Pseudo Panel Data Models With Cohort Interactive Effects

Juodis, Arturas

Published in:
Journal of Business & Economic Statistics

DOI:
10.1080/07350015.2015.1137759

Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

Pseudo Panel Data Models With Cohort Interactive Effects

Artūras Juodis

To cite this article: Artūras Juodis (2018) Pseudo Panel Data Models With Cohort Interactive Effects, Journal of Business & Economic Statistics, 36:1, 47-61, DOI: 10.1080/07350015.2015.1137759

To link to this article: https://doi.org/10.1080/07350015.2015.1137759

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Pseudo Panel Data Models With Cohort Interactive Effects

Artūras Juodis

Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands (*a.juodis@rug.nl*)

When genuine panel data samples are not available, repeated cross-sectional surveys can be used to form so-called pseudo panels. In this article, we investigate the properties of linear pseudo panel data estimators with fixed number of cohorts and time observations. We extend standard linear pseudo panel data setup to models with factor residuals by adapting the quasi-differencing approach developed for genuine panels. In a Monte Carlo study, we find that the proposed procedure has good finite sample properties in situations with endogeneity, cohort interactive effects, and near nonidentification. Finally, as an illustration the proposed method is applied to data from Ecuador to study labor supply elasticity. Supplementary materials for this article are available online.

KEY WORDS: Cohort interactive effects; Labor supply elasticity; Pseudo panel data; Weak identification.

1. INTRODUCTION

Over the last three decades, panel data techniques proved to be of high value for both micro and macro economists. Nevertheless, genuine microeconomic panel data can still be difficult and costly to obtain and administer. The nonavailability of genuine panel datasets can be especially problematic for developing countries with a limited amount of administrative data that tracks individuals over time. In such cases, repeated cross-section surveys can be used to form so-called pseudo panels.

Models for this type of data in economics were introduced by Deaton (1985), with early contributions by Verbeek and Nijman (1992) and Moffitt (1993) among others. Although pseudo panel data models have not been analyzed as extensively as their genuine counterparts, the volume of literature on these types of models is increasing. For some recent theoretical articles and reviews, readers may be referred to McKenzie (2004), Verbeek and Vella (2005), Inoue (2008), and Verbeek (2008).

Existing estimation methods for linear pseudo panel data models assume that the unobserved individual heterogeneity can be properly captured using the standard additive error component structure. However, in some cases this assumption might be too restrictive to properly describe the data at hand. For genuine panel data models, there is a substantial literature available on models that use a multiplicative error component structure of the factor type to capture the unobserved individual characteristics in a more flexible way, see, for example, Pesaran (2006), Bai (2009), Sarafidis, Yamagata, and Robertson (2009), and the survey of Sarafidis and Wansbeek (2012).

The key component of pseudo panel analysis is the use of cohort-based data in estimation. We use the generic term "cohorts" to describe any grouping structure based on variables like gender, race, or age. In this article, we introduce a factor structure to linear pseudo panel data models with a fixed number of time periods and cohorts. We provide several theoretical contributions to the existing pseudo panel data literature. First, we propose a generalized method of moments (GMM) estimator based on the quasi-differencing (QD) approach of Ahn, Lee, and Schmidt (2013). Second, we discuss identification, estimation, and inference properties of this estimator for potentially unbalanced samples. In addition to the theoretical results of the novel estimator, an extensive Monte Carlo simulation study is conducted to assess the finite sample properties. We focus on the robustness of the proposed estimator with respect to endogenous variables, cohort interactive effects, and weak identification.

As an empirical illustration for our method, we apply our estimator in an analysis of the labor supply elasticity in Ecuador over the period of 2007–2013. We use annual survey data to construct 10 cohorts based on the corresponding heads of the household that work full time. To account for possible general nonlinear trends in labor supply, we allow for a nonadditive factor structure using the newly developed estimator.

Here we briefly introduce our notation. The usual $\text{vec}(\cdot)$ operator denotes the column stacking operator. The commutation matrix $\mathbf{K}_{ab}$ is defined in such a way that for any $[a \times b]$ matrix $\boldsymbol{A}$, $\text{vec}(\boldsymbol{A}') = \mathbf{K}_{ab}\,\text{vec}(\boldsymbol{A})$. $\otimes$ denotes the Kronecker product satisfying the property $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\,\text{vec}(\boldsymbol{B})$ and $\boldsymbol{\iota}_T$ is a $[T \times 1]$ vector of ones. For some set $\mathcal{A}$, we denote its cardinality by $|\mathcal{A}|$. Finally, $1_{(\cdot)}$ is the usual indicator function. For further details regarding the notation used in this article see Abadir and Magnus (2002).

2. MODEL

In this article, we consider a linear panel data model with group-specific membership

$$y_{i,t} = \boldsymbol{\beta}'\boldsymbol{w}_{s,t} + \boldsymbol{\zeta}'z_{i,t} + \nu_{i,t}, \quad \nu_{i,t} = \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t},$$
$$\text{E}[\varepsilon_{i,t}] = 0, \quad i \in \mathcal{I}_{s,t}, \tag{2.1}$$

where $\mathcal{I}_{s,t}$ is the set of all individuals (in total $N_{s,t}$) that are in group $s = 1, \ldots, S$ at time $t = 1, \ldots, T$, $\boldsymbol{w}_{s,t}$ is a

Journal of Business & Economic Statistics
January 2018, Vol. 36, No. 1
DOI: 10.1080/07350015.2015.1137759
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

K_w-dimensional vector of group-time-specific covariates, and $z_{i,t}$ is a K_z-dimensional vector of individual-specific explanatory variables. Thus, in total there are $K_z + K_w = K$ parameters of interest for observed explanatory variables. We denote the combined parameter vector by $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\zeta}')'$. For ease of exposition we shall assume at this stage that $z_{i,t}$ does not contain any lags of $y_{i,t}$. The extension of (2.1) to dynamic models is discussed in the supplementary material.

We use the generic term cohorts to describe any grouping structure based on some selection variable. In the literature variables like gender, race, region of residence, and most popularly age are used to define the group participation, see Verbeek (2008) and McKenzie (2004).

To allow for individual-specific unobserved characteristics, $u_{i,t}$ contains the multifactor error term $\boldsymbol{\lambda}_i' \boldsymbol{f}_t = \sum_{l=1}^{L} \lambda_i^{(l)} f_t^{(l)}$. The L-dimensional vectors $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ are individual-specific factor loadings and time-specific factors, respectively. The standard two-component (fixed effects (FE)) model (as in, e.g., McKenzie 2001, 2004, Inoue 2008, and Verbeek 2008) can be obtained by setting $\boldsymbol{f}_t$ to some constant $\boldsymbol{c}$, such that $\boldsymbol{\lambda}_i' \boldsymbol{f}_t = \delta_i, \quad \forall t$.

Estimation of the model in (2.1) is straightforward if $\mathrm{E}[v_{i,t}|z_{i,t}] = 0$ and can be performed by using pooled cross-sectional ordinary least squares (OLS). However, in most cases of empirical interest these conditions can be violated as the unobserved individual characteristics $\boldsymbol{\lambda}_i$ are correlated with observed individual characteristics $z_{i,t}$. Hence, if the correlation is nonzero we have to rely on either general "external" instruments or pseudo panel techniques that use the cohort structure of the dataset as instruments, as originally suggested by Deaton (1985). This article deals with the latter type of estimators.

Before defining the estimators considered in this article, we discuss the notation first. All estimators discussed in this article can be expressed solely in terms of the matrices/vectors-containing cross-sectional averages. Observations at the individual level i, on the other hand, are only used for the estimation of the asymptotic variance-covariance matrices. By taking the cross-sectional average for some group s at time t, we obtain the following aggregated equation

$$\bar{y}_{s,t} = \boldsymbol{\theta}' \bar{\boldsymbol{x}}_{s,t} + \bar{v}_{s,t}, \quad s = 1, \ldots, S, \quad t = 1, \ldots, T. \tag{2.2}$$

Here, we denote

$$\bar{y}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{I}_{s,t}} y_{i,t}, \quad \bar{\boldsymbol{x}}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{I}_{s,t}} \boldsymbol{x}_{i,s,t},$$

$$\bar{\boldsymbol{\lambda}}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{I}_{s,t}} \boldsymbol{\lambda}_i, \quad \bar{\varepsilon}_{s,t} = \frac{1}{N_{s,t}} \sum_{i \in \mathcal{I}_{s,t}} \varepsilon_{i,t},$$

$$\bar{v}_{s,t} = \bar{\boldsymbol{\lambda}}_{s,t}' \boldsymbol{f}_t + \bar{\varepsilon}_{s,t}, \quad \boldsymbol{x}_{i,s,t} = (\boldsymbol{w}_{s,t}', z_{i,t}')'.$$

After performing cross-sectional averaging, we stack all observations over the time-dimension for some cohort s

$$\boldsymbol{y}_s = \boldsymbol{X}_s \boldsymbol{\theta} + \boldsymbol{v}_s, \tag{2.3}$$

where $\boldsymbol{X}_s = (\bar{\boldsymbol{x}}_{s,1}, \ldots, \bar{\boldsymbol{x}}_{s,T})'$, a $[T \times K]$-dimensional matrix, and similarly for T-dimensional vectors $\boldsymbol{y}_s$ and $\boldsymbol{u}_s$. Finally, we can stack observations for all cohorts

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{v}, \tag{2.4}$$

where the corresponding s specific vectors/matrices are stacked on top of each other, for example, $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_S')'$.

It is important that already at this point we discuss the asymptotic setup that one can use to derive the theoretical results. Using the terminology of Verbeek (2008) we formulate commonly used asymptotic schemes:

Type I. $N_{s,t} \to \infty$. T and S are fixed (this article);
Type II. $N_{s,t}$ and T fixed but $S \to \infty$;
Type III. $N_{s,t} \to \infty$ and $T \to \infty$ but S fixed.

A well-known implication (see, e.g., Inoue 2008) of the Type I asymptotics is the robustness of the estimator based on cross-sectional averages to the presence of endogenous explanatory variables. However, discussed later in this article, robustness to endogeneity is only achieved under the assumption of strong identification. Another implication for our analysis is that under Type I (unlike Type II) asymptotics, the estimator that is discussed in this article does not suffer from the "many instrument" bias as in Bekker (1994) and Bekker and van der Ploeg (2005). The intuition behind these properties is discussed later in the article.

3. COHORT INTERACTIVE EFFECTS

3.1 Inconsistency of the Conventional Fixed Effects Estimator

In this section, we show that the conventional fixed effects-type estimator for pseudo panel data models is inconsistent if $v_{i,t}$ has a factor structure. The conventional estimator GMM (or fixed effects) estimator, as in Dargay (2007) and Inoue (2008), is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l} = \left(\boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\boldsymbol{y}, \tag{3.1}$$

where we use the subscript l to emphasize that this estimator is linear. The use of $[ST \times ST]$ matrix $\boldsymbol{M} = \boldsymbol{I}_S \otimes (\boldsymbol{I}_T - (1/T)\boldsymbol{\iota}_T \boldsymbol{\iota}_T')$ can be motivated if the unobserved heterogeneity is of the form $\mathrm{E}[\boldsymbol{\lambda}_i' \boldsymbol{f}_t | i \in \mathcal{I}_{s,t}] = \delta_s$. Finally, $\boldsymbol{\Omega}$ is some prespecified $[ST \times ST]$ positive-definite weighting matrix.

This estimator remains consistent provided that $E[\boldsymbol{\lambda}_i' \boldsymbol{f}_t | i \in \mathcal{I}_{s,t}] = \delta_s$, as in this case $\boldsymbol{M}\boldsymbol{u} \xrightarrow{p} \boldsymbol{0}_{S(T-1)}$. This condition in some cases can be too restrictive as it imposes that all cohorts respond similarly to common shocks (on average). However, it can still be reasonable to maintain the less restrictive assumption that $\mathrm{E}[\boldsymbol{\lambda}_i | i \in \mathcal{I}_{s,t}] = \boldsymbol{\lambda}_s$, such that

$$\mathrm{E}[\boldsymbol{\lambda}_i' \boldsymbol{f}_t | i \in \mathcal{I}_{s,t}] = \boldsymbol{\lambda}_s' \boldsymbol{f}_t. \tag{3.2}$$

Under this assumption all individuals i in cohort s have an error-component structure with common time-varying mean, or in other words, a cohort interactive effects structure.

Before characterizing the asymptotic properties of the $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l}$ estimator under the cohort interactive effects structure in (3.2), we define

$$\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)', \quad \boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_S)'. \tag{3.3}$$

Here $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ are $[T \times L]$ and $[S \times L]$ matrices of factors and cohort factor loadings, respectively. As a special case, in the fixed effects model, both $\boldsymbol{F} = \boldsymbol{c}$ and $\boldsymbol{\Lambda} = \boldsymbol{\delta}$ are T- and S-dimensional vectors. For the more general model, we define $u_{i,t}$

to be

$$u_{i,t} \equiv \nu_{i,t} - \boldsymbol{\lambda}_s' \boldsymbol{f}_t, \quad i \in \mathcal{I}_{s,t}, \tag{3.4}$$

such that the newly combined error term has mean zero, that is, $\mathrm{E}[u_{i,t}] = 0$, $i \in \mathcal{I}_{s,t}$. Using this notation we state formally the assumptions we impose on the error terms $u_{i,t}$.

