{F}', \\ \mathrm{d}^2\boldsymbol{\Sigma} &= 2(\mathrm{d}\boldsymbol{F}\mathrm{d}\boldsymbol{F}'), \\ \mathrm{d}\boldsymbol{v}_i &= -\boldsymbol{W}_i(\mathrm{d}\boldsymbol{\gamma}) - \mathrm{d}(\boldsymbol{F})\boldsymbol{\Phi}\boldsymbol{z}_i - \boldsymbol{F}\mathrm{d}(\boldsymbol{\Phi})\boldsymbol{z}_i, \\ \mathrm{d}^2\boldsymbol{v}_i &= -2(\mathrm{d}(\boldsymbol{F})\mathrm{d}(\boldsymbol{\Phi})\boldsymbol{z}_i).\end{aligned}$$

Denoting as $\boldsymbol{V}(\boldsymbol{\theta})$ the following $[N \times T]$ matrix (with the ith row being simply $\boldsymbol{v}_i'$)

$$\boldsymbol{V}(\boldsymbol{\theta}) = \frac{1}{N}\left(\boldsymbol{Y} - \alpha\boldsymbol{Y}_{-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{X}_k - \boldsymbol{Z}\boldsymbol{\Phi}'\boldsymbol{F}'\right),$$

then the score vector, using matrix notation rather than sums, is simply given by

$$\nabla(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Y}_{-1}\right) \\ \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{X}_1\right) \\ \vdots \\ \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{X}_K\right) \\ -0.5\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right) \\ -\mathrm{vec}\left(\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}\right) + \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Z}\boldsymbol{\Phi}'\right) \\ \mathrm{vec}\left(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{V}(\boldsymbol{\theta})'\boldsymbol{Z}\right) \end{pmatrix}.$$

Appendix A.5. Specifics: Hessians of likelihood based-estimators

Observe that the general structure of the likelihood function is given by

$$-\frac{2}{N}\ell(\theta) = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\big(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{S}(\boldsymbol{\theta})\big).$$

Using the rules for differentials (see, e.g., Magnus and Neudecker, 2007) the first differential of the two components is given by

$$\begin{aligned}\mathrm{d}\log|\boldsymbol{\Sigma}| &= \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\big),\\ \mathrm{dtr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big) &= -\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\big),\end{aligned}$$

where for simplicity the dependence on $\boldsymbol{\theta}$ has been dropped. By the chain rule for differentials it follows similarly that the second differential for the log-determinant is of the form

$$\mathrm{d}^2\log|\boldsymbol{\Sigma}| = \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{\Sigma})\big) - \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\big),$$

while the trace component is given by

$$\begin{aligned}\mathrm{d}^2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big) &= 2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big) - 2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\big)\\ &\quad - \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\big).\end{aligned}$$

We can combine both terms such that

$$\begin{aligned}-\frac{2}{N}\mathrm{d}^2\ell(\theta) &= \mathrm{tr}\big(\big(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\big)\,\mathrm{d}^2\boldsymbol{\Sigma}\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\big)\\ &\quad + \mathrm{tr}\big(\big(2\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\big)\,(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\big) - 2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\big).\end{aligned}$$

Note that, evaluated at any consistent estimate of $\hat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} &= {}_p(1),\\ 2\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-1} + {}_p(1).\end{aligned}$$

Hence from the asymptotic point of view, this is equivalent to considering the following consistent estimate of the Hessian:

$$-\frac{2}{N}\mathrm{d}^2\ell(\theta) = \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}^2\boldsymbol{S})\big) + \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\big) - 2\mathrm{tr}\big(\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{S})\big).$$

In our Monte Carlo study, we will make use of these facts and ignore the ${}_p(1)$ terms. Now let us consider the differentials of $\boldsymbol{S}$ in more detail. We have

$$\begin{aligned}\mathrm{d}\boldsymbol{S} &= \frac{1}{N}\sum_{i=1}^{N}\big(\boldsymbol{v}_i\mathrm{d}(\boldsymbol{v}_i)' + \mathrm{d}(\boldsymbol{v}_i)\boldsymbol{v}_i'\big),\\ \mathrm{d}^2\boldsymbol{S} &= \frac{1}{N}\sum_{i=1}^{N}\big(2\mathrm{d}(\boldsymbol{v}_i)\mathrm{d}(\boldsymbol{v}_i)' + \mathrm{d}^2(\boldsymbol{v}_i)\boldsymbol{v}_i' + \boldsymbol{v}_i\mathrm{d}^2(\boldsymbol{v}_i)'\big).\end{aligned}$$

Note that if evaluated at any consistent estimator of $\hat{\boldsymbol{\theta}}$

$$\frac{1}{N}\sum_{i=1}^{N}\big(\mathrm{d}^2(\boldsymbol{v}_i)\boldsymbol{v}_i' + \boldsymbol{v}_i\mathrm{d}^2(\boldsymbol{v}_i)'\big) = {}_p(1).$$

However, in our Monte Carlo study, we retain the corresponding terms in the formula of the estimate for the Hessian matrix. Furthermore, note that

$$\mathrm{vecd}\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{v}_i\otimes\boldsymbol{I}_T + \boldsymbol{I}_T\otimes\boldsymbol{v}_i\right)\mathrm{d}(\boldsymbol{v}_i).$$

Appendix B. Monte Carlo results

Table B.1. Linearized estimator of Hayakawa (2012) with strict exogeneity assumption.

