Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 449–466
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Efficient estimation and inference in linear pseudo-panel data models

Atsushi Inoue

Department of Economics, #997-1873 East Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z1

Available online 15 August 2007

Abstract

We consider