(A.1) $N_{s,t} \to \infty$, $\forall s, t$; $\exists N \to \infty$ s.t. $N_{s,t}/N \to \pi_{s,t}$, and $0 < \min \pi_{s,t} < \max \pi_{s,t} < \infty$. T and S are fixed (Type I asymptotics).

(A.2) $u_{i,t}$ are i.h.d. with finite $2+\delta$ moment, for $\delta > 0$, such that $\sqrt{N_{s,t}}\bar{u}_{s,t} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{s,t})$ jointly $\forall s, t$ with $0 < \min \sigma^2_{s,t} \leq \max \sigma^2_{s,t} < \infty$.

Assumption (A.1) states that the number of individuals per cohort at any time t should be large and asymptotically nonnegligible as compared to N, while the number of cohorts and time periods is fixed.

Remark 1. Note that in (A.1) instead of explicitly assuming that $N = \sum_{t=1}^{T} \sum_{s=1}^{S} N_{s,t}$ we allow for some generic N. The estimators and the test statistics considered in this article are invariant to a particular choice of N. In general, one can think of N to be the sum (as in Inoue 2008), average or even any particular value of $N_{s,t}$ (as in McKenzie 2004).

In Assumption (A.2), unlike Inoue (2008), we do not impose the iid assumption, but allow for heteroscedasticity between individuals and over time. Furthermore, this assumption can be relaxed by allowing a certain degree of spatial dependence between individuals of the same cohort. In that case consistency (or inconsistency) properties of all estimators discussed in this article are not altered, but knowledge about the exact structure of the spatial dependence is required for correct inference.

Similarly to the model with cohort fixed effects, one can rewrite the stacked equation for $\boldsymbol{y}$ as

$$\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{\nu} = \mathrm{vec}\,(\boldsymbol{F\Lambda}') + \boldsymbol{X\theta} + \boldsymbol{u}. \tag{3.5}$$

Now define the probability limit of the regressors matrix $\boldsymbol{X}$ as $\boldsymbol{X}_\infty$, that is, $\boldsymbol{X}_\infty \equiv \mathrm{plim}_{N\to\infty} \boldsymbol{X}$. Here, the regressor matrix $\boldsymbol{X}_\infty = (\boldsymbol{W}, \boldsymbol{Z}_\infty)$, has a typical st'th row element given by $(\boldsymbol{w}'_{s,t}, \lim_{N\to\infty}(1/N_{s,t}) \sum_{i=1}^{N_{s,t}} \mathrm{E}[z'_{i,t}|i \in \mathcal{I}_{s,t}])$. If the $z_{i,t}$ are iid for all $i \in \mathcal{I}_{s,t}$ the st'th row is simply given by $(\boldsymbol{w}'_{s,t}, \mathrm{E}[z'_{i,t}|i \in \mathcal{I}_{s,t}])$.

In this case under high-level Assumptions (A.1)–(A.2), the following result applies to the $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l}$ estimator.

Proposition 1. If (3.2) holds, $\boldsymbol{X}_\infty$ and $\boldsymbol{F\Lambda}'$ are deterministic, then under Assumptions (A.1)–(A.2)

$$\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l} - \boldsymbol{\theta}_0 \xrightarrow{p} \left(\boldsymbol{X}'_\infty \boldsymbol{M\Omega M X}_\infty\right)^{-1} \boldsymbol{X}'_\infty \boldsymbol{M\Omega M}\,\mathrm{vec}\,(\boldsymbol{F\Lambda}').$$

Proof. In Appendix A.1. □

Thus, the GMM/FE estimator converges in probability to a value that depends on unobserved factors ($\boldsymbol{F}$) and on cohort factor loadings ($\boldsymbol{\Lambda}$). Note that, in principle, it is possible that both limiting quantities have zero mean (hence $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l}$ can be asymptotically unbiased), if one assumes $\boldsymbol{F\Lambda}'$ to be stochastic. Technical reasons behind the assumption that some of the quantities have to be deterministic are discussed later in this article in more detail.

Remark 2. One can further show that the conclusions of Proposition 1 also hold for the linear estimator that allows for additive cohort and time effects (with $\boldsymbol{M} = (\boldsymbol{I}_S - (1/S)\boldsymbol{\iota}_S\boldsymbol{\iota}_S') \otimes (\boldsymbol{I}_T - (1/T)\boldsymbol{\iota}_T\boldsymbol{\iota}_T'))$.

3.2 Assumptions and Estimation

Given that the $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}l}$ estimator is in general inconsistent in the presence of the multifactor error structure, another estimation strategy is needed to obtain consistent estimates of $\boldsymbol{\theta}$. For this purpose, we adopt the quasi-differencing (QD) approach of Ahn, Lee, and Schmidt (2001, 2013) that is tailored for genuine panel data models with fixed T. Their approach suggests the use of the transformation matrix $\boldsymbol{M}_s(\boldsymbol{\phi})$ that depends on the unknown parameter vector $\boldsymbol{\phi}$ so that (for $T > L$)

$$\boldsymbol{M}_s(\boldsymbol{\phi})\boldsymbol{F} = \mathbf{O}_{(T-L)\times L}. \tag{3.6}$$

In other words, one has to introduce the additional parameter vector $\boldsymbol{\phi}$ to remove the unobserved factors $\boldsymbol{F}$ from the model. Unlike the standard setup with fixed effects δ_s only, where the factors and consequently the corresponding transformation matrix are known (up to a constant), the $\boldsymbol{M}_s(\boldsymbol{\phi})$ matrix is unknown and depends on $\boldsymbol{\phi}$, which has to be estimated jointly with $\boldsymbol{\theta}$.

Observe that for each $[L \times L]$ invertible matrix $\boldsymbol{A}$, we have

$$\boldsymbol{F\lambda}_s = (\boldsymbol{FA})\left(\boldsymbol{A}^{-1}\boldsymbol{\lambda}_s\right) = \boldsymbol{F}^*\boldsymbol{\lambda}^*_s. \tag{3.7}$$

To avoid this rotational indeterminacy (or in other words, nonuniqueness to multiplication), we can normalize $\boldsymbol{F}^* = (\boldsymbol{\Phi}', -\boldsymbol{I}_L)'$ (assuming that the lower $[L \times L]$ block ($\boldsymbol{F}_{L\times L}$) of $\boldsymbol{F}$ matrix is of full rank). One can then set $\boldsymbol{M}_s(\boldsymbol{\phi})$ to be

$$\boldsymbol{M}_s(\boldsymbol{\phi}) = (\boldsymbol{I}_{T-L}, \boldsymbol{\Phi}), \tag{3.8}$$

where $\boldsymbol{\phi} = \mathrm{vec}\,(\boldsymbol{\Phi})$. Analogously to the fixed effects transformation matrix, we define the stacked version of this matrix using the Kronecker product, that is, $\boldsymbol{M}(\boldsymbol{\phi}) = \boldsymbol{I}_S \otimes \boldsymbol{M}_s(\boldsymbol{\phi})$. Note that other normalization schemes are also possible, for further details please refer to the supplementary material.

Given the transformation matrix $\boldsymbol{M}(\boldsymbol{\phi})$, we define the nonlinear GMM estimator $\hat{\boldsymbol{\gamma}}_{\mathrm{GMM}n} = (\hat{\boldsymbol{\theta}}'_{\mathrm{GMM}n}, \hat{\boldsymbol{\phi}}'_{\mathrm{GMM}n})'$ as the global minimizer of the following objective function

$$f(\boldsymbol{\gamma}) = \frac{1}{2}\left[(\boldsymbol{y} - \boldsymbol{X\theta})'\boldsymbol{M}(\boldsymbol{\phi})'\boldsymbol{\Omega M}(\boldsymbol{\phi})(\boldsymbol{y} - \boldsymbol{X\theta})\right], \tag{3.9}$$

for some prespecified $[S(T-L) \times S(T-L)]$ positive definite weighting matrix $\boldsymbol{\Omega}$. The corresponding gradient of the objective function in (3.9) is given by

$$\nabla f(\boldsymbol{\gamma}) = \begin{pmatrix} -\boldsymbol{X}'\boldsymbol{M}(\boldsymbol{\phi})' \\ \boldsymbol{Q}'\left((\boldsymbol{y} - \boldsymbol{X\theta}) \otimes \boldsymbol{I}_{S(T-L)}\right) \end{pmatrix}$$
$$\times \boldsymbol{\Omega M}(\boldsymbol{\phi})(\boldsymbol{y} - \boldsymbol{X\theta}) = \begin{pmatrix} \boldsymbol{D}_\theta(\boldsymbol{\gamma})' \\ \boldsymbol{D}_\phi(\boldsymbol{\gamma})' \end{pmatrix} \boldsymbol{\Omega M}(\boldsymbol{\phi})(\boldsymbol{y} - \boldsymbol{X\theta}).$$

Here $\boldsymbol{D}_\gamma(\boldsymbol{\gamma}) = (\boldsymbol{D}_\theta(\boldsymbol{\gamma}), \boldsymbol{D}_\phi(\boldsymbol{\gamma}))$ is the Jacobian matrix of the moment conditions $\boldsymbol{M}(\boldsymbol{\phi})(\boldsymbol{y} - \boldsymbol{X\theta})$, evaluated at some $\boldsymbol{\gamma}$ (when evaluated at the true value $\boldsymbol{\gamma}_0$ we suppress the dependence on

$\boldsymbol{\gamma}$). Finally, the $\boldsymbol{Q}$ is an $[S^2T(T-L)\times(T-L)L]$ selection matrix of the following form:

$$\boldsymbol{Q}=(((\boldsymbol{I}_S\otimes\mathbf{K}_{TS})(\text{vec}\,\boldsymbol{I}_S\otimes\boldsymbol{I}_T))\otimes\boldsymbol{I}_{T-L})\boldsymbol{V},$$

$$\boldsymbol{V}=\left(\begin{pmatrix}\mathbf{O}_{(T-L)\times L}\\ \boldsymbol{I}_L\end{pmatrix}\otimes\boldsymbol{I}_{T-L}\right),$$

with zeros and ones as elements.

Before proceeding, we extend the set of the high-level assumptions that are sufficient for proving the asymptotic results for $\hat{\boldsymbol{\gamma}}_{\text{GMM}n}$.

(A.3) $\boldsymbol{\gamma}=(\boldsymbol{\theta}',\boldsymbol{\phi}')'$; $\boldsymbol{\gamma}\in\boldsymbol{\Gamma}\subset\mathbb{R}^{K+(T-L)L}$ and $\boldsymbol{\gamma}_0\in$ interior $(\boldsymbol{\Gamma})$. The parameter space $\boldsymbol{\Gamma}$ is compact.

(A.4) $\text{rk}[\text{plim}_{N\to\infty}\boldsymbol{D}_\gamma]=K+(T-L)L$. The $\boldsymbol{X}_\infty$ matrix is deterministic.

(A.5) (a) $L=L_0<\min\{S,T\}$, while $(S-L)(T-L)>K$ with L_0 being the true number of factors with nonzero mean factor loadings. (b) $\text{rk}(\boldsymbol{\Lambda})=L_0$. (c) $\boldsymbol{F}^{-1}_{L\times L}$ exists. (d) $\boldsymbol{F}$ and $\boldsymbol{\Lambda}$ are deterministic matrices.

(A.6) The model is asymptotically identified: $\text{plim}_{N\to\infty}$ $\boldsymbol{M}(\boldsymbol{\phi})(\boldsymbol{y}-\boldsymbol{X\theta})=\mathbf{0}_{S(T-L)}$ implies $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$.

The probability limit of the (transposed) Jacobian $\boldsymbol{D}'_\gamma$ in (A.4) can be expressed in the following way:

$$\text{plim}_{N\to\infty}\boldsymbol{D}'_\gamma=\begin{pmatrix}-\boldsymbol{X}'_\infty\boldsymbol{M}(\boldsymbol{\phi}_0)'\\ \boldsymbol{Q}'(\text{vec}\,(\boldsymbol{F\Lambda}')\otimes\boldsymbol{I}_{S(T-L)})\end{pmatrix}.$$

(A.4) is the strong identification assumption commonly used in the standard GMM setting. This assumption is quite restrictive even when $\boldsymbol{\phi}$ is known, see discussion in Verbeek (2008). The importance of this problem is illustrated in Section 4.2 as well as in the Monte Carlo section of this article. However, we leave the properties of the identification-robust inference procedures of, for example, Kleibergen (2005) for future research.