Designs					GMM 1 step								GMM 2 step								J
					α				β				α				β				
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	−0.004	0.030	0.097	0.060	−0.005	0.030	0.099	0.057	−0.003	0.025	0.077	0.120	−0.008	0.024	0.076	0.111	0.125
200	4	0.4	0.0	0.3	−0.059	0.065	0.160	0.214	−0.160	0.168	0.247	0.504	−0.032	0.051	0.142	0.306	−0.189	0.190	0.215	0.812	0.239
200	4	0.4	0.6	0.0	−0.012	0.031	0.109	0.079	0.000	0.029	0.103	0.068	−0.007	0.026	0.085	0.147	−0.005	0.025	0.080	0.128	0.133
200	4	0.4	0.6	0.3	−0.085	0.086	0.181	0.291	−0.160	0.174	0.262	0.503	−0.059	0.065	0.150	0.404	−0.195	0.196	0.228	0.831	0.216
200	4	0.8	0.0	0.0	−0.060	0.077	0.216	0.193	−0.010	0.025	0.084	0.085	−0.060	0.074	0.209	0.281	−0.014	0.023	0.077	0.194	0.179
200	4	0.8	0.0	0.3	−0.322	0.322	0.301	0.768	−0.125	0.127	0.134	0.643	−0.348	0.348	0.320	0.930	−0.157	0.157	0.096	0.905	0.098
200	4	0.8	0.6	0.0	−0.075	0.090	0.242	0.236	−0.008	0.025	0.084	0.095	−0.072	0.088	0.243	0.345	−0.017	0.026	0.076	0.207	0.193
200	4	0.8	0.6	0.3	−0.347	0.347	0.305	0.761	−0.126	0.130	0.134	0.627	−0.380	0.380	0.334	0.938	−0.157	0.157	0.089	0.905	0.082
200	8	0.4	0.0	0.0	−0.003	0.022	0.075	0.094	0.000	0.023	0.078	0.092	0.000	0.015	0.048	0.339	−0.003	0.015	0.047	0.333	0.108
200	8	0.4	0.0	0.3	−0.064	0.070	0.168	0.279	−0.015	0.063	0.219	0.157	−0.029	0.039	0.102	0.525	−0.058	0.068	0.142	0.695	0.642
200	8	0.4	0.6	0.0	−0.012	0.024	0.092	0.117	0.010	0.022	0.091	0.113	−0.006	0.017	0.056	0.372	0.004	0.015	0.051	0.331	0.114
200	8	0.4	0.6	0.3	−0.080	0.080	0.200	0.374	−0.007	0.063	0.267	0.186	−0.042	0.044	0.117	0.584	−0.057	0.073	0.164	0.707	0.583
200	8	0.8	0.0	0.0	−0.024	0.029	0.092	0.165	−0.003	0.015	0.051	0.080	−0.020	0.024	0.071	0.433	−0.005	0.011	0.036	0.311	0.118
200	8	0.8	0.0	0.3	−0.201	0.201	0.179	0.820	−0.048	0.074	0.215	0.317	−0.193	0.193	0.149	0.991	−0.086	0.095	0.126	0.852	0.600
200	8	0.8	0.6	0.0	−0.029	0.033	0.106	0.216	0.004	0.015	0.063	0.111	−0.025	0.027	0.079	0.476	−0.002	0.011	0.038	0.319	0.104
200	8	0.8	0.6	0.3	−0.208	0.208	0.185	0.884	−0.048	0.077	0.252	0.340	−0.200	0.200	0.137	0.996	−0.089	0.097	0.137	0.869	0.508
800	4	0.4	0.0	0.0	−0.005	0.028	0.102	0.081	−0.007	0.032	0.117	0.078	−0.002	0.023	0.074	0.143	−0.006	0.023	0.076	0.117	0.149
800	4	0.4	0.0	0.3	−0.066	0.069	0.122	0.478	−0.192	0.194	0.227	0.726	−0.037	0.055	0.128	0.603	−0.215	0.215	0.178	0.979	0.818
800	4	0.4	0.6	0.0	−0.008	0.028	0.108	0.093	−0.003	0.033	0.114	0.087	−0.004	0.023	0.083	0.160	−0.005	0.024	0.084	0.142	0.160
800	4	0.4	0.6	0.3	−0.078	0.078	0.125	0.549	−0.200	0.203	0.194	0.773	−0.054	0.057	0.118	0.605	−0.229	0.229	0.175	0.980	0.732
800	4	0.8	0.0	0.0	−0.082	0.098	0.302	0.255	−0.020	0.035	0.123	0.144	−0.073	0.087	0.292	0.339	−0.021	0.031	0.122	0.266	0.203
800	4	0.8	0.0	0.3	−0.389	0.389	0.307	0.892	−0.148	0.149	0.121	0.806	−0.436	0.436	0.321	0.981	−0.178	0.178	0.067	0.995	0.549
800	4	0.8	0.6	0.0	−0.106	0.118	0.316	0.307	−0.022	0.037	0.120	0.156	−0.099	0.112	0.341	0.422	−0.028	0.036	0.118	0.312	0.233
800	4	0.8	0.6	0.3	−0.409	0.409	0.311	0.887	−0.151	0.152	0.107	0.824	−0.458	0.458	0.308	0.985	−0.182	0.182	0.051	0.991	0.436
800	8	0.4	0.0	0.0	−0.003	0.019	0.079	0.088	−0.002	0.024	0.099	0.112	0.000	0.011	0.035	0.208	−0.004	0.012	0.039	0.199	0.167
800	8	0.4	0.0	0.3	−0.066	0.069	0.117	0.515	−0.019	0.052	0.157	0.290	−0.013	0.025	0.066	0.528	−0.085	0.087	0.089	0.915	1
800	8	0.4	0.6	0.0	−0.007	0.020	0.077	0.106	0.002	0.022	0.092	0.113	−0.003	0.012	0.036	0.209	−0.002	0.012	0.037	0.173	0.163
800	8	0.4	0.6	0.3	−0.072	0.073	0.117	0.585	−0.027	0.053	0.166	0.314	−0.019	0.024	0.057	0.511	−0.094	0.096	0.083	0.952	1
800	8	0.8	0.0	0.0	−0.027	0.029	0.107	0.242	−0.004	0.019	0.071	0.091	−0.023	0.024	0.075	0.415	−0.008	0.012	0.040	0.250	0.193
800	8	0.8	0.0	0.3	−0.185	0.185	0.141	0.884	−0.057	0.067	0.125	0.531	−0.182	0.182	0.140	0.984	−0.103	0.104	0.089	0.974	1
800	8	0.8	0.6	0.0	−0.031	0.033	0.112	0.275	−0.003	0.019	0.071	0.091	−0.025	0.026	0.079	0.459	−0.008	0.013	0.039	0.259	0.192
800	8	0.8	0.6	0.3	−0.192	0.192	0.136	0.926	−0.062	0.073	0.133	0.572	−0.188	0.188	0.124	0.993	−0.109	0.110	0.076	0.977	1

Bias is the median bias of the estimator; RmedSE is the root median squared error; sStd is the quasi standard deviation; Size is the empirical rejection frequencies of the *t*-test for the parameter of interest. All results are based on 2000 Monte Carlo replications.

Table B.2. GMM estimator of Ahn et al. (2013).

					GMM 1 step								GMM 2 step								
Designs					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	0.001	0.028	0.087	0.075	−0.002	0.026	0.085	0.056	−0.001	0.022	0.067	0.137	0.000	0.021	0.065	0.102	0.097
200	4	0.4	0.0	0.3	−0.001	0.055	0.200	0.109	−0.005	0.057	0.199	0.111	−0.007	0.038	0.134	0.148	0.000	0.041	0.137	0.158	0.085
200	4	0.4	0.6	0.0	−0.005	0.029	0.097	0.094	0.004	0.025	0.083	0.063	−0.004	0.023	0.074	0.150	0.002	0.020	0.063	0.091	0.094
200	4	0.4	0.6	0.3	−0.020	0.048	0.211	0.134	0.013	0.049	0.217	0.117	−0.013	0.037	0.134	0.141	0.005	0.037	0.127	0.138	0.081
200	4	0.8	0.0	0.0	−0.004	0.029	0.107	0.096	−0.001	0.016	0.056	0.058	−0.005	0.022	0.083	0.146	0.000	0.013	0.045	0.099	0.102
200	4	0.8	0.0	0.3	−0.014	0.043	0.424	0.182	−0.004	0.038	0.292	0.166	−0.013	0.034	0.373	0.197	−0.003	0.029	0.270	0.198	0.122
200	4	0.8	0.6	0.0	−0.007	0.032	0.117	0.110	0.003	0.016	0.053	0.067	−0.007	0.022	0.086	0.142	0.002	0.013	0.044	0.092	0.106
200	4	0.8	0.6	0.3	−0.016	0.039	0.323	0.168	0.006	0.034	0.125	0.098	−0.013	0.032	0.273	0.193	0.001	0.027	0.103	0.151	0.107
200	8	0.4	0.0	0.0	−0.001	0.022	0.077	0.109	0.000	0.022	0.081	0.100	−0.001	0.015	0.049	0.315	0.000	0.014	0.045	0.257	0.106
200	8	0.4	0.0	0.3	0.008	0.054	0.205	0.133	−0.011	0.057	0.219	0.128	0.001	0.029	0.105	0.341	−0.002	0.029	0.102	0.332	0.078
200	8	0.4	0.6	0.0	−0.006	0.024	0.092	0.142	0.004	0.020	0.076	0.100	−0.004	0.017	0.058	0.356	0.002	0.013	0.043	0.239	0.085
200	8	0.4	0.6	0.3	−0.014	0.046	0.235	0.144	0.010	0.047	0.246	0.141	−0.007	0.027	0.116	0.323	0.006	0.027	0.110	0.296	0.091
200	8	0.8	0.0	0.0	−0.005	0.021	0.072	0.104	0.001	0.013	0.044	0.063	−0.002	0.015	0.050	0.288	0.001	0.009	0.028	0.197	0.095
200	8	0.8	0.0	0.3	−0.005	0.035	0.133	0.099	0.003	0.037	0.133	0.096	−0.004	0.022	0.079	0.280	0.002	0.023	0.076	0.263	0.074
200	8	0.8	0.6	0.0	−0.006	0.021	0.080	0.113	0.002	0.012	0.045	0.076	−0.003	0.015	0.054	0.295	0.001	0.008	0.027	0.195	0.093
200	8	0.8	0.6	0.3	−0.010	0.033	0.134	0.118	0.010	0.036	0.146	0.113	−0.005	0.021	0.075	0.264	0.006	0.023	0.076	0.241	0.075
800	4	0.4	0.0	0.0	−0.002	0.025	0.085	0.090	0.002	0.029	0.105	0.092	−0.001	0.018	0.057	0.123	0.001	0.021	0.068	0.120	0.096
800	4	0.4	0.0	0.3	−0.002	0.033	0.124	0.106	−0.001	0.033	0.126	0.119	−0.003	0.021	0.070	0.122	0.000	0.022	0.072	0.124	0.105
800	4	0.4	0.6	0.0	−0.005	0.024	0.086	0.102	0.005	0.025	0.097	0.086	−0.003	0.019	0.060	0.136	0.002	0.019	0.064	0.091	0.096
800	4	0.4	0.6	0.3	−0.008	0.028	0.115	0.111	0.005	0.027	0.121	0.111	−0.005	0.019	0.063	0.110	0.002	0.019	0.066	0.109	0.100
800	4	0.8	0.0	0.0	−0.004	0.020	0.076	0.096	0.000	0.018	0.059	0.078	−0.004	0.017	0.058	0.136	0.000	0.015	0.048	0.093	0.088
800	4	0.8	0.0	0.3	−0.005	0.022	0.094	0.127	−0.002	0.021	0.079	0.124	−0.004	0.017	0.067	0.132	−0.001	0.016	0.059	0.130	0.111
800	4	0.8	0.6	0.0	−0.006	0.019	0.073	0.101	0.001	0.019	0.063	0.064	−0.005	0.016	0.065	0.143	0.000	0.016	0.052	0.085	0.090
800	4	0.8	0.6	0.3	−0.006	0.021	0.089	0.127	0.002	0.021	0.074	0.098	−0.005	0.017	0.070	0.138	0.000	0.017	0.054	0.115	0.106
800	8	0.4	0.0	0.0	0.001	0.022	0.083	0.136	−0.001	0.027	0.111	0.123	−0.001	0.010	0.035	0.220	0.000	0.013	0.041	0.186	0.141
800	8	0.4	0.0	0.3	0.003	0.029	0.115	0.109	−0.004	0.030	0.118	0.119	−0.001	0.012	0.040	0.176	0.001	0.012	0.040	0.173	0.123
800	8	0.4	0.6	0.0	−0.004	0.019	0.079	0.143	0.003	0.021	0.088	0.113	−0.001	0.012	0.038	0.237	0.001	0.012	0.037	0.154	0.139
800	8	0.4	0.6	0.3	−0.005	0.023	0.117	0.133	0.004	0.024	0.120	0.126	−0.002	0.012	0.039	0.170	0.002	0.012	0.038	0.150	0.114
800	8	0.8	0.0	0.0	−0.002	0.013	0.045	0.083	0.000	0.015	0.051	0.076	−0.001	0.008	0.027	0.175	0.000	0.009	0.027	0.125	0.110
800	8	0.8	0.0	0.3	−0.002	0.017	0.063	0.083	0.001	0.017	0.063	0.083	−0.001	0.009	0.029	0.137	0.001	0.010	0.030	0.134	0.097
800	8	0.8	0.6	0.0	−0.003	0.013	0.046	0.087	0.000	0.015	0.052	0.083	−0.001	0.008	0.030	0.183	0.000	0.009	0.027	0.115	0.116
800	8	0.8	0.6	0.3	−0.003	0.015	0.056	0.093	0.002	0.016	0.063	0.088	−0.001	0.008	0.027	0.117	0.001	0.009	0.028	0.108	0.095

See Table B.1.

Table B.3. QD GMM estimator of Holtz-Eakin et al. (1988).