For genuine panel data models, Ahn, Lee, and Schmidt (2013) formulated (A.5) slightly differently: "*… denotes the number of the individual-specific effects that are correlated with the regressors…*." In our case, only factors with nonzero mean factor loadings are of interest for estimation, as factors with zero mean factor loadings cannot be identified from cross-sectional averages. Hence, it is possible that the genuine panel data estimators would identify more factors than the pseudo panel data estimator even if applied to the same dataset.

Assumption (A.6) imposes asymptotic global identification for the objective function. In Section 4.3, we provide detailed examples when this condition can be violated. Note that Assumptions (A.1)–(A.6) do not impose any exogeneity restriction of $z_{i,t}$ with respect to $u_{i,t}$, thus elements of $z_{i,t}$ are allowed to be endogenous, as noted at the end of Section 2.

Remark 3. Although possible, in this article we assume that $\boldsymbol{w}_{s,t}$ does not contain regressors that are either constant over time (e.g., cohort effects) or over cohorts (e.g., time effects). For the nonlinear presented in this article, the identification of the parameters for such variables (one can think of these regressors as "low rank," see, e.g., Moon and Weidner 2015) can be problematic. That is mainly due to the possible collinearity with time-varying unobserved factors. On the other hand, if one does not include these regressors in $\boldsymbol{w}_{s,t}$, the nonlinear estimator for θ remains consistent, but not necessarily efficient (see also the discussion in Ahn, Lee, and Schmidt 2013).

Remark 4. In this article, we treat both cohort specific variables $\boldsymbol{w}_{s,t}$ and the cohort-interactive effects component $\boldsymbol{F\Lambda}'$ as deterministic. Equivalently, Assumptions (A.1)–(A.6) can be formulated conditional on these quantities, but in that case one has to rely on limit theory developed in Kuersteiner and Prucha (2013, 2015) to obtain the limiting distribution. Our treatment of the unobserved quantities is similar to the genuine panel data models for fixed T, where $\boldsymbol{F}$ is usually treated as deterministic (as in Ahn, Lee, and Schmidt 2013; Robertson and Sarafidis 2015). The deterministic treatment of variables is only needed to avoid technicalities, without any effect on the way estimation and inference are performed (as emphasized by Kuersteiner and Prucha 2013, 2015).

Assumptions (A.1)–(A.6) are sufficient to obtain the following asymptotic representation of $\hat{\boldsymbol{\gamma}}_{\text{GMM}n}$.

Proposition 2. Suppose that Assumptions (A.1)–(A.6) are satisfied. Then $\hat{\boldsymbol{\gamma}}_{\text{GMM}n}$ has the following asymptotic representation:

$$\sqrt{N}(\hat{\boldsymbol{\gamma}}_{\text{GMM}n}-\boldsymbol{\gamma}_0)\xrightarrow{d}\text{plim}_{N\to\infty}\left(\left(\boldsymbol{D}'_\gamma\boldsymbol{\Omega}\boldsymbol{D}_\gamma\right)^{-1}\boldsymbol{D}'_\gamma\right)\times\boldsymbol{\Omega}\boldsymbol{M}(\boldsymbol{\phi}_0)\boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi},\qquad(3.10)$$

where $\boldsymbol{\xi}\sim\mathcal{N}(\mathbf{0}_{ST},\boldsymbol{I}_{ST})$ and $\boldsymbol{\Sigma}$ is an $[ST\times ST]$ diagonal matrix with the typical $(s-1)T+t$ diagonal element given by $\sigma^2_{s,t}/\pi_{s,t}$.

These results can be proved using standard arguments, for example, as in Newey and McFadden (1994).

Remark 5. Similar to the original setup of Ahn, Lee, and Schmidt (2013) for genuine panels, the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_{\text{GMM}n}$ is well defined only in the case of the true value of $L=L_0$ as imposed by Assumption (A.5).

The asymptotic variance-covariance matrix (treating $\boldsymbol{X}_\infty$, $\boldsymbol{F}$, and $\boldsymbol{\Lambda}$ as deterministic) is minimized at $\boldsymbol{\Omega}_{\text{opt}}=(\boldsymbol{M}(\boldsymbol{\phi}_0)\boldsymbol{\Sigma}\boldsymbol{M}(\boldsymbol{\phi}_0)')^{-1}$. In that case, the asymptotic variance-covariance matrix of the $\hat{\boldsymbol{\theta}}_{\text{GMM}n}$ is given by

$$\text{plim}_{N\to\infty}\left(\boldsymbol{D}'_\theta\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{M}_{\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{D}_\phi}\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{D}_\theta\right)^{-1},\qquad(3.11)$$

where $\boldsymbol{M}_{\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{D}_\phi}=\boldsymbol{I}_{S(T-L)}-\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{D}_\phi(\boldsymbol{D}'_\phi\boldsymbol{\Omega}_{\text{opt}}\boldsymbol{D}_\phi)^{-1}\boldsymbol{D}'_\phi\boldsymbol{\Omega}^{1/2}_{\text{opt}}$ is the usual "residual maker" projection matrix that projects off the column space of $\boldsymbol{\Omega}^{1/2}_{\text{opt}}\boldsymbol{D}_\phi$.

Remark 6. Depending on the assumptions made for the error terms (heteroscedasticity or homoscedasticity over time and cohorts) and regressors (static or dynamic model), the formulas for consistent estimation of $\boldsymbol{\Sigma}$ can be used without modifications as in Inoue (2008). The typical $(s-1)T+t$ diagonal element of the $\hat{\boldsymbol{\Sigma}}$ matrix is equal to $\hat{q}^2_{s,t}=(N/N_{s,t})\hat{\sigma}^2_{s,t}$, where

$$\hat{\sigma}^2_{s,t}=\frac{1}{N_{s,t}}\sum_{i\in\mathcal{I}_{s,t}}(y_{i,t}-\boldsymbol{x}'_{i,s,t}\hat{\boldsymbol{\theta}}_1)^2-\left(\frac{1}{N_{s,t}}\sum_{i\in\mathcal{I}_{s,t}}(y_{i,t}-\boldsymbol{x}'_{i,s,t}\hat{\boldsymbol{\theta}}_1)\right)^2\qquad(3.12)$$

for some consistent initial estimator $\hat{\boldsymbol{\theta}}_1$ (e.g., the estimator that uses the identity matrix for $\boldsymbol{\Omega}$).

For a fixed value of S and T, the unconditional (treating deterministic quantities as stochastic) distribution of $\sqrt{N}(\hat{\boldsymbol{\gamma}}_{\text{GMM}n} - \boldsymbol{\gamma}_0)$ is not multivariate normal and depends on factors, cohort-specific factor loadings, and cohort-specific regressors $\boldsymbol{w}_{s,t}$ in the limit. Note that the limiting distribution of the linear GMM estimator $\hat{\boldsymbol{\gamma}}_{\text{GMM}l}$ (as in Inoue 2008) is also normal only conditionally on cohort-specific regressors $\boldsymbol{w}_{s,t}$, while unconditionally it is not. Hence, the conditioning argument is not unique to the nonlinear estimator.

Note that the number of rows in the $\boldsymbol{Q}$ matrix is quadratic in both S and T. Thus, even for moderate dimensions, numerical computations might become cumbersome. Given that under Assumptions (A.1)–(A.6) the $\boldsymbol{\Sigma}$ matrix is diagonal (or block-diagonal if dynamics are allowed), we can also limit our attention to the block diagonal (over cohorts) $\boldsymbol{\Omega}$ weighting matrix. In this case, the objective function can be substantially simplified

$$f(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{s=1}^{S}(\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\theta})'\boldsymbol{M}_s(\boldsymbol{\phi})'\boldsymbol{\Omega}_s\boldsymbol{M}_s(\boldsymbol{\phi})(\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\theta}). \quad (3.13)$$

Here as before $\boldsymbol{M}_s(\boldsymbol{\phi}) = (\boldsymbol{I}_{T-L}, \boldsymbol{\Phi})$, while $\boldsymbol{\Omega}_s$ is the sth block of the block-diagonal matrix $\boldsymbol{\Omega}_s$. The gradient can be also expressed as a sum

$$\nabla f(\boldsymbol{\gamma}) = \sum_{s=1}^{S}\begin{pmatrix} -\boldsymbol{X}_s'\boldsymbol{M}_s(\boldsymbol{\phi})' \\ \boldsymbol{V}'\left((\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\theta}) \otimes \boldsymbol{I}_{T-L}\right)\end{pmatrix}\boldsymbol{\Omega}_s\boldsymbol{M}_s(\boldsymbol{\phi})(\boldsymbol{y}_s - \boldsymbol{X}_s\boldsymbol{\theta}),$$

with $\boldsymbol{V}$ as defined previously. This simplification of the objective function is used in Section 5 while conducting the Monte Carlo study as well as in the empirical exercise.

As an alternative to the quasi-differencing approach with respect to factors, one can also construct a similar estimator by quasi-differencing over cohorts. In this case, one has to look for redefined $\boldsymbol{\phi}$ s.t.

$$\boldsymbol{M}(\boldsymbol{\phi})(\boldsymbol{\Lambda} \otimes \boldsymbol{I}_T) = \boldsymbol{0}_{T(S-L)}. \quad (3.14)$$

In this case, the $\boldsymbol{\phi}$ vector is of dimension $(S-L)L$ rather than $(T-L)L$. This alternative QD transformation might be of particular interest if $T >> S$ and it is reasonable to consider a large N, T asymptotic framework as in McKenzie (2004) (Type III). As a result, the number of parameters does not grow as both N, T increase.

Remark 7. The possibility to perform quasi-differencing with respect to either $\boldsymbol{\Lambda}$ or $\boldsymbol{F}$ is similar in spirit to the estimation procedure of Robertson and Sarafidis (2015) for genuine panel data models. For the model studied in Robertson and Sarafidis (2015), quasi-differencing can be performed either with respect to the $\boldsymbol{F}$ or $\boldsymbol{G}$ matrices (where $\boldsymbol{G}$ depends on the covariance between the factor loadings and the instruments).

Remark 8. Note that if we estimate the factor loadings instead of the factors themselves, the weighting matrix $\boldsymbol{\Omega}$ is not block diagonal over the T dimension. Furthermore, if the $\boldsymbol{\Sigma}$ matrix is not diagonal, the optimal $\boldsymbol{\Omega}$ in the second step is not even block diagonal in the S dimension. In this case, the GMM objective function does not simplify as in (3.13).

4. TESTING, MODEL SELECTION, AND IDENTIFICATION

In this section, we briefly discuss how hypothesis testing and the selection of the number of factors can be performed under the conditions of Proposition 2. Later we discuss some examples, where one or more of these conditions can be potentially violated. Particularly, we discuss the issues of local and global identification.

4.1 Testing and Model Selection

Given that the estimator derived under Assumptions (A.1)–(A.6) has a well-defined asymptotic normal limit, hypothesis testing is conducted in the usual way. First of all, the t- and Wald statistics can be used to test parameter restrictions. Second, we can consider the Wald test for the $\mathbf{H}_0 : f_t = c \neq 0, \forall t = 1, \ldots, T$ (hence $\phi_t = -1$) of the fixed effects model:

$$W = N\left(\hat{\boldsymbol{\phi}}_{\text{GMM}n} + \boldsymbol{\iota}_{T-1}\right)'\left(\widehat{\text{Avar}\,\hat{\boldsymbol{\phi}}_{\text{GMM}n}}\right)^{-1} \times \left(\hat{\boldsymbol{\phi}}_{\text{GMM}n} + \boldsymbol{\iota}_{T-1}\right) \xrightarrow{d} \chi^2(T-1). \quad (4.1)$$

Here, the consistent estimator of the variance-covariance matrix of $\hat{\boldsymbol{\phi}}_{\text{GMM}n}$ is given by

$$\widehat{\text{Avar}\,\hat{\boldsymbol{\phi}}_{\text{GMM}n}} = \left(\boldsymbol{D}_{\phi}(\hat{\boldsymbol{\gamma}})'\hat{\boldsymbol{\Omega}}^{1/2}\boldsymbol{M}_{\hat{\boldsymbol{\Omega}}^{1/2}\boldsymbol{D}_{\theta}(\hat{\boldsymbol{\gamma}})}\hat{\boldsymbol{\Omega}}^{1/2}\boldsymbol{D}_{\phi}(\hat{\boldsymbol{\gamma}})\right)^{-1}, \quad (4.2)$$

with $\hat{\boldsymbol{\Omega}} = (\boldsymbol{M}(\hat{\boldsymbol{\phi}}_1)\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}_1)\boldsymbol{M}(\hat{\boldsymbol{\phi}}_1)')^{-1}$ evaluated at a consistent one-step estimator $\hat{\boldsymbol{\gamma}}_1$ (e.g., using the identity weighting matrix $\boldsymbol{\Omega}$). Given that the number of degrees of freedom grows linearly with T, one can suspect that some loss of power for moderate values of T might occur.