Designs					GMM 1 step								GMM 2 step								J
					α				β				α				β				
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	0.003	0.028	0.088	0.090	0.002	0.028	0.094	0.077	0.000	0.025	0.080	0.177	0.001	0.026	0.081	0.170	0.110
200	4	0.4	0.0	0.3	−0.011	0.057	0.218	0.141	0.011	0.057	0.226	0.145	−0.008	0.047	0.166	0.198	0.008	0.050	0.171	0.195	0.108
200	4	0.4	0.6	0.0	−0.002	0.032	0.138	0.149	0.008	0.029	0.472	0.158	−0.002	0.028	0.108	0.227	0.006	0.026	0.209	0.212	0.084
200	4	0.4	0.6	0.3	−0.029	0.057	0.519	0.249	0.032	0.060	0.628	0.239	−0.021	0.051	0.466	0.298	0.024	0.053	0.559	0.277	0.149
200	4	0.8	0.0	0.0	−0.001	0.030	0.102	0.101	0.004	0.018	0.060	0.076	0.000	0.024	0.084	0.183	0.001	0.016	0.051	0.156	0.125
200	4	0.8	0.0	0.3	−0.020	0.048	0.432	0.215	−0.001	0.041	0.228	0.176	−0.011	0.037	0.202	0.243	−0.001	0.034	0.122	0.214	0.147
200	4	0.8	0.6	0.0	−0.004	0.033	0.122	0.146	0.006	0.019	0.070	0.112	−0.002	0.027	0.102	0.212	0.005	0.016	0.059	0.182	0.123
200	4	0.8	0.6	0.3	−0.031	0.053	0.625	0.280	0.017	0.042	0.747	0.212	−0.019	0.044	0.587	0.329	0.015	0.040	0.674	0.275	0.142
200	8	0.4	0.0	0.0	−0.002	0.026	0.080	0.157	0.002	0.025	0.080	0.136	−0.002	0.018	0.056	0.385	0.001	0.017	0.053	0.315	0.307
200	8	0.4	0.0	0.3	−0.010	0.057	0.273	0.210	0.010	0.056	0.288	0.204	−0.009	0.036	0.148	0.449	0.008	0.036	0.151	0.420	0.237
200	8	0.4	0.6	0.0	−0.007	0.029	0.182	0.262	0.006	0.023	0.519	0.192	−0.005	0.021	0.117	0.462	0.004	0.017	0.370	0.343	0.169
200	8	0.4	0.6	0.3	−0.035	0.057	0.485	0.315	0.032	0.054	0.559	0.277	−0.022	0.040	0.376	0.489	0.020	0.035	0.427	0.428	0.240
200	8	0.8	0.0	0.0	−0.002	0.023	0.073	0.162	0.001	0.015	0.045	0.105	−0.001	0.018	0.056	0.411	0.001	0.010	0.032	0.269	0.286
200	8	0.8	0.0	0.3	−0.016	0.040	0.204	0.234	0.003	0.042	0.153	0.194	−0.012	0.029	0.150	0.432	0.001	0.030	0.102	0.376	0.210
200	8	0.8	0.6	0.0	−0.005	0.026	0.085	0.225	0.002	0.014	0.046	0.114	−0.003	0.019	0.065	0.443	0.003	0.010	0.031	0.259	0.215
200	8	0.8	0.6	0.3	−0.031	0.049	0.649	0.335	0.026	0.047	0.795	0.279	−0.023	0.034	0.551	0.473	0.017	0.032	0.650	0.393	0.270
800	4	0.4	0.0	0.0	0.000	0.013	0.041	0.077	0.000	0.013	0.040	0.068	−0.001	0.011	0.034	0.135	0.000	0.011	0.034	0.126	0.100
800	4	0.4	0.0	0.3	−0.003	0.026	0.086	0.101	0.003	0.027	0.089	0.102	−0.002	0.020	0.067	0.139	0.002	0.020	0.067	0.133	0.102
800	4	0.4	0.6	0.0	−0.003	0.015	0.055	0.124	0.002	0.014	0.049	0.119	−0.001	0.013	0.044	0.159	0.001	0.011	0.040	0.135	0.110
800	4	0.4	0.6	0.3	−0.010	0.026	0.391	0.175	0.008	0.025	0.474	0.173	−0.006	0.021	0.285	0.205	0.005	0.021	0.369	0.195	0.161
800	4	0.8	0.0	0.0	−0.001	0.015	0.046	0.086	0.001	0.009	0.027	0.068	−0.001	0.011	0.036	0.138	0.000	0.007	0.021	0.121	0.112
800	4	0.8	0.0	0.3	−0.008	0.022	0.123	0.172	0.002	0.019	0.066	0.135	−0.003	0.017	0.065	0.166	0.000	0.014	0.047	0.153	0.138
800	4	0.8	0.6	0.0	−0.001	0.015	0.053	0.100	0.001	0.009	0.028	0.076	−0.001	0.011	0.039	0.138	0.001	0.007	0.021	0.123	0.108
800	4	0.8	0.6	0.3	−0.011	0.024	0.402	0.227	0.005	0.019	0.163	0.176	−0.006	0.018	0.324	0.234	0.003	0.015	0.157	0.205	0.170
800	8	0.4	0.0	0.0	0.000	0.012	0.037	0.121	0.000	0.012	0.035	0.099	0.000	0.007	0.022	0.241	0.000	0.007	0.020	0.197	0.303
800	8	0.4	0.0	0.3	−0.003	0.025	0.091	0.148	0.003	0.026	0.092	0.144	−0.002	0.013	0.044	0.266	0.001	0.013	0.043	0.235	0.234
800	8	0.4	0.6	0.0	−0.002	0.013	0.060	0.200	0.002	0.011	0.049	0.157	−0.001	0.008	0.030	0.304	0.001	0.007	0.024	0.206	0.234
800	8	0.4	0.6	0.3	−0.009	0.024	0.331	0.230	0.008	0.024	0.328	0.210	−0.004	0.013	0.080	0.288	0.003	0.012	0.072	0.237	0.230
800	8	0.8	0.0	0.0	0.000	0.011	0.035	0.116	0.000	0.008	0.022	0.087	0.000	0.007	0.022	0.262	0.000	0.004	0.013	0.178	0.261
800	8	0.8	0.0	0.3	−0.005	0.018	0.071	0.179	0.001	0.019	0.067	0.149	−0.003	0.011	0.042	0.264	0.000	0.011	0.035	0.205	0.205
800	8	0.8	0.6	0.0	−0.001	0.012	0.040	0.165	0.000	0.007	0.022	0.089	−0.001	0.007	0.025	0.290	0.000	0.004	0.013	0.153	0.217
800	8	0.8	0.6	0.3	−0.008	0.020	0.616	0.240	0.007	0.020	0.768	0.189	−0.004	0.012	0.479	0.301	0.003	0.012	0.552	0.232	0.245

See Table B.1.

Table B.4. FIVU estimator of Robertson and Sarafidis (2015).

Designs					GMM 1 step								GMM 2 step								J
					α				β				α				β				
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	0.001	0.023	0.068	0.064	−0.002	0.022	0.065	0.048	0.000	0.021	0.061	0.073	−0.001	0.021	0.060	0.061	0.031
200	4	0.4	0.0	0.3	0.008	0.045	0.132	0.072	−0.004	0.043	0.136	0.068	−0.003	0.036	0.111	0.085	0.001	0.038	0.113	0.085	0.031
200	4	0.4	0.6	0.0	0.000	0.023	0.069	0.063	0.001	0.020	0.060	0.041	0.000	0.022	0.064	0.079	0.000	0.019	0.057	0.064	0.029
200	4	0.4	0.6	0.3	−0.008	0.036	0.107	0.064	0.006	0.036	0.116	0.064	−0.006	0.033	0.100	0.068	0.003	0.034	0.102	0.079	0.031
200	4	0.8	0.0	0.0	0.000	0.024	0.075	0.063	0.000	0.014	0.042	0.053	−0.001	0.020	0.061	0.070	0.001	0.012	0.040	0.069	0.035
200	4	0.8	0.0	0.3	−0.003	0.030	0.099	0.060	0.003	0.026	0.088	0.065	−0.003	0.028	0.089	0.076	0.002	0.024	0.079	0.080	0.038
200	4	0.8	0.6	0.0	−0.002	0.025	0.079	0.063	0.001	0.013	0.041	0.043	−0.002	0.020	0.066	0.071	0.002	0.012	0.038	0.066	0.033
200	4	0.8	0.6	0.3	−0.006	0.029	0.093	0.068	0.004	0.026	0.082	0.069	−0.004	0.028	0.089	0.084	0.002	0.025	0.079	0.085	0.035
200	8	0.4	0.0	0.0	0.002	0.014	0.042	0.072	−0.002	0.013	0.041	0.071	0.001	0.012	0.036	0.182	0.000	0.011	0.034	0.160	0.032
200	8	0.4	0.0	0.3	0.012	0.034	0.097	0.080	−0.014	0.034	0.099	0.085	0.004	0.021	0.063	0.173	−0.004	0.022	0.065	0.180	0.035
200	8	0.4	0.6	0.0	0.000	0.014	0.042	0.065	0.000	0.012	0.035	0.061	0.000	0.013	0.037	0.179	0.000	0.011	0.033	0.135	0.032
200	8	0.4	0.6	0.3	−0.004	0.025	0.080	0.056	0.003	0.026	0.079	0.054	−0.002	0.020	0.060	0.174	0.002	0.020	0.061	0.158	0.034
200	8	0.8	0.0	0.0	−0.001	0.013	0.038	0.053	0.000	0.008	0.025	0.050	0.000	0.011	0.034	0.168	0.000	0.007	0.023	0.143	0.037
200	8	0.8	0.0	0.3	−0.001	0.022	0.066	0.051	0.001	0.023	0.068	0.048	−0.001	0.018	0.054	0.163	0.001	0.018	0.057	0.155	0.036
200	8	0.8	0.6	0.0	−0.001	0.014	0.039	0.051	0.000	0.008	0.023	0.055	0.000	0.012	0.035	0.164	0.001	0.007	0.022	0.140	0.037
200	8	0.8	0.6	0.3	−0.004	0.020	0.060	0.048	0.005	0.023	0.066	0.048	−0.003	0.018	0.053	0.156	0.002	0.019	0.057	0.153	0.030
800	4	0.4	0.0	0.0	0.000	0.020	0.061	0.060	0.000	0.022	0.073	0.066	0.000	0.017	0.051	0.069	−0.001	0.020	0.060	0.069	0.052
800	4	0.4	0.0	0.3	0.002	0.024	0.078	0.072	−0.001	0.024	0.081	0.068	−0.001	0.020	0.059	0.059	0.000	0.020	0.061	0.063	0.055
800	4	0.4	0.6	0.0	−0.002	0.019	0.055	0.068	0.002	0.019	0.058	0.056	−0.001	0.017	0.053	0.074	0.002	0.018	0.057	0.066	0.050
800	4	0.4	0.6	0.3	−0.004	0.021	0.063	0.064	0.002	0.020	0.067	0.059	−0.002	0.018	0.054	0.060	0.001	0.018	0.055	0.065	0.046
800	4	0.8	0.0	0.0	−0.002	0.016	0.053	0.058	0.000	0.015	0.047	0.050	−0.001	0.016	0.048	0.067	0.000	0.013	0.042	0.056	0.050
800	4	0.8	0.0	0.3	−0.002	0.017	0.055	0.056	0.001	0.017	0.053	0.053	−0.002	0.015	0.048	0.058	0.001	0.015	0.047	0.052	0.051
800	4	0.8	0.6	0.0	−0.004	0.015	0.051	0.071	0.000	0.016	0.049	0.058	−0.003	0.014	0.047	0.077	0.001	0.015	0.046	0.059	0.048
800	4	0.8	0.6	0.3	−0.004	0.016	0.052	0.069	0.002	0.016	0.050	0.059	−0.002	0.015	0.047	0.066	0.000	0.015	0.046	0.058	0.049
800	8	0.4	0.0	0.0	0.002	0.013	0.038	0.056	−0.003	0.017	0.050	0.066	0.000	0.008	0.025	0.079	0.000	0.010	0.031	0.081	0.050
800	8	0.4	0.0	0.3	0.005	0.018	0.055	0.063	−0.007	0.019	0.055	0.064	0.000	0.010	0.030	0.080	0.000	0.010	0.031	0.083	0.047
800	8	0.4	0.6	0.0	−0.001	0.011	0.031	0.054	0.000	0.012	0.035	0.055	−0.001	0.009	0.026	0.078	0.001	0.010	0.030	0.080	0.055
800	8	0.4	0.6	0.3	−0.001	0.013	0.039	0.054	0.000	0.013	0.038	0.052	−0.001	0.010	0.029	0.078	0.001	0.010	0.030	0.077	0.051
800	8	0.8	0.0	0.0	−0.001	0.008	0.026	0.049	0.000	0.010	0.030	0.059	0.000	0.007	0.021	0.077	0.000	0.008	0.024	0.080	0.050
800	8	0.8	0.0	0.3	0.000	0.011	0.034	0.050	0.001	0.011	0.034	0.056	0.000	0.008	0.024	0.065	0.000	0.008	0.025	0.079	0.052
800	8	0.8	0.6	0.0	−0.001	0.008	0.025	0.050	−0.001	0.009	0.028	0.057	−0.001	0.007	0.021	0.084	0.000	0.008	0.024	0.079	0.051
800	8	0.8	0.6	0.3	−0.001	0.009	0.029	0.056	0.000	0.010	0.031	0.053	0.000	0.007	0.024	0.076	0.000	0.008	0.025	0.073	0.059

See Table B.1.

Table B.5. FIVR estimator of Robertson and Sarafidis (2015).