Similar to any standard GMM estimation problem it can be shown that under (A.1)–(A.6) the criterion function has a limiting chi-square distribution (provided that $(S-L)(T-L) - K > 0$, and accordingly modified for unbalanced and/or dynamic models):

$$J_N(L) = N(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})'\boldsymbol{M}(\hat{\boldsymbol{\phi}})'\boldsymbol{\Omega}_{\text{opt}}\boldsymbol{M}(\hat{\boldsymbol{\phi}})(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2_{(S-L)(T-L)-K}, \quad (4.3)$$

if $L = L_0$. Here $J_N(L)$ denotes the corresponding J-test for the model with L factors. Testing for the number of unobserved factors can be performed sequentially as in Ahn, Lee, and Schmidt (2013) or using a Bayesian information criterion (BIC) model selection criterion. One starts with $\mathbf{H}_0 : L_0 = 0$ and if the null hypothesis is rejected proceeds with $\mathbf{H}_0 : L_0 = 1$. The sequential procedure for a globally identified model, can be motivated by the fact that for $L < L_0$ (for any positive definite $\boldsymbol{\Omega}$)

$$J_N(L) = N(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})'\boldsymbol{M}(\hat{\boldsymbol{\phi}})'\boldsymbol{\Omega}\boldsymbol{M}(\hat{\boldsymbol{\phi}})(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}) \to \infty. \quad (4.4)$$

However, as we further discuss in the section on global identification, the sequential procedure and BIC can fail to consistently estimate the true number of factors if the global identification assumption is violated, as for some data-generating processes (DGPs) in this case $J_N(L=0) \xrightarrow{d} \chi^2_{ST-K}$. Alternatively, for model selection we use the Schwartz information criterion (BIC) of the following form

$$S_N(L) = J_N(L) - a \ln(N)((T-L)(S-L)-K). \tag{4.5}$$

For further details please refer to Propositions 2 and 3 of Ahn, Lee, and Schmidt (2013).

Remark 9. While $J_N(L)$ is invariant to a particular choice of "N," this is not the case for $S_N(L)$. In the empirical section we consider two BIC criteria based on $N = \sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$ as, for example, in Inoue (2008) and $N = (1/ST)\sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$.

4.2 Identification: Local and Weak

At first, we consider the local asymptotic identification condition summarized in (A.4). For the asymptotic distribution to be properly defined, the Jacobian sub-matrix $\boldsymbol{M}(\boldsymbol{\phi}_0)\boldsymbol{X}_\infty$ should have full column rank K. This condition is more restrictive than the analogous condition for the model with only fixed effects, where it is necessary that $\boldsymbol{M}\boldsymbol{X}_\infty$ (where $\boldsymbol{M} = \boldsymbol{I}_S \otimes (\boldsymbol{I}_T - (1/T)\boldsymbol{\iota}_T\boldsymbol{\iota}_T'))$ is of full column rank. For example, the rank condition for the nonlinear GMM estimator is not satisfied in a model with one individual specific regressor of the following form:

$$z_{i,t} = \boldsymbol{f}_t'\boldsymbol{\lambda}_i^z + \varepsilon_{i,t}^z, \quad \varepsilon_{i,t}^z \sim (0, \sigma_{i,t}^2), \tag{4.6}$$

and as a result one has

$$\operatorname{plim}_{N\to\infty} \boldsymbol{M}(\boldsymbol{\phi}_0)\boldsymbol{z} = \boldsymbol{0}_{S(T-L)}. \tag{4.7}$$

In this example, the cross-sectional averages of $z_{i,t}$ asymptotically lie in the space spanned by $\boldsymbol{F}$. In the fixed effects model (i.e., $\boldsymbol{f}_t'\boldsymbol{\lambda}_s^z = \delta_s$), this condition is violated if the mean of $z_{i,t}$ does not sufficiently vary over time. On the other hand, if the factors $\boldsymbol{f}_t$ in the equation for $z_{i,t}$ differ from the corresponding factors in $y_{i,t}$ then the asymptotic identification condition is satisfied.

Remark 10. In the genuine panel data literature, it is sometimes assumed that factor loadings in the equation for $z_{i,t}$ have a different (nonzero) mean than the corresponding factor loadings in the equation for $y_{i,t}$, for example, the common correlated effects (CCE) estimator of Pesaran (2006) for $L > 1$ (in our case that implies $\mathrm{E}[\boldsymbol{\lambda}_i^z] \neq \boldsymbol{\lambda}_s \neq \boldsymbol{0}_L$). In such a case it might be tempting to use the Quasi-differencing approach with respect to cohort factor loadings (as in (3.14)) rather than factors to circumvent the problem with local identification. However, analogously to the derivations in Section 4.3, one can show that for this setup the global (rather than local) identification assumption is violated.

The full rank Jacobian condition plays an important role in proving the consistency and asymptotic normality of the GMM estimator. To further illustrate the importance of this assumption for pseudo panel models, we consider the simplified example with one endogenous regressor.

Example 1 (One Regressor).

$$y_{i,t} = \zeta z_{i,t} + u_{i,t}, \quad u_{i,t} = \rho\left(z_{i,t} - \frac{\mu_{s,t}}{N_{s,t}^{\gamma}}\right) + \varepsilon_{i,t} \quad i \in \mathcal{I}_{s,t},$$

$$z_{i,t} = \frac{\mu_{s,t}}{N_{s,t}^{\gamma}} + \varepsilon_{i,t}^{(z)}, \quad \varepsilon_{i,t}, \varepsilon_{i,t}^{(z)} \sim \mathrm{iid}(0,1).$$

For simplicity, we assume that $\boldsymbol{\lambda}_s = \boldsymbol{0}_L, \forall s$ and $\mu_{s,t}$ are deterministic, so that the first-step estimator of ζ is given by (3.1) with $\boldsymbol{\Omega} = \boldsymbol{I}_{ST}$ and $\boldsymbol{M} = \boldsymbol{I}_{ST}$. Then we can state the following result:

Proposition 3. Let the assumptions in the One Regressor example be satisfied, then

$$(\hat{\zeta} - \zeta_0) - \rho \xrightarrow{d} \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\pi_{s,t}^{-1} z_{s,t}\varepsilon_{s,t}}{\sum_{s=1}^{S}\sum_{t=1}^{T}\pi_{s,t}^{-1} z_{s,t}^2}, \quad \text{if } \gamma \geq 1/2.$$

$$N^{1/2-\gamma}(\hat{\zeta} - \zeta_0) \xrightarrow{d} \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\pi_{s,t}^{-(0.5+\gamma)}\mu_{s,t}u_{s,t}}{\sum_{s=1}^{S}\sum_{t=1}^{T}\pi_{s,t}^{-2\gamma}\mu_{s,t}^2}, \quad \text{if } 0 \leq \gamma < 1/2.$$

Proof. In Appendix A.1. □

Here the limiting random variables with subscript s, t are defined as

$$z_{s,t} \sim 1_{(\gamma=1/2)}\mu_{s,t} + \mathcal{N}(0,1), \quad \varepsilon_{s,t} \sim -1_{(\gamma=1/2)}\rho\mu_{s,t} + \mathcal{N}(0,1),$$

with $u_{s,t} \sim \mathcal{N}(0, 1+\rho^2)$, while all Gaussian random variables are mutually independent. As a result, the nonscaled estimator of the structural parameter ζ converges in distribution a random variable centered at $\zeta_0 + \rho$ under weak-instrument asymptotics (if all $\mu_{s,t} = 0$ for $\gamma = 1/2$). On the other hand, for semiweak (or semistrong) identification ($0 \leq \gamma < 1/2$), the estimator retains the asymptotic normal limit centered at the true value ζ_0 but with slower rate of convergence $N^{1/2-\gamma}$.

In the supplementary material, we discuss how one can use the result from this example to analyze the panel AR(1) model. The important lesson that we learn from this example is that in cases where the rank condition can be potentially locally violated, endogeneity starts to play an important role. This is in sharp contrast to the full-rank assumption. In other words, endogenous regressors play a role even if one considers the Type I asymptotic scheme. We investigate implications of this example for the more detailed model in the Monte Carlo section of this article.

4.3 Identification: Global

In addition to the full rank condition of the Jacobian matrix, the model has to be globally identified, as formally summarized in Assumption (A.6). We start this section with the most trivial case by considering the setup as in (4.6), but with

$$\mathrm{E}[\boldsymbol{\lambda}_i^z] = \kappa\boldsymbol{\lambda}_s, \forall i \in \mathcal{I}_{s,t}, \quad \kappa \neq 0. \tag{4.8}$$

Then the required global identification condition for the model with one regressor is of the very simple form

$$\boldsymbol{M}(\boldsymbol{\phi})\left((\boldsymbol{I}_S \otimes \boldsymbol{F})\operatorname{vec}(\boldsymbol{\Lambda}')(\kappa(\zeta_0 - \zeta) + 1)\right) = \boldsymbol{0}_{S(T-L)}. \tag{4.9}$$

This condition is satisfied for $\zeta = \zeta_0 + 1/\kappa$, irrespective of the value of $\boldsymbol{\phi}$. In this case, if one performs sequential selection of factors, the model with $L = 0$ can be selected, as the $J_N(L = 0)$ statistic for this model has a nondegenerate chi-squared limit. The corresponding (inconsistent) estimator satisfies $\hat{\zeta} \xrightarrow{p} \zeta_0 + 1/\kappa$.

Now consider a slightly less restrictive DGP for $z_{i,t}$

$$z_{i,t} = (\dot{\boldsymbol{f}}_t)'\boldsymbol{\lambda}_i^z + \varepsilon_{i,t}^z, \tag{4.10}$$

where we assume that $\dot{\boldsymbol{f}}_t \neq \boldsymbol{f}_t$, but the factor loadings are assumed to satisfy (4.8). Defining $\dot{\boldsymbol{F}} \equiv (\kappa \dot{\boldsymbol{f}}_1, \ldots, \kappa \dot{\boldsymbol{f}}_T)'$ the global identification condition can then be formulated in the following way:

$$\text{plim}_{N\to\infty} \boldsymbol{M}(\boldsymbol{\phi})(\boldsymbol{y} - \boldsymbol{X}\zeta) = \boldsymbol{M}(\boldsymbol{\phi})\,\text{vec}(((\zeta_0 - \zeta)\dot{\boldsymbol{F}} + \boldsymbol{F})\boldsymbol{\Lambda}') = \boldsymbol{0}_{S(T-L)}. \quad (4.11)$$

If we can further assume that the appropriate $[L \times L]$ block of the $\tilde{\boldsymbol{F}} = (\zeta_0 - \zeta)\dot{\boldsymbol{F}} + \boldsymbol{F}$ matrix is invertible, then each parameter value from the set

$$\boldsymbol{\Gamma} = \{\boldsymbol{\gamma} = (\zeta, \boldsymbol{\phi}')' \in \mathbb{R}^{(T-L)L+1} : \zeta \in \mathbb{R}, \quad \boldsymbol{\Phi} = -\tilde{\boldsymbol{F}}_{[(T-L)\times L]}\tilde{\boldsymbol{F}}^{-1}_{[L\times L]}\} \quad (4.12)$$

satisfies (4.11). In the supplementary material, we show that the AR(1) model can be written as (4.10), highlighting the global identification issues of the model without any other regressors. Finally, one can show that similar result is valid if we consider the example in (4.8) with quasi-differencing with respect to factor loadings rather than factors.