Designs					GMM 1 step								GMM 2 step								J
					α				β				α				β				
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	0.001	0.019	0.058	0.068	−0.002	0.020	0.060	0.058	0.000	0.016	0.047	0.081	−0.001	0.018	0.052	0.081	0.035
200	4	0.4	0.0	0.3	0.008	0.037	0.113	0.081	−0.006	0.038	0.122	0.071	−0.002	0.027	0.083	0.081	−0.001	0.030	0.090	0.080	0.033
200	4	0.4	0.6	0.0	0.000	0.019	0.057	0.061	0.000	0.019	0.055	0.046	0.000	0.016	0.048	0.081	0.000	0.017	0.051	0.073	0.031
200	4	0.4	0.6	0.3	−0.002	0.031	0.095	0.062	0.003	0.034	0.106	0.065	−0.001	0.026	0.079	0.068	0.000	0.029	0.088	0.077	0.032
200	4	0.8	0.0	0.0	0.001	0.017	0.055	0.066	0.000	0.012	0.038	0.063	0.000	0.014	0.044	0.072	0.000	0.011	0.035	0.085	0.035
200	4	0.8	0.0	0.3	0.000	0.023	0.073	0.061	0.002	0.024	0.076	0.057	0.000	0.021	0.061	0.067	0.000	0.022	0.067	0.082	0.039
200	4	0.8	0.6	0.0	−0.001	0.018	0.054	0.059	0.000	0.012	0.037	0.060	0.000	0.014	0.044	0.068	0.000	0.011	0.035	0.086	0.038
200	4	0.8	0.6	0.3	−0.001	0.023	0.071	0.062	0.002	0.024	0.076	0.066	0.000	0.021	0.062	0.071	0.000	0.022	0.072	0.084	0.041
200	8	0.4	0.0	0.0	0.001	0.012	0.037	0.069	−0.002	0.013	0.039	0.068	0.001	0.011	0.031	0.181	−0.001	0.011	0.033	0.172	0.043
200	8	0.4	0.0	0.3	0.015	0.034	0.095	0.086	−0.017	0.036	0.099	0.087	0.005	0.020	0.057	0.214	−0.006	0.021	0.061	0.215	0.043
200	8	0.4	0.6	0.0	0.000	0.012	0.036	0.067	−0.001	0.011	0.033	0.062	0.001	0.011	0.032	0.189	0.000	0.011	0.032	0.163	0.040
200	8	0.4	0.6	0.3	−0.002	0.025	0.077	0.054	0.001	0.027	0.080	0.051	−0.001	0.018	0.055	0.197	0.001	0.020	0.060	0.186	0.038
200	8	0.8	0.0	0.0	0.000	0.011	0.032	0.054	0.000	0.008	0.023	0.051	0.001	0.009	0.028	0.179	0.000	0.007	0.022	0.155	0.037
200	8	0.8	0.0	0.3	0.000	0.019	0.057	0.047	0.000	0.022	0.066	0.045	0.001	0.015	0.046	0.183	0.000	0.018	0.054	0.174	0.037
200	8	0.8	0.6	0.0	0.000	0.011	0.031	0.054	0.000	0.007	0.022	0.051	0.001	0.009	0.028	0.181	0.000	0.007	0.022	0.159	0.036
200	8	0.8	0.6	0.3	−0.003	0.018	0.055	0.051	0.004	0.022	0.066	0.046	−0.001	0.016	0.047	0.176	0.002	0.018	0.056	0.177	0.038
800	4	0.4	0.0	0.0	−0.001	0.015	0.045	0.059	0.000	0.019	0.061	0.063	−0.001	0.012	0.036	0.066	0.001	0.016	0.049	0.066	0.051
800	4	0.4	0.0	0.3	0.000	0.021	0.064	0.068	−0.001	0.022	0.070	0.066	−0.001	0.015	0.044	0.068	0.000	0.016	0.049	0.060	0.048
800	4	0.4	0.6	0.0	−0.001	0.013	0.041	0.059	0.002	0.017	0.052	0.051	−0.001	0.012	0.037	0.062	0.001	0.015	0.048	0.056	0.051
800	4	0.4	0.6	0.3	−0.002	0.017	0.051	0.062	0.002	0.018	0.058	0.059	−0.001	0.014	0.043	0.059	0.001	0.016	0.050	0.058	0.048
800	4	0.8	0.0	0.0	−0.001	0.011	0.034	0.061	0.000	0.014	0.043	0.056	0.000	0.010	0.030	0.075	0.000	0.012	0.038	0.060	0.045
800	4	0.8	0.0	0.3	0.000	0.014	0.042	0.051	0.000	0.015	0.046	0.052	−0.001	0.011	0.035	0.062	0.000	0.013	0.040	0.059	0.047
800	4	0.8	0.6	0.0	−0.001	0.011	0.033	0.069	0.000	0.015	0.044	0.056	0.000	0.010	0.029	0.075	0.000	0.014	0.042	0.062	0.042
800	4	0.8	0.6	0.3	−0.001	0.013	0.041	0.064	0.002	0.015	0.048	0.056	0.000	0.011	0.035	0.057	0.000	0.014	0.042	0.059	0.044
800	8	0.4	0.0	0.0	0.001	0.011	0.033	0.050	−0.002	0.015	0.047	0.064	0.000	0.007	0.020	0.093	0.000	0.010	0.028	0.082	0.054
800	8	0.4	0.0	0.3	0.005	0.017	0.053	0.070	−0.006	0.018	0.056	0.073	0.000	0.008	0.026	0.082	0.000	0.010	0.028	0.082	0.054
800	8	0.4	0.6	0.0	0.000	0.009	0.026	0.051	−0.001	0.011	0.033	0.054	0.000	0.007	0.021	0.079	0.000	0.009	0.028	0.077	0.053
800	8	0.4	0.6	0.3	−0.001	0.012	0.037	0.052	0.000	0.013	0.038	0.056	0.000	0.009	0.026	0.078	0.000	0.009	0.029	0.079	0.053
800	8	0.8	0.0	0.0	0.000	0.006	0.019	0.053	0.000	0.010	0.028	0.061	0.000	0.005	0.016	0.082	0.000	0.007	0.023	0.080	0.053
800	8	0.8	0.0	0.3	0.000	0.010	0.029	0.055	0.000	0.011	0.032	0.054	0.000	0.007	0.020	0.078	0.000	0.008	0.023	0.081	0.053
800	8	0.8	0.6	0.0	0.000	0.006	0.018	0.051	0.000	0.009	0.028	0.057	0.000	0.005	0.016	0.078	0.000	0.008	0.024	0.080	0.050
800	8	0.8	0.6	0.3	0.000	0.009	0.027	0.055	0.000	0.010	0.031	0.055	0.000	0.007	0.021	0.079	0.000	0.008	0.024	0.079	0.049

See Table B.1.

Table B.6. Projection GMM estimator of Hayakawa (2012) with weak exogeneity.

Designs					GMM 1 step								GMM 2 step								J
					α				β				α				β				
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Size
200	4	0.4	0.0	0.0	0.000	0.025	0.076	0.058	−0.001	0.023	0.075	0.053	−0.002	0.023	0.072	0.087	0.002	0.023	0.072	0.074	0.020
200	4	0.4	0.0	0.3	0.003	0.056	0.181	0.078	−0.003	0.054	0.172	0.083	−0.011	0.055	0.171	0.113	0.007	0.050	0.166	0.113	0.026
200	4	0.4	0.6	0.0	−0.001	0.026	0.081	0.077	0.003	0.021	0.070	0.062	−0.003	0.026	0.081	0.106	0.005	0.022	0.073	0.097	0.028
200	4	0.4	0.6	0.3	−0.016	0.055	0.191	0.106	0.015	0.052	0.191	0.097	−0.019	0.056	0.206	0.153	0.016	0.050	0.199	0.141	0.021
200	4	0.8	0.0	0.0	−0.001	0.033	0.107	0.073	0.001	0.016	0.050	0.047	−0.001	0.031	0.101	0.092	0.002	0.015	0.046	0.062	0.020
200	4	0.8	0.0	0.3	−0.009	0.050	0.179	0.088	0.002	0.034	0.116	0.079	−0.013	0.052	0.179	0.132	0.004	0.035	0.117	0.111	0.033
200	4	0.8	0.6	0.0	−0.003	0.032	0.104	0.069	0.003	0.015	0.050	0.053	−0.005	0.032	0.108	0.108	0.004	0.015	0.051	0.088	0.021
200	4	0.8	0.6	0.3	−0.013	0.056	0.212	0.106	0.009	0.041	0.142	0.084	−0.018	0.059	0.253	0.167	0.010	0.041	0.150	0.122	0.025
200	8	0.4	0.0	0.0	0.001	0.015	0.046	0.075	−0.001	0.014	0.046	0.075	0.000	0.013	0.039	0.143	0.000	0.012	0.038	0.131	0.018
200	8	0.4	0.0	0.3	0.015	0.045	0.134	0.089	−0.015	0.044	0.135	0.099	0.002	0.031	0.093	0.147	−0.002	0.031	0.092	0.145	0.021
200	8	0.4	0.6	0.0	0.000	0.014	0.044	0.063	0.001	0.012	0.037	0.056	0.000	0.014	0.042	0.144	0.001	0.012	0.035	0.118	0.028
200	8	0.4	0.6	0.3	−0.008	0.038	0.120	0.066	0.008	0.038	0.118	0.051	−0.006	0.031	0.089	0.136	0.006	0.030	0.089	0.128	0.029
200	8	0.8	0.0	0.0	−0.001	0.016	0.050	0.059	0.001	0.009	0.028	0.068	−0.001	0.016	0.046	0.140	0.001	0.008	0.026	0.129	0.021
200	8	0.8	0.0	0.3	−0.001	0.033	0.104	0.052	0.002	0.030	0.094	0.056	−0.004	0.031	0.090	0.128	0.004	0.027	0.081	0.123	0.019
200	8	0.8	0.6	0.0	−0.001	0.015	0.046	0.045	0.000	0.009	0.025	0.058	−0.002	0.015	0.047	0.136	0.001	0.008	0.025	0.118	0.026
200	8	0.8	0.6	0.3	−0.010	0.041	0.125	0.059	0.007	0.038	0.121	0.059	−0.009	0.035	0.106	0.135	0.008	0.033	0.100	0.138	0.026
800	4	0.4	0.0	0.0	−0.001	0.026	0.074	0.065	0.000	0.026	0.082	0.079	−0.001	0.021	0.064	0.072	0.001	0.023	0.069	0.073	0.035
800	4	0.4	0.0	0.3	0.000	0.032	0.105	0.072	−0.001	0.031	0.102	0.075	−0.003	0.028	0.090	0.079	0.003	0.027	0.088	0.074	0.037
800	4	0.4	0.6	0.0	−0.004	0.024	0.079	0.086	0.004	0.022	0.072	0.069	−0.004	0.024	0.077	0.103	0.004	0.022	0.074	0.095	0.038
800	4	0.4	0.6	0.3	−0.006	0.031	0.107	0.081	0.006	0.030	0.108	0.072	−0.005	0.027	0.089	0.078	0.004	0.025	0.089	0.075	0.042
800	4	0.8	0.0	0.0	−0.006	0.037	0.113	0.066	0.001	0.019	0.057	0.054	−0.007	0.036	0.108	0.098	0.002	0.017	0.052	0.059	0.036
800	4	0.8	0.0	0.3	−0.003	0.028	0.094	0.067	0.001	0.021	0.067	0.060	−0.004	0.025	0.088	0.074	0.001	0.020	0.063	0.066	0.044
800	4	0.8	0.6	0.0	−0.007	0.028	0.105	0.081	0.004	0.021	0.069	0.071	−0.008	0.028	0.105	0.096	0.003	0.020	0.067	0.084	0.045
800	4	0.8	0.6	0.3	−0.007	0.031	0.115	0.077	0.005	0.026	0.091	0.064	−0.007	0.030	0.106	0.085	0.005	0.024	0.083	0.072	0.043
800	8	0.4	0.0	0.0	0.003	0.018	0.053	0.105	−0.003	0.022	0.065	0.118	0.000	0.011	0.032	0.082	0.000	0.012	0.036	0.077	0.030
800	8	0.4	0.0	0.3	0.008	0.025	0.073	0.072	−0.007	0.025	0.072	0.074	0.000	0.016	0.048	0.076	0.000	0.016	0.047	0.073	0.027
800	8	0.4	0.6	0.0	−0.001	0.014	0.041	0.060	0.000	0.013	0.040	0.055	−0.001	0.012	0.037	0.084	0.001	0.012	0.035	0.079	0.042
800	8	0.4	0.6	0.3	−0.003	0.022	0.068	0.054	0.003	0.022	0.068	0.056	−0.001	0.017	0.047	0.067	0.001	0.016	0.047	0.073	0.035
800	8	0.8	0.0	0.0	−0.001	0.019	0.056	0.068	0.000	0.013	0.039	0.079	−0.002	0.015	0.046	0.087	0.001	0.010	0.029	0.075	0.031
800	8	0.8	0.0	0.3	0.000	0.018	0.055	0.057	0.000	0.016	0.048	0.057	−0.001	0.014	0.042	0.080	0.001	0.012	0.036	0.079	0.037
800	8	0.8	0.6	0.0	0.000	0.015	0.048	0.046	0.001	0.011	0.033	0.055	−0.001	0.014	0.046	0.083	0.001	0.010	0.030	0.073	0.040
800	8	0.8	0.6	0.3	−0.002	0.020	0.062	0.053	0.001	0.019	0.058	0.055	−0.001	0.015	0.047	0.071	0.000	0.015	0.043	0.069	0.041

See Table B.1.

Table B.7. Conditional likelihood estimator of Bai (2013b).

Designs					Strict								Weak							
					α				β				α				β			
N	T	α	ρ	δ	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size	Bias	RMedSE	qStd	Size
200	4	0.4	0.0	0.0	0.001	0.013	0.040	0.052	−0.001	0.013	0.038	0.050	−0.001	0.013	0.039	0.059	0.002	0.013	0.036	0.066
200	4	0.4	0.0	0.3	0.003	0.027	0.081	0.150	−0.015	0.031	0.103	0.207	−0.001	0.025	0.074	0.127	0.000	0.027	0.078	0.161
200	4	0.4	0.6	0.0	0.000	0.014	0.040	0.053	0.000	0.013	0.038	0.052	−0.011	0.017	0.043	0.129	0.024	0.024	0.039	0.302
200	4	0.4	0.6	0.3	0.000	0.025	0.074	0.109	−0.006	0.029	0.090	0.167	−0.040	0.042	0.081	0.350	0.050	0.051	0.078	0.445
200	4	0.8	0.0	0.0	0.000	0.013	0.040	0.054	0.000	0.009	0.026	0.059	0.000	0.013	0.039	0.052	−0.002	0.009	0.025	0.069
200	4	0.8	0.0	0.3	−0.005	0.026	0.225	0.234	−0.016	0.030	0.134	0.313	0.000	0.019	0.058	0.093	0.000	0.020	0.060	0.142
200	4	0.8	0.6	0.0	0.000	0.013	0.039	0.048	0.000	0.009	0.027	0.059	−0.005	0.014	0.041	0.066	0.011	0.012	0.026	0.166
200	4	0.8	0.6	0.3	−0.003	0.022	0.075	0.162	−0.002	0.026	0.082	0.194	−0.025	0.028	0.060	0.205	0.035	0.035	0.055	0.347
200	8	0.4	0.0	0.0	0.000	0.008	0.024	0.056	0.000	0.008	0.024	0.051	−0.001	0.008	0.024	0.053	0.001	0.008	0.024	0.064
200	8	0.4	0.0	0.3	0.005	0.015	0.045	0.086	−0.005	0.016	0.047	0.096	0.001	0.016	0.049	0.120	−0.003	0.018	0.053	0.144
200	8	0.4	0.6	0.0	0.000	0.009	0.025	0.059	0.000	0.008	0.024	0.057	−0.006	0.010	0.025	0.088	0.012	0.013	0.025	0.190
200	8	0.4	0.6	0.3	0.003	0.015	0.044	0.076	−0.003	0.016	0.047	0.090	−0.023	0.025	0.054	0.290	0.027	0.028	0.057	0.328
200	8	0.8	0.0	0.0	0.000	0.008	0.024	0.053	0.000	0.006	0.017	0.062	0.000	0.008	0.024	0.050	−0.001	0.006	0.018	0.064
200	8	0.8	0.0	0.3	−0.008	0.015	0.044	0.131	0.007	0.018	0.054	0.155	0.000	0.015	0.047	0.148	−0.001	0.019	0.057	0.179
200	8	0.8	0.6	0.0	0.000	0.008	0.024	0.052	0.000	0.006	0.017	0.059	−0.003	0.008	0.024	0.065	0.006	0.007	0.017	0.122
200	8	0.8	0.6	0.3	−0.009	0.015	0.042	0.128	0.011	0.019	0.052	0.150	−0.021	0.022	0.050	0.256	0.027	0.029	0.057	0.332
800	4	0.4	0.0	0.0	0.000	0.010	0.031	0.060	0.001	0.012	0.035	0.051	−0.003	0.011	0.033	0.095	0.004	0.015	0.043	0.172
800	4	0.4	0.0	0.3	0.002	0.022	0.072	0.339	−0.014	0.028	0.116	0.438	0.001	0.020	0.061	0.301	−0.002	0.025	0.078	0.415
800	4	0.4	0.6	0.0	0.000	0.010	0.031	0.057	0.001	0.012	0.035	0.052	−0.025	0.026	0.051	0.449	0.064	0.064	0.076	0.798
800	4	0.4	0.6	0.3	−0.002	0.021	0.063	0.297	−0.003	0.028	0.099	0.409	−0.044	0.044	0.073	0.642	0.056	0.056	0.074	0.741
800	4	0.8	0.0	0.0	−0.001	0.009	0.027	0.057	0.000	0.010	0.030	0.049	0.000	0.009	0.027	0.058	−0.006	0.012	0.038	0.182
800	4	0.8	0.0	0.3	−0.008	0.024	0.250	0.448	−0.019	0.035	0.170	0.578	−0.002	0.016	0.049	0.263	0.001	0.022	0.067	0.411
800	4	0.8	0.6	0.0	0.000	0.009	0.027	0.055	0.000	0.010	0.032	0.052	−0.007	0.011	0.030	0.134	0.040	0.040	0.045	0.722
800	4	0.8	0.6	0.3	−0.008	0.022	0.077	0.388	0.005	0.031	0.110	0.516	−0.034	0.034	0.049	0.616	0.049	0.049	0.055	0.779
800	8	0.4	0.0	0.0	0.000	0.006	0.018	0.058	0.000	0.008	0.023	0.056	−0.001	0.007	0.020	0.081	0.002	0.010	0.029	0.154
800	8	0.4	0.0	0.3	0.005	0.011	0.030	0.211	−0.006	0.012	0.035	0.241	0.002	0.013	0.039	0.314	−0.004	0.016	0.048	0.385
800	8	0.4	0.6	0.0	−0.001	0.006	0.019	0.054	0.000	0.008	0.023	0.050	−0.014	0.015	0.025	0.403	0.035	0.035	0.044	0.708
800	8	0.4	0.6	0.3	0.003	0.010	0.029	0.175	−0.003	0.012	0.035	0.224	−0.026	0.026	0.047	0.586	0.030	0.031	0.054	0.629
800	8	0.8	0.0	0.0	0.000	0.005	0.015	0.050	0.000	0.007	0.019	0.058	0.000	0.005	0.015	0.052	−0.004	0.008	0.024	0.163
800	8	0.8	0.0	0.3	−0.006	0.009	0.024	0.212	0.007	0.012	0.035	0.301	0.000	0.010	0.031	0.270	−0.001	0.014	0.041	0.369
800	8	0.8	0.6	0.0	0.000	0.005	0.015	0.046	0.000	0.007	0.020	0.057	−0.004	0.006	0.016	0.118	0.021	0.021	0.031	0.553
800	8	0.8	0.6	0.3	−0.007	0.010	0.023	0.214	0.010	0.013	0.033	0.323	−0.020	0.020	0.032	0.568	0.026	0.027	0.040	0.655