5. MONTE CARLO STUDY

5.1 Setup

The main goal of this Monte Carlo study is to investigate the effects of possibly endogenous regressors in the nearly singular designs with factors. By doing so, we expand the literature to models with unobserved factors and cases where asymptotic identification assumption might be (locally) violated. The summary of the Monte Carlo setup is provided below.

$$y_{i,t} = f_t\lambda_s + \beta w_{s,t} + \zeta z_{i,t} + \varepsilon_{i,t},$$

$$\varepsilon_{i,t} = \rho(z_{i,t} - \mu_{s,t}) + \sqrt{1-\rho^2}\sqrt{1-\sigma^2_{f\lambda}}\tilde{\eta}_{i,t}, \quad \tilde{\eta}_i \sim \mathcal{N}(0,1),$$

$$z_{i,t} = \mu_{s,t} + \sqrt{1-\sigma^2_\mu}\tilde{z}_{i,t}, \quad \tilde{z}_{i,t} \sim \mathcal{N}(0,1), \quad \mu_{s,t} = 1 + \sigma_\mu \mathcal{N}(0,1),$$

$$w_{s,t} \sim \mathcal{N}(0,1), \quad \bar{\lambda}_s \sim \mathcal{U}(0,1),$$

$$\lambda_s \sim \mathcal{N}(\bar{\lambda}_s, \sigma^2_{\lambda_s}), \quad f_t = 1 + \sigma_f u^{(f)}_t, \quad u^{(f)}_t = \alpha_f u^{(f)}_{t-1} + \varepsilon^{(f)}_t,$$

$$\varepsilon^{(f)}_t \sim \mathcal{N}(0, 1-\alpha_f^2), \quad u^{(f)}_0 \sim \mathcal{N}(0,1).$$

Similarly to the Monte Carlo setup of Inoue (2008), we normalize $\text{var}\,(f_t\lambda_s + \varepsilon_{i,t}) = 1$. The following parameter space is considered:

$$T = 5, \quad S = 10, \quad \boldsymbol{\theta}_0 = (1,1)', \quad \alpha_f = 0.5,$$

$$\bar{N} = \{150; 300\}, \quad \sigma^2_{f\lambda} = \{0.1; 0.5\},$$

$$\sigma^2_\mu = \{0; 0.05; 0.3\}, \quad \rho = \{0; 0.3\}, \quad \sigma_f = \{0; 0.1\}.$$

To ensure that $\text{var}\,(f_t\lambda_s) = \sigma^2_{f\lambda}$, the $\sigma^2_{\lambda_s}$ is set to

$$\sigma^2_{\lambda_s} = \frac{\sigma^2_{f\lambda} - \sigma^2_f\bar{\lambda}^2_s}{\sigma^2_f + 1}. \quad (5.1)$$

The fact that we normalize $\text{var}\,(f_t\lambda_s + \varepsilon_{i,t}) = 1$ ensures easy interpretation of $\sigma^2_{f\lambda}$, that is, the fraction of the total error term explained by the factor structure. Easy interpretability of the Monte Carlo parameters is a desirable property of a good Monte Carlo study, see, for example, Kiviet (2012). Note that the values of $\sigma^2_{f\lambda}$ and σ^2_f are relatively small to emphasize that the importance of the time-varying factor even when this contribution (to the total error structure) is small.

Results for five different estimators are presented. We denote two linear GMM estimators that do not allow any cohort effects ($\boldsymbol{M} = \boldsymbol{I}_{ST}$, i.e., $L = 0$) or only time-invarying cohort effects ($\boldsymbol{M} = \boldsymbol{I}_S \otimes (\boldsymbol{I}_T - (1/T)\boldsymbol{\iota}_T\boldsymbol{\iota}_T')$), by GMMl0 and GMMl1 (that are obtained as in (3.1) but with the optimal weighting matrix). Furthermore, we use abbreviations GMMn1, GMMn2, and GMMo to denote the two-step nonlinear GMM with $L = L_0 = 1$, nonlinear GMM with $L = 2$ (both solutions to (3.9)) and GMM based on BIC selection criteria, respectively. We also present results for the linear estimator that allows for time and cohort fixed effects, GMMl2.

The key parameters of this Monte Carlo study are

- (σ^2_μ) This parameter controls the degree of singularity of the Jacobian matrix. Larger value of σ^2_μ, translates into a larger amount of time-series heterogeneity in $z_{i,t}$. For $\sigma^2_\mu \approx 0$, $\sigma^2_f \approx 0$ the Jacobian is (near) singular and one can expect standard asymptotic inference techniques to perform poorly in finite samples, see Section 4.2.
- (ρ) This parameter controls the correlation between $z_{i,t}$ and the error term $\varepsilon_{i,t}$. Based on results from Section 4.2, for $\sigma^2_\mu = 0$ the bias of GMMl1 and GMMn1 is proportional to ρ.
- (σ^2_f) When this parameter is nonzero, one time-varying factor is present in the model. As a result, in this case the GMMl1 and GMMl2 estimators are inconsistent.

Note that all s, t specific variables $\{\lambda_s, f_t, w_{s,t}, \mu_{s,t}\}$ are simulated in every Monte Carlo replication so that the limiting distribution of the estimator is only conditionally normal. This is done to emphasize that in Assumptions (A.1)–(A.6) these quantities are assumed to be deterministic only for technical reasons. On the other hand, the $\bar{\lambda}_s$'s are generated only once in each design to make sure that the GMMl0 estimator (estimator without any factors, i.e., $L = 0$) is biased in finite samples. However, the results for other estimators do not change quantitatively or qualitatively if all $\bar{\lambda}_s = 0$. Similar to Inoue (2008), other distributional assumptions for $\{\lambda_s, \mu_{s,t}, w_{s,t}\}$ can be considered, but the given setup is sufficient for our purposes. $N_{s,t}$ is set to be $\lfloor \pi_{s,t}\bar{N}ST \rfloor$, where $\pi_{s,t} \sim U(0,1)$. Note, by generating λ_s, f_t in each replication we deviate from the theoretical discussions in this article, but are more in line with genuine panel data literature and the setup of Inoue (2008).

All results are presented for the two-step estimators (where necessary) with the asymptotically optimal weighting matrix, under the assumption that $\sigma^2_{s,t} = \sigma^2$. In this section, we discuss the results for the ζ parameter only; results for β are available from the author upon request.

Remark 11. Note that, for the given setup the GMMl0 estimator is always inconsistent and biased, where the second result is due to $\bar{\lambda}_s$ being nonzero. The GMMl1 estimator, on the other hand, is always unbiased but inconsistent if $\sigma^2_f \neq 0$. In this article, we do not consider any designs where for $\sigma^2_f \neq 0$ and $\rho = 0$ the GMMl1 estimator is both inconsistent and biased. For unbiasedness it is sufficient to have $\text{E}[f_t] = c$ and $\text{E}[z_{i,t}] = \tilde{c}$ for all t, this condition is satisfied in our Monte Carlo setup. The GMMl2 estimator remains unbiased even if $\mathrm{E}[f] \neq c$ for some t.

Remark 12. In the supplementary material, we present additional Monte Carlo evidence for the model with two factors. Results suggest that the variance of factors plays a more substantial role in the two-factor model as it is necessary to identify two time-varying factors from cross-sectional averages.

5.2 Results: Estimation

In this section, we summarize the bias and root mean square estimation (RMSE) properties of the estimators as presented in Table B.1. All results in the tables are grouped by the corresponding values of the σ_μ^2 variable. Here we briefly summarize the main findings.

- The GMMl0 estimator is severely biased in finite samples, driving the corresponding values of the RMSE.
- The results for GMMl1 and GMMl2 estimator are quantitatively and qualitatively similar. Hence, the effect of including the time effects on RMSE is marginal.
- Finite sample bias of all GMM estimators (except GMML0) is proportional to ρ, when these estimators are asymptotically (locally) unidentified ($\sigma_\mu^2 = 0$). The bias quickly disappears once $\sigma_\mu^2 > 0$, for example, already for $\sigma_\mu^2 = 0.05$ the bias is small.
- The effects of asymptotic nonidentification show up clearly once we consider the corresponding values of the RMSE (even for $\rho = 0$), as in this case $\bar{N}$ has no effect on RMSE. The slightest increase in σ_μ^2 substantially reduces the RMSE and $\bar{N}$ starts to play a role.
- Although inconsistent for $\sigma_f^2 \neq 0$, the GMMl1 estimator remains unbiased. For small values of $\sigma_{f\lambda}^2$ it has small bias and RMSE similar to the consistent GMMn1 estimator.
- Designs with $\sigma_{f\lambda}^2 = 0.5$ show substantial improvements in terms of both bias and RMSE for consistent estimators. If time-varying factor is present ($\sigma_f^2 \neq 0$), the RMSE of GMMl1 is larger than that of GMMn1. Hence, one needs a sufficiently strong factor for GMMn1 to have smaller RMSE, as compared to GMMl1.
- The RMSE of GMMn2 ($L > L_0$) tends to be larger than of correctly specified GMMn1 ($L = L_0$).

5.3 Results: Testing and Model Selection

In this section, inferential properties of the estimators are considered and discussed. As discussed in Section 4.1, one can use the Wald test to test the fixed effects assumption. In Table B.2, we also present results for two Hausman tests, that we denote by h0 and h1. Here h0 is the test statistic for *GMMl0 versus GMMl1*, and h1 tests *GMMl1 versus GMMn1*. The h0 almost in all design rejects close to 100% of all Monte Carlo replications. All test statistics have a nominal size of 5% (with the exception of the tests based on the GMMn2 estimator, as no asymptotic results for that estimator are available).

- The Wald test is superior in terms of both size and power as compared to the h1 test. For low values of $\sigma_{f\lambda}^2$ and $\bar{N}$, the former test is slightly size-distorted. Similar conclusions apply to the GMMl1 estimator and the corresponding J-test; the fixed effects model is rejected more frequently for larger values of $\sigma_{f\lambda}^2$ and $\bar{N}$.
- A larger of σ_μ^2 does not seem to influence the Wald test a lot, while an increase in $\sigma_{f\lambda}^2$ and/or $\bar{N}$ has a positive impact on size.
- The size of the J- and t-tests for the GMMl1 and GMMl2 estimators is similar when both estimators are consistent. On the other hand, when $\sigma_f^2 \neq 0$, the tests based on GMMl2 reject the null hypothesis less frequently than the corresponding tests for GMMl1 estimator. Hence, the J-test based on GMMl2 is less powerful to detect model misspecifications.
- Some size distortions are present for the J-test of the consistent GMMn1 and GMMo estimator, if $\sigma_\mu^2 = 0.05$. In general, distortions disappear with larger values of $\bar{N}$, $\sigma_{f\lambda}^2$, and σ_μ^2.
- For the GMMn2 estimator, the J-test is always undersized, which is driven by the fact that we estimate more factors than present in the model.
- Local nonidentification ($\sigma_\mu^2 = 0.00$) has no major impact on the J-test. However, one can observe that this test tends to under reject this can be partially explained using the results in Dovonon and Renault (2009).
- If the GMMl1 estimator is consistent, the t-test has better size properties than the corresponding test for GMMn1. On the other hand, when this estimator is inconsistent, the rejection frequencies slowly approach 1 as $\bar{N}$, σ_μ^2, and/or $\sigma_{f\lambda}^2$ increase.
- Empirical rejection frequencies of the t-test for GMMn1, GMMo, and GMMn2 approach the nominal size of 5% for larger values of $\bar{N}$, σ_μ^2, and/or $\sigma_{f\lambda}^2$.
- If all GMM estimators are (locally) nonidentified ($\sigma_\mu^2 = 0.0$), the properties of the t- test are driven by ρ, for example, for $\rho = 0.3$ rejection frequencies are close to 1.

The results for BIC criteria can be found in the last column of Table B.2 ($\#L = 1$), where each number indicates the fraction of Monte Carlo replications in which the correct number of factors was selected ($L_0 = 1$ in this case). Here, we adopt a procedure similar to Ahn, Lee, and Schmidt (2013) and set $a = 0.75/\ln(5)$, and $N = (1/ST)\sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$. The results based on $N = \sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$ are similar and available from the author upon request.

We can see that the BIC-based model selection procedure performs well in general. In 8 out of 48 designs, the proportion of correctly specified L is marginally lower than 95% while in the majority of cases this is above 98%. The results are not highly sensitive to the choice of the design parameters, but it is clearly seen that a higher relative weight of unobserved factor components as represented by the $\sigma_{f\lambda}^2$ parameter has a positive effect on the model selection procedure. Finally, we do not document any substantial effects on model selection when $\rho \neq 0$ and $\sigma_\mu^2 = 0$.

6. EMPIRICAL ILLUSTRATION

6.1 The ENEMDU Dataset

In this section, we estimate the labor supply elasticity in Ecuador for working males that are also the heads of the household, based on the data from 2007 to 2013. To accommodate possible common shocks, we estimate the model with cohort interactive effects. We use annual data from the *National Employment and Unemployment Survey* (ENEMDU) collected by the National Institute of Statistics and Census of Ecuador. The dataset contains information at the household level, with information provided about all individuals of age 5 and above. We consider only surveys from the fourth quarter of each year as it contains the largest number of observations, which is also representative for annual observations. This is partially done to ensure that each cohort contains at least 100 individuals.

Like some other studies (e.g., Antman and McKenzie 2007), we study the labor market participation of prime aged males (26–55) occupying a single job. We restrict our sample to males who work for at least 30 hr but no more than 60 hr per week to minimize the number of potential outliers. Moreover, as we are only interested in the intensive margin (the number of hours worked) of the labor supply, the observations with a lower number of hours worked (corresponding to part-time workers) are not of prime interest. A joint study of the extensive (decision to work full/part time) and intensive margin (the number of hours worked) is complicated due to the scarcity of the available explanatory variables. To obtain real rather than nominal income, we deflated individual income using the annual Consumer Price Index (CPI) at the national level.

Before proceeding with estimation and model specification, we discuss how we define cohorts in our study. Similar to González and Sala (2015), we define cohorts solely based on the age of the individual. In total, we construct 10 cohorts of equal age intervals based on individuals born in 1952–1981, therefore each cohort represents a 3-year interval. Alternatively, one could define cohorts based on 5-year intervals and/or geographical location. That strategy, on the other hand, would substantially reduce the average number of observations per cohort or the total number of cohorts. This is a common tradeoff faced by practitioners when dealing with pseudo panel datasets, see, for example, Verbeek and Vella (2005) and Verbeek (2008). Hence, for simplicity and ease of exposition we consider only results based on 3-year intervals. As discussed by Inoue (2008) the adequacy of the model as well as the definition of cohorts can be investigated by means of the J-test.

6.2 Results

As a basic setup, we are interested in the model of the following form:

$$\log \text{hours}_{i,t} = \gamma \log \text{wage}_{i,t} + \boldsymbol{\beta}' z_{i,t} + \boldsymbol{\theta}' \boldsymbol{q}_{i,t} + u_{i,t},$$
$$\text{E}[u_{i,t}] = 0, \quad i \in \mathcal{I}_{s,t}. \tag{6.1}$$

Here, $\log \text{hours}_{i,t}$ is the logarithm of the weekly hours worked by individual i while $\log \text{wage}_{i,t}$ is the real hourly wage. Models of similar form were extensively estimated using genuine panel data methods, see, for example, Ziliak (1997). Furthermore, we assume that the regressors in $z_{i,t}$ are observed by the econometrician, however $\boldsymbol{q}_{i,t}$ are unobserved but can be well approximated by

$$\boldsymbol{q}_{i,t} = \boldsymbol{\Lambda}_s^{(q)} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_{i,t}, \quad \bar{\boldsymbol{\varepsilon}}_{s,t} \xrightarrow{p} \mathbf{0}_{K_q}. \tag{6.2}$$

We would like to stress that we do not assume that $\text{E}[z_{i,t}\boldsymbol{q}_{i,t}'] = \mathbf{O}_{K_z \times K_q}$ or $\text{E}[\boldsymbol{q}_{i,t} \log \text{wage}_{i,t}] = \mathbf{0}_{K_q}$ hence we can allow for endogeneity in our framework. For example, due to the nonavailability of consumption data in Ecuador, we assume that this endogenous variable is a part of the $\boldsymbol{q}_{i,t}$ variables rather than of $z_{i,t}$.

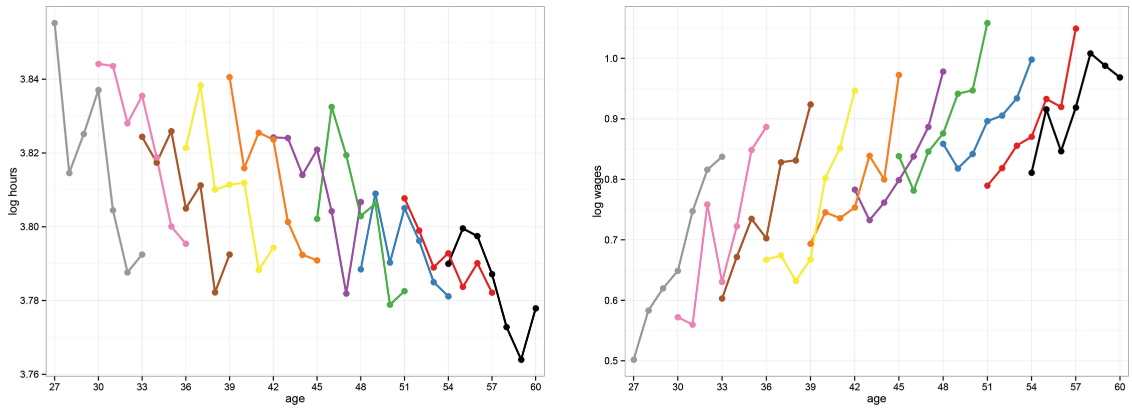
Combining both equations, we obtain a simple model to study the labor supply elasticity in the intensive margin that can be summarized by the following equation:

$$\log \text{hours}_{i,t} = \boldsymbol{\lambda}_s' \boldsymbol{f}_t + \gamma \log \text{wage}_{i,t} + \beta z_{i,t} + \upsilon_{i,t},$$
$$\upsilon_{i,t} = u_{i,t} + \boldsymbol{\theta}' \boldsymbol{\varepsilon}_{i,t}, \quad i \in \mathcal{I}_{s,t}, \tag{6.3}$$

while $\boldsymbol{\lambda}_s' = \boldsymbol{\theta}' \boldsymbol{\Lambda}_s^{(q)}$. The only control variable that we include in our model that is of particular interest on its own (and is available in the dataset) is the total number of reported individuals under the age of 16 in a given household. By including this variable, we follow some other studies, particularly Peterman (2014), and control for the household composition. The averages of $\log \text{hours}_{i,t}$ worked and $\log \text{wage}_{i,t}$ are presented in Figure 1, while average values of $z_{i,t}$ and the number of observations per cohort are summarized in Tables C1–C2 in Appendix C.

As discussed in Sections 4.1 and 4.3, J-test of GMMl0 estimator can be used as a possible indication for global nonidentification. For this model specification, the J-test is equal to $J_N(L = 0) = 698.91$, which indicates a clear rejection based on the critical values from $\chi^2(68)$ at any conventional significance level. That provides some indication that global identification failure for this dataset is not very likely. In addition to the estimators considered in the Monte Carlo section, we also include the linear GMM estimator (GMMl2) that allows for cohort and time effects in the specification of the error component structure.

As we can see from the estimation results in Table 1, the GMMl1 model specification is rejected based on the J-test. This conclusion is also confirmed using the Wald test to test the null hypothesis that the factor is constant over time (see also Figure C.1 in Appendix C). Hence, based on the statistical procedures we consider, the assumption that cohorts were not affected by any internal or external shocks is difficult to justify. For GMMl1, the estimated elasticity coefficient of log wage is negative and significant at any conventional significance level. On the other hand, if we allow for cohort interactive effects the estimated elasticity coefficient is smaller in magnitude and does not significantly differ from zero at the 1% significance level (for the GMMn2 estimator). Based on the BIC model selection criteria, the model specification with one factor is preferred. We also find that the results for the GMMl2 estimator are both quantitatively and qualitatively similar to nonlinear estimators with one and two factors. Furthermore, the J-test for GMMl2 does not reject the null hypothesis of correct model specification (unlike GMMl1). This is partially in-line with Monte Carlo findings where the J-test for GMMl2 was found to have lower power to reject the null hypothesis in the presence of multiplicative factors. Turning our attention to the estimated coefficient



(a) Average log *hours* worked. (b) Average log *wage*.

Figure 1. Age of each cohort is defined as the middle point in the interval.

Table 1. $T = 7$, $S = 10$. Results are based on two-step estimates using the optimal weighting matrix in the second step. Based only on heads of the household. ** indicates statistical significance at the 5% level and *** at the 1% level $\gamma_0, \beta_0 = 0$. $J(\text{GMMl0}) = 577.84$. BIC_1 and BIC_2 use $N = \sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$ and $N = (1/ST)\sum_{s=1}^{S}\sum_{t=1}^{T} N_{s,t}$, respectively.

Variable	GMMl1	GMMl2	GMMn1	GMMn2
log wage	−0.130***	−0.076***	−0.075***	−0.076**
# kids	0.003	−0.009	−0.008	−0.011
df	58	52	52	38
J	88.05***	57.98	56.97	44.92
BIC_1			−187.35	−133.62
BIC_2			−84.41	−58.39
Wald(FE)	33.874***			

of #*kids* we can see that the results differ slightly between estimators. In all cases, we find that estimates are not significant at any conventional significance level.

Although methodologically our study is simpler and spans a shorter (and different) time period, we can compare our results with those in González and Sala (2015). They found that for some Latin American countries, particularly Paraguay, the estimate of the labor supply elasticity is strongly negative, while for others, it was found to be positive (Argentina, and after sample restriction, Uruguay). Our estimate of the elasticity coefficient places Ecuador closer to countries like Paraguay than to Argentina or Chile. As a robustness check in the supplementary material, we also provide results based on a *linear-log* specification with qualitatively similar conclusions.

7. CONCLUSIONS

In this article, we have studied the properties of available estimation techniques for linear pseudo panel data models. We have extended the pseudo panel data literature to models with possible cohort interactive effects. To overcome inconsistency of the usual FE estimator, we have introduced the approach of Ahn, Lee, and Schmidt (2013) to pseudo panel data models. The consistency and conditional asymptotic normality of the new estimator was proved for pseudo panels with a fixed number of time series observations and cohorts. Furthermore, we have discussed the estimation and identification for datasets with a cohort-specific number of time observations.

Results from the extensive Monte Carlo study suggest that the estimator that accounts for the multiplicative structure of the cohort effects has good finite sample properties for small values of S and T. The results, however, can be sensitive to the relative importance of the unobserved factors in the total error component structure.

As an empirical illustration, we have studied labor supply elasticity based on data from Ecuador. In our analysis, we have found that the model with multiplicative error structure provides a better fit to the data than its counterpart with fixed effects.

As thoroughly discussed by McKenzie (2004) and Verbeek (2008), different types of asymptotic approximations are available for pseudo panel data models, depending on their dimensions. In this article, we have mainly investigated the effects of the error terms with multifactor structure assuming that the number of cohorts and the time dimension is fixed. This assumption is only sensible for models with limited number of cohorts but a large number of observations per cohort. However, given the limited scope of this article we leave the rigorous analysis of other asymptotic schemes for models with multifactor error structure for future research.

APPENDIX A: THEORETICAL RESULTS

A.1 Proofs

Proof of Proposition 1. The result of this proposition follows directly given the DGP is given by

$$\boldsymbol{y} = \text{vec}\,(\boldsymbol{F}\boldsymbol{\Lambda}') + \boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{u}.$$

Plugging in the expression for $\boldsymbol{y}$ into the formula for

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{GMM}l} - \boldsymbol{\theta}_0 &= \left(\boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}(\text{vec}\,(\boldsymbol{F}\boldsymbol{\Lambda}') + \boldsymbol{u}) \\ &= \left(\boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}(\text{vec}\,(\boldsymbol{F}\boldsymbol{\Lambda}')) + o_P(1) \\ &= \left(\boldsymbol{X}_\infty'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\boldsymbol{X}_\infty\right)^{-1}\boldsymbol{X}_\infty'\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}\,\text{vec}\,(\boldsymbol{F}\boldsymbol{\Lambda}') + o_P(1).\end{aligned}$$

Here the second line follows, as by assumption (A.2) $\text{plim}_{N\to\infty} \boldsymbol{u} = \mathbf{0}_{ST}$. Finally, using the notation $\text{plim}_{N\to\infty} \boldsymbol{X} = \boldsymbol{X}_\infty$ we obtain the final result. □

Proof of Proposition 2. The proof of this proposition is a straightforward modification of the proof for a simple IV estimator.

CASE I: $\gamma \geq 0.5$:

$$
\begin{aligned}
&\hat{\zeta} - \zeta_0 - \rho \\
&= \frac{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}(\bar{\varepsilon}_{s,t} - \rho(\mu_{s,t}/N_{s,t}^{\gamma}))}{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}^2} \\
&= \frac{N}{N}\frac{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}(\bar{\varepsilon}_{s,t} - \rho(\mu_{s,t}/N_{s,t}^{\gamma}))}{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}^2} \\
&= \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}(N/N_{s,t})(\sqrt{N_{s,t}}\bar{z}_{s,t})(\sqrt{N_{s,t}}\bar{\varepsilon}_{s,t} - \rho(\mu_{s,t}/N_{s,t}^{\gamma+0.5}))}{\sum_{s=1}^{S}\sum_{t=1}^{T}(N/N_{s,t})\left(\sqrt{N_{s,t}}\bar{z}_{s,t}\right)^2}.
\end{aligned}
$$

From here the desired result follows given that

$$
\begin{aligned}
&N/N_{s,t} \to \pi_{s,t}^{-1}, \\
&\sqrt{N_{s,t}}\bar{z}_{s,t} \xrightarrow{d} z_{s,t} \\
&\sqrt{N_{s,t}}\bar{\varepsilon}_{s,t} - \rho(\mu_{s,t}/N_{s,t}^{\gamma+0.5}) \xrightarrow{d} \varepsilon_{s,t}
\end{aligned}
$$

as we assume that all idiosyncratic components are iid and hence the usual central limit theorem (CLT) applies.