See Table B.1.

Acknowledgments

This article greatly benefited from comments made by Esfandiar Maasoumi (Editor), an Associate Editor and by two anonymous referees. We would also like to thank seminar participants at Monash University, the University of Amsterdam and the University of South Australia for useful comments. Part of this article was written while the first author enjoyed the hospitality of the Department of Econometrics and Business Statistics at Monash University.

Funding

Financial support from the NWO MaGW grant "Likelihood-Based Inference in Dynamic Panel Data Models with Endogenous Covariates" is gratefully acknowledged by the first author.

References

Abadir, K. M., Magnus, J. R. (2002). Notation in econometrics: A proposal for a standard. *Econometrics Journal* 5:76–90.

Abrevaya, J. (2013). The projection approach for unbalanced panel data. *The Econometrics Journal* 16:161–178.

Ahn, S. C. (2015). Comment on "IV estimation of panels with factor residuals" by D. Robertson and V. Sarafidis. *Journal of Econometrics* 185:542–544.

Ahn, S. C., Lee, Y. H., Schmidt, P. (2001). Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101:219–255.

Ahn, S. C., Lee, Y. H., Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174:1–14.

Alvarez, J., Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71(4):1121–1159.

Anderson, D. (1999). The aggregate burden of crime. *Journal of Law and Economics* 42:611–642.

Anderson, T. W., Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18:47–82.

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* 58:277–297.

Bai, J. (2013a). Fixed-effects dynamic panel models, a factor analytical method. *Econometrica* 81:285–314.

Bai, J. (2013b). Likelihood approach to dynamic panel models with interactive effects. Working Paper.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 76:169–217.

Blundell, R. W., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87:115–143.

Bond, S., Windmeijer, F. (2002). Projection estimators for autoregressive panel data models. *The Econometrics Journal* 5:457–479.

Bun, M. J. G., Carree, M. A., Juodis, A. (2016). On maximum likelihood estimation of dynamic panel data models. *Journal of Business and Economic Statistics*. Forthcoming.

Bun, M. J. G., Kiviet, J. F. (2006). The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models. *Journal of Econometrics* 132:409–444.

Bun, M. J. G., Poldermans, R. W. (2015). Weak identification robust inference in dynamic panel data models. Mimeo.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18:5–46.

Chudik, A., Pesaran, M. H., Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal* 14:C45–C90.

Doornik, J. (2009). *An Object-Oriented Matrix Language Ox 6*. London: Timberlake Consultants Press.

Ehrlich, I. (1975). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy* 81:521–565.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator for panel ar(p) models when both n and t are large. *Econometric Theory* 25:873–890.

Hayakawa, K. (2012). Gmm estimation of short dynamic panel data model with interactive fixed effects. *Journal of the Japan Statistical Society* 42:109–123.

Hayakawa, K. (2016). On the effect of weighting matrix in gmm specification test. *Journal of Statistical Planning and Inference*. Forthcoming.

Holtz-Eakin, D., Newey, W. K., Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56:1371–1395.

Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109:107–150.

Juodis, A. (2016a). First difference transformation in panel var models: Robustness, estimation and inference. *Econometric Reviews*. Forthcoming.

Juodis, A. (2016b). Pseudo panel data models with cohort interactive effects. *Journal of Business & Economic Statistics*. Forthcoming.

Juodis, A., Sarafidis, V. (2015). A simple estimator for short panels with common factors. UvA-Econometrics working paper 2015/03.

Kruiniger, H. (2008). Not so fixed effects: Correlated structural breaks in panel data. Unpublished manuscript, Queen Marry, University of London.

Kruiniger, H. (2013). Quasi ml estimation of the panel ar(1) model with arbitrary initial conditions. *Journal of Econometrics* 173:175–188.

Magnus, J. R., Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester–New-York–Weinheim–Brisbane–Singapore–Toronto: John Wiley & Sons.

McDonald, T., Morling, S. (2011). The australian economy and the global downturn: Reasons for resilience. *Economic Roundup: Treasury Report*, 2.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* 46:69–85.

Nauges, C., Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique* 70:54–75.

Robertson, D., Sarafidis, V. (2015). Iv estimation of panels with factor residuals. *Journal of Econometrics* 185:526–541.

Robertson, D., Sarafidis, V., Westerlund, J. (2014). Gmm unit root inference in generally trending and cross-correlated dynamic panels. Working Paper.

Sarafidis, V. (2016). Neighborhood gmm estimation of dynamic panel data models. *Computational Statistics and Data Analysis* 100:526–544.

Sarafidis, V., Robertson, D. (2009). On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal* 12:62–81.

Sarafidis, V., Wansbeek, T. J. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31:483–531.

Sarafidis, V., Yamagata, T., Robertson, D. (2009). A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* 148:149–161.

Westerlund, J., Norkutė, M. (2014). A factor analytical method to interactive effects dynamic panel models with or without unit root. Working Paper 2014:12.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics* 126:25–51.