CASE II: $\gamma \in [0; 0.5)$:

$$
\begin{aligned}
N^{1/2-\gamma}(\hat{\zeta} - \zeta_0) &= \frac{N^{1/2+\gamma}}{N^{2\gamma}}\frac{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}(\bar{u}_{s,t})}{\sum_{s=1}^{S}\sum_{t=1}^{T} \bar{z}_{s,t}^2} \\
&= \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}(N/N_{s,t})^{1/2+\gamma}(N_{s,t}^{\gamma}\bar{z}_{s,t})(\sqrt{N_{s,t}}\bar{u}_{s,t})}{\sum_{s=1}^{S}\sum_{t=1}^{T}(N/N_{s,t})^{2\gamma}(N_{s,t}^{\gamma}\bar{z}_{s,t})^2}.
\end{aligned}
$$

By means of Slutsky's Theorem for the denominator

$$
N_{s,t}^{\gamma}\bar{z}_{s,t} = \mu_{s,t} + N_{s,t}^{\gamma-1/2}(\sqrt{N_{s,t}}\bar{\varepsilon}_{s,t}^{(z)}) \xrightarrow{p} \mu_{s,t}.
$$

For the numerator, simple CLT for iid data applies

$$
\sqrt{N_{s,t}}\bar{u}_{s,t} = \rho\sqrt{N_{s,t}}\bar{\varepsilon}_{s,t}^{z} + \sqrt{N_{s,t}}\bar{\varepsilon}_{s,t} \xrightarrow{d} \rho\mathcal{N}(0,1) + \mathcal{N}(0,1).
$$

The desired result follows by combining the results for the numerator and denominator

$$
N^{1/2-\gamma}(\hat{\zeta} - \zeta_0) \xrightarrow{d} \frac{\sum_{s=1}^{S}\sum_{t=1}^{T} \pi_{s,t}^{-(0.5+\gamma)}\mu_{s,t}u_{s,t}}{\sum_{s=1}^{S}\sum_{t=1}^{T} \pi_{s,t}^{-2\gamma}\mu_{s,t}^2},
$$

with $u_{s,t} \sim \mathcal{N}(0, 1+\rho^2)$. □

APPENDIX B: MONTE CARLO RESULTS

Table B.1. Estimation results for $T = 5$, $S = 10$. 10,000 MC replications. For $\zeta_0 = 1$

		Bias						RMSE					
σ_μ^2	$\{\sigma_{f\lambda}^2; \rho; \bar{N}; \sigma_f\}$	L_0	L_1^{FE}	L_2^{FE}	L_1	L_2	$\hat{L}$	L_0	L_1^{FE}	L_2^{FE}	L_1	L_2	$\hat{L}$
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	0.58	0.00	0.00	0.03	0.02	0.03	0.59	0.16	0.17	0.21	0.25	0.22
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	0.43	0.00	0.00	0.01	0.01	0.01	0.44	0.19	0.18	0.17	0.21	0.18
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	0.39	0.00	0.00	0.02	0.02	0.02	0.40	0.16	0.16	0.20	0.25	0.20
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	0.48	0.00	0.00	0.01	0.00	0.00	0.49	0.21	0.19	0.17	0.20	0.16
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	0.66	0.30	0.30	0.34	0.33	0.33	0.67	0.34	0.34	0.39	0.41	0.39
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	0.47	0.30	0.30	0.23	0.23	0.22	0.48	0.35	0.35	0.30	0.32	0.30
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	0.61	0.30	0.30	0.32	0.32	0.32	0.61	0.34	0.34	0.37	0.40	0.37
0.00	{ 0.1 ; 0.3 ; 300 ; 0.1 }	0.47	0.30	0.30	0.17	0.19	0.17	0.48	0.36	0.36	0.25	0.29	0.24
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	0.58	0.00	0.00	0.00	0.00	0.00	0.62	0.12	0.12	0.13	0.17	0.13
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	0.47	0.00	0.00	0.00	0.00	0.00	0.53	0.18	0.18	0.10	0.14	0.11
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	0.59	0.00	0.00	0.00	0.00	0.00	0.64	0.12	0.12	0.13	0.17	0.13
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	0.51	0.00	0.00	0.00	0.00	0.00	0.56	0.24	0.22	0.10	0.13	0.09
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	0.51	0.30	0.30	0.30	0.30	0.30	0.56	0.32	0.32	0.33	0.35	0.33
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	0.35	0.30	0.30	0.18	0.20	0.18	0.42	0.35	0.35	0.21	0.25	0.22
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	0.56	0.30	0.30	0.30	0.30	0.30	0.61	0.32	0.32	0.33	0.35	0.33
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	0.47	0.30	0.30	0.14	0.16	0.14	0.52	0.38	0.37	0.18	0.22	0.17
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	0.35	0.00	0.00	0.00	0.00	0.00	0.36	0.06	0.06	0.07	0.09	0.07
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	0.51	0.00	0.00	0.01	0.00	0.00	0.52	0.07	0.07	0.09	0.09	0.07
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	0.47	0.00	0.00	0.00	0.00	0.00	0.48	0.04	0.04	0.05	0.06	0.05
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	0.53	0.00	0.00	0.00	0.00	0.00	0.54	0.06	0.05	0.06	0.06	0.05
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	0.52	0.04	0.04	0.04	0.04	0.04	0.53	0.07	0.07	0.08	0.10	0.08
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	0.44	0.04	0.04	0.04	0.04	0.04	0.45	0.08	0.08	0.08	0.09	0.07
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	0.49	0.02	0.02	0.02	0.02	0.02	0.50	0.05	0.05	0.05	0.07	0.05
0.05	{ 0.1 ; 0.3 ; 300 ; 0.1 }	0.42	0.02	0.02	0.02	0.02	0.02	0.43	0.06	0.06	0.05	0.06	0.05
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	0.46	0.00	0.00	0.00	0.00	0.00	0.51	0.04	0.05	0.05	0.06	0.05
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	0.51	0.00	0.00	0.00	0.00	0.00	0.56	0.07	0.07	0.05	0.06	0.05
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	0.44	0.00	0.00	0.00	0.00	0.00	0.49	0.03	0.03	0.03	0.05	0.03
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	0.42	0.00	0.00	0.00	0.00	0.00	0.47	0.07	0.06	0.03	0.04	0.03
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	0.57	0.04	0.04	0.04	0.04	0.04	0.61	0.06	0.06	0.06	0.08	0.06
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	0.41	0.04	0.04	0.03	0.04	0.03	0.47	0.08	0.08	0.06	0.07	0.06
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	0.47	0.02	0.02	0.02	0.02	0.02	0.52	0.04	0.04	0.04	0.05	0.04
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	0.44	0.02	0.02	0.02	0.02	0.02	0.50	0.07	0.07	0.04	0.05	0.04
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	0.45	0.00	0.00	0.00	0.00	0.00	0.46	0.02	0.03	0.03	0.04	0.03
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	0.37	0.00	0.00	0.00	0.00	0.00	0.38	0.03	0.03	0.03	0.04	0.03
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	0.39	0.00	0.00	0.00	0.00	0.00	0.40	0.02	0.02	0.03	0.03	0.02
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	0.35	0.00	0.00	0.00	0.00	0.00	0.36	0.02	0.02	0.02	0.03	0.02
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	0.33	0.00	0.00	0.01	0.01	0.01	0.34	0.02	0.03	0.03	0.04	0.03
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	0.42	0.01	0.00	0.01	0.01	0.01	0.43	0.03	0.03	0.04	0.04	0.03
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	0.44	0.00	0.00	0.00	0.00	0.00	0.45	0.02	0.02	0.03	0.03	0.02
0.30	{ 0.1 ; 0.3 ; 300 ; 0.1 }	0.31	0.00	0.00	0.00	0.00	0.00	0.32	0.02	0.02	0.02	0.03	0.02
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	0.38	0.00	0.00	0.00	0.00	0.00	0.42	0.02	0.02	0.02	0.03	0.02
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	0.33	0.00	0.00	0.00	0.00	0.00	0.38	0.03	0.03	0.02	0.03	0.02
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	0.44	0.00	0.00	0.00	0.00	0.00	0.48	0.01	0.01	0.01	0.02	0.01
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	0.49	0.00	0.00	0.00	0.00	0.00	0.53	0.03	0.03	0.02	0.02	0.01
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	0.45	0.01	0.01	0.01	0.01	0.01	0.49	0.02	0.02	0.02	0.03	0.02
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	0.49	0.01	0.01	0.01	0.01	0.01	0.52	0.03	0.03	0.03	0.03	0.02
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	0.33	0.00	0.00	0.00	0.00	0.00	0.38	0.01	0.01	0.02	0.02	0.01
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	0.38	0.00	0.00	0.00	0.00	0.00	0.42	0.03	0.03	0.02	0.02	0.01

NOTE: Here "L_0" is the "GMMl0" estimator; "L_1^{FE}" and "L_2^{FE}" are the linear "GMMl1" and "GMMl2" estimators; "L_1" and "L_2" are the nonlinear "GMMn1" and "GMMn2" estimators; "$\hat{L}$" is the "GMMo" estimator with optimal number of factors based on BIC.

Table B.2. Testing results for $T = 5$, $S = 10$. 10,000 MC replications. For $\zeta_0 = 1$

		L_0		L_1^{FE}		L_2^{FE}		L_1		L_2		$\hat{L}$					
σ_μ^2	$\{\sigma_{f\lambda}^2; \rho; \bar{N}; \sigma_f\}$	J	t	J	t	J	t	J	t	J	t	J	t	W	$h0$	$h1$	$\#L = 1$
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	1	1	0.05	0.04	0.05	0.04	0.08	0.14	0.02	0.14	0.04	0.14	0.08	0.99	0.30	0.94
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	1	1	0.39	0.09	0.21	0.06	0.09	0.13	0.02	0.13	0.05	0.13	0.62	0.96	0.47	0.94
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	1	1	0.05	0.04	0.05	0.04	0.06	0.15	0.02	0.16	0.04	0.15	0.08	0.97	0.27	0.98
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	1	1	0.63	0.13	0.34	0.08	0.07	0.10	0.02	0.12	0.05	0.10	0.80	0.98	0.57	0.98
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	1	1	0.04	0.55	0.05	0.51	0.08	0.61	0.02	0.49	0.03	0.60	0.13	0.94	0.29	0.94
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	1	1	0.43	0.55	0.19	0.51	0.12	0.50	0.03	0.41	0.06	0.48	0.65	0.88	0.50	0.92
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	1	1	0.05	0.55	0.04	0.51	0.05	0.59	0.01	0.49	0.04	0.59	0.08	0.96	0.27	0.98
0.00	{ 0.1 ; 0.3 ; 300 ; 0.1 }	1	1	0.67	0.54	0.39	0.50	0.09	0.38	0.03	0.33	0.07	0.38	0.83	0.92	0.60	0.97
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	1	1	0.05	0.04	0.05	0.04	0.05	0.06	0.02	0.08	0.04	0.07	0.05	0.99	0.24	0.97
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	1	0.99	0.80	0.18	0.69	0.15	0.07	0.06	0.02	0.07	0.05	0.06	0.93	0.99	0.68	0.97
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	1	1	0.05	0.04	0.05	0.04	0.05	0.06	0.02	0.07	0.04	0.06	0.05	1	0.23	0.99
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	1	1	0.91	0.27	0.82	0.21	0.06	0.06	0.02	0.07	0.05	0.06	0.98	0.99	0.76	0.99
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	1	1	0.05	0.78	0.05	0.73	0.05	0.75	0.02	0.60	0.04	0.75	0.07	0.97	0.27	0.97
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	1	0.98	0.78	0.70	0.69	0.68	0.09	0.53	0.03	0.44	0.06	0.53	0.91	0.96	0.69	0.95
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	1	1	0.05	0.78	0.05	0.74	0.05	0.75	0.02	0.59	0.04	0.75	0.06	0.98	0.26	0.99
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	1	1	0.92	0.67	0.85	0.66	0.09	0.42	0.03	0.36	0.07	0.42	0.97	0.98	0.79	0.98
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	1	1	0.06	0.05	0.06	0.05	0.10	0.08	0.03	0.09	0.06	0.07	0.12	1	0.27	0.94
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	1	1	0.48	0.11	0.19	0.07	0.11	0.08	0.03	0.08	0.06	0.07	0.73	1	0.60	0.93
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	1	1	0.06	0.06	0.05	0.05	0.07	0.06	0.02	0.07	0.05	0.06	0.08	1	0.18	0.98
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	1	1	0.73	0.18	0.36	0.11	0.08	0.07	0.03	0.08	0.06	0.06	0.90	1	0.72	0.98
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	1	1	0.07	0.12	0.06	0.11	0.11	0.14	0.04	0.13	0.06	0.13	0.10	1	0.24	0.94
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	1	1	0.47	0.17	0.26	0.14	0.11	0.14	0.04	0.13	0.07	0.13	0.73	1	0.58	0.94
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	1	1	0.06	0.09	0.06	0.09	0.08	0.09	0.03	0.10	0.06	0.09	0.07	1	0.19	0.98
0.05	{ 0.1 ; 0.3 ; 300 ; 0.1 }	1	1	0.64	0.19	0.38	0.14	0.09	0.09	0.03	0.10	0.07	0.09	0.85	1	0.67	0.98
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	1	0.99	0.06	0.05	0.06	0.05	0.07	0.06	0.03	0.07	0.05	0.06	0.07	1	0.15	0.97
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	1	1	0.83	0.24	0.70	0.19	0.07	0.06	0.03	0.07	0.05	0.06	0.94	1	0.79	0.97
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	1	1	0.06	0.05	0.06	0.05	0.07	0.06	0.02	0.07	0.06	0.06	0.06	1	0.13	0.99
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	1	0.99	0.92	0.32	0.85	0.27	0.06	0.06	0.02	0.07	0.05	0.06	0.97	1	0.86	0.99
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	1	1	0.08	0.16	0.08	0.15	0.09	0.16	0.03	0.15	0.06	0.16	0.07	1	0.17	0.97
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	1	0.99	0.79	0.29	0.70	0.26	0.08	0.15	0.03	0.14	0.06	0.15	0.92	0.99	0.76	0.97
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	1	1	0.07	0.11	0.06	0.10	0.07	0.11	0.03	0.11	0.06	0.11	0.07	1	0.14	0.99
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	1	0.99	0.92	0.34	0.85	0.30	0.07	0.11	0.03	0.11	0.06	0.10	0.98	1	0.86	0.99
	{ 0.1 ; 0.0 ; 150 ; 0.0 }	1	1	0.06	0.05	0.06	0.05	0.09	0.06	0.03	0.08	0.06	0.06	0.08	1	0.17	0.96
	{ 0.1 ; 0.0 ; 150 ; 0.1 }	1	1	0.48	0.12	0.26	0.09	0.09	0.06	0.03	0.07	0.06	0.06	0.75	1	0.59	0.95
	{ 0.1 ; 0.0 ; 300 ; 0.0 }	1	1	0.06	0.05	0.05	0.05	0.08	0.06	0.03	0.07	0.06	0.06	0.08	1	0.15	0.98
	{ 0.1 ; 0.0 ; 300 ; 0.1 }	1	1	0.68	0.17	0.44	0.12	0.07	0.05	0.03	0.07	0.06	0.05	0.87	1	0.70	0.99
	{ 0.1 ; 0.3 ; 150 ; 0.0 }	1	1	0.06	0.06	0.06	0.06	0.10	0.07	0.04	0.08	0.06	0.07	0.11	1	0.21	0.95
	{ 0.1 ; 0.3 ; 150 ; 0.1 }	1	1	0.53	0.14	0.23	0.09	0.10	0.07	0.03	0.08	0.06	0.06	0.77	1	0.61	0.95
	{ 0.1 ; 0.3 ; 300 ; 0.0 }	1	1	0.06	0.05	0.06	0.05	0.07	0.06	0.03	0.07	0.06	0.06	0.07	1	0.12	0.99
0.30	{ 0.1 ; 0.3 ; 300 ; 0.1 }	1	1	0.62	0.16	0.39	0.12	0.08	0.06	0.03	0.07	0.06	0.06	0.85	1	0.68	0.98
	{ 0.5 ; 0.0 ; 150 ; 0.0 }	1	0.99	0.06	0.05	0.06	0.05	0.07	0.05	0.02	0.07	0.05	0.06	0.06	1	0.12	0.98
	{ 0.5 ; 0.0 ; 150 ; 0.1 }	1	0.99	0.80	0.24	0.70	0.19	0.07	0.06	0.03	0.07	0.05	0.06	0.93	1	0.78	0.97
	{ 0.5 ; 0.0 ; 300 ; 0.0 }	1	1	0.05	0.05	0.05	0.05	0.06	0.05	0.02	0.07	0.05	0.05	0.06	1	0.10	0.99
	{ 0.5 ; 0.0 ; 300 ; 0.1 }	1	1	0.94	0.37	0.85	0.27	0.06	0.05	0.02	0.06	0.05	0.05	0.98	1	0.89	0.99
	{ 0.5 ; 0.3 ; 150 ; 0.0 }	1	1	0.06	0.07	0.06	0.07	0.07	0.07	0.03	0.08	0.06	0.07	0.07	1	0.12	0.97
	{ 0.5 ; 0.3 ; 150 ; 0.1 }	1	1	0.85	0.27	0.69	0.20	0.07	0.07	0.03	0.08	0.06	0.07	0.95	1	0.81	0.97
	{ 0.5 ; 0.3 ; 300 ; 0.0 }	1	0.99	0.06	0.06	0.05	0.06	0.06	0.06	0.03	0.07	0.05	0.06	0.06	1	0.11	0.99
	{ 0.5 ; 0.3 ; 300 ; 0.1 }	1	0.99	0.92	0.34	0.86	0.29	0.06	0.06	0.03	0.07	0.06	0.06	0.98	1	0.87	0.99

NOTE: See Table B.1.

APPENDIX C: THE ENEMDU DATASET

Table C.1. The number of observations per cohort in a particular year

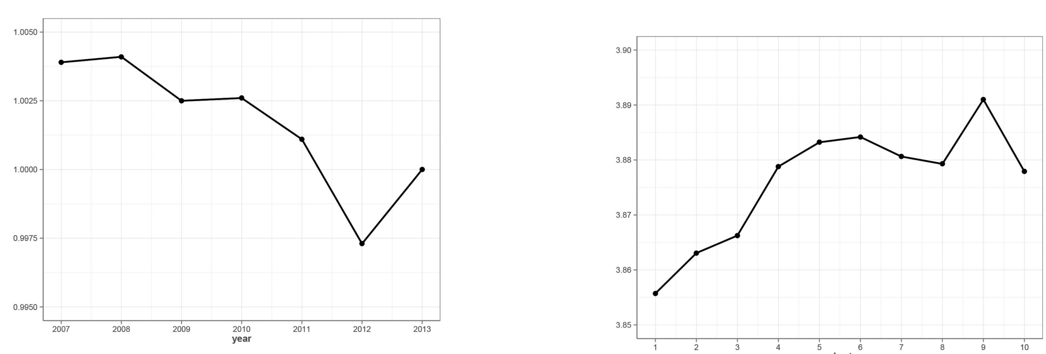
	1	2	3	4	5	6	7	8	9	10
2007	264	283	276	308	300	351	322	277	235	191
2008	288	329	383	383	366	353	361	298	267	211
2009	260	317	320	340	335	338	280	241	229	185
2010	318	341	390	417	396	398	390	272	285	202
2011	338	361	339	420	372	356	412	369	321	297
2012	338	359	411	429	420	402	441	387	342	301
2013	240	304	363	437	432	447	525	518	473	467

NOTE: Here 1 denotes the oldest and 10 youngest cohort, respectively.

Table C.2. The average number of individuals of age under 16 in a particular household

	1	2	3	4	5	6	7	8	9	10
2007	1.08	1.30	1.57	1.94	2.32	2.21	2.37	2.27	2.08	1.57
2008	0.92	1.47	1.41	1.87	2.18	2.09	2.46	2.18	2.12	1.70
2009	0.93	1.15	1.42	1.60	1.98	1.99	2.39	2.27	2.18	1.60
2010	0.90	1.08	1.30	1.54	1.86	2.05	2.17	2.18	2.37	1.79
2011	0.78	0.82	1.06	1.25	1.55	1.79	1.94	2.17	2.20	1.93
2012	0.80	0.97	0.97	1.21	1.40	1.86	1.92	2.21	2.21	1.94
2013	0.81	0.86	1.04	1.14	1.28	1.77	1.85	2.23	2.27	2.01

NOTE: Here 1 denotes the oldest and 10 youngest cohort, respectively.



(a) Estimated factors based on $f_T = 1$ normalization.

(b) Estimated factor loadings.

Figure C.1. Estimated factors and cohort factor loadings based on model with one time-varying factor in the log – log specification. Age of each cohort is defined as the middle point in the interval. The factor loadings are estimated using $\widehat{\text{vec}\,\boldsymbol{\Lambda}} = (\boldsymbol{I}_S \otimes ((\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'))(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\theta}})$.

SUPPLEMENTARY MATERIALS

Supplemental material contains additional material for identification and estimation of dynamic (possibly unbalanced) pseudo panels, additional Monte Carlo simulations for the model with one and two factors, and results for the linear-log specification in the empirical illustration.

ACKNOWLEDGMENTS

The author thanks the editor (Rong Chen), two anonymous referees, and an associate editor, whose comments changed the article substantially. The author also greatly benefitted from the comments made by his supervisors Peter Boswijk and Maurice Bun. Furthermore, the author is grateful to Simon Broda, Eglė Grublytė, Vasilis Sarafidis, Joakim Westerlund, and participants at Deakin University, University of Melbourne, Tinbergen Institute Amsterdam, University of Amsterdam, University of Groningen, Utrecht University, NESG 2014 (Tilburg), IAAE 2014 (London), and Bristol Econometric Study Group 2014 for insightful comments and discussions. The author is also grateful to Alberto Basantes-Arias for research assistance with the ENEMDU dataset. Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged.

[Received November 2014. Revised November 2015.]

REFERENCES

Abadir, K. M., and Magnus, J. R. (2002), "Notation in Econometrics: A Proposal for a Standard," *Econometrics Journal*, 5, 76–90. [47]

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001), "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219–255. [49]

——— (2013), "Panel Data Models With Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1–14. [47,49,50,51,52,54,56]

Antman, F., and McKenzie, D. J. (2007), "Earnings Mobility and Measurement Error: A Pseudo Panel Approach," *Economic Development and Cultural Change*, 56, 125–161. [55]

Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [47]

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–681. [48]

Bekker, P. A., and van der Ploeg, J. (2005), "Instrumental Variable Estimation Based On Grouped Data," *Statistica Neerlandica*, 59, 239–267. [48]

Dargay, J. (2007), "The Effect of Prices and Income on Car Travel in the UK," *Transportation Research Part A*, 41, 949–960. [48]

Deaton, A. (1985), "Panel Data From Time Series of Cross-Sections," *Journal of Econometrics*, 30, 109–126. [47,48]

Dovonon, P., and Renault, E. (2009), "GMM Overidentification Test With First Order Underidentification," mimeo. [54]

González, R., and Sala, H. (2015), "The Frisch Elasticity in the Mercosur Countries: A Pseudo-Panel Approach," *Development Policy Review*, 33, 107–131. [55,56]

Inoue, A. (2008), "Efficient Estimation and Inference in Linear Pseudo-Panel Data Models," *Journal of Econometrics*, 142, 449–466. [47,48,49,50,52,53,55]

Kiviet, J. F. (2012), *Monte Carlo Simulation for Econometricians* (*Foundations and Trends in Econometrics,* Vol. 5), NOW the Essence of Knowledge, London/New York: Delft. [53]

Kleibergen, F. R. (2005), "Testing Parameters in GMM Without Assuming that They are Identified," *Econometrica*, 73, 1103–1123. [50]

Kuersteiner, G., and Prucha, I. R. (2013), "Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity," *Journal of Econometrics*, 174, 107–126. [50]

——— (2015), "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," CESifo Working Paper No. 5445. [50]

McKenzie, D. J. (2001), "Estimation of AR(1) Models With Unequally Spaced Pseudo-Panels," *Econometrics Journal*, 4, 89–108. [48]

——— (2004), "Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels," *Journal of Econometrics*, 120, 235–262. [47,48,49,51,56]

Moffitt, R. (1993), "Identification and Estimation of Dynamic Models With a Time Series of Repeated Cross-Sections," *Journal of Econometrics*, 59, 99–123. [47]

Moon, H. R., and Weidner, M. (2015), "Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579. [50]

Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. 4), eds. J. Heckman, and E. Leamer, Amsterdam: Elsevier, pp. 2111–2245 [50]

Pesaran, H. M. (2006), "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [47,52]

Peterman, W. B. (2014), "Reconciling Micro and Macro Estimates of the Frisch Labor Supply Elasticity: A Sensitivity Analysis," mimeo. [55]

Robertson, D., and Sarafidis, V. (2015), "IV Estimation of Panels With Factor Residuals," *Journal of Econometrics*, 185, 526–541. [50,51]

Sarafidis, V., and Wansbeek, T. J. (2012), "Cross-Sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531. [47]

Sarafidis, V., Yamagata, T., and Robertson, D. (2009), "A Test of Cross Section Dependence for a Linear Dynamic Panel Model With Regressors," *Journal of Econometrics*, 148, 149–161. [47]

Verbeek, M. (2008), "Pseudo-Panels and Repeated Cross-Sections," in *The Econometrics of Panel Data (Advanced Studies in Theoretical and Applied Econometrics,* Vol. 46), eds. L. Matyas, P. Sevestre, J. Marquez, A. Spanos, F. Adams, P. Balestra, M. Dagenais, D. Kendrick, J. Paelinck, R. Pindyck, and W. Welfe, Springer Verlag. [47,48,50,55,56]

Verbeek, M., and Nijman, T. (1992), "Can Cohort Data Be Treated As Genuine Panel Data?" *Empirical Economics*, 17, 9–23. [47]

Verbeek, M., and Vella, F. (2005), "Estimating Dynamic Models From Repeated Cross-Sections," *Journal of Econometrics*, 127, 83–102. [47,55]

Ziliak, J. P. (1997), "Efficient Estimation With Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business & Economic Statistics*, 15, 419–431. [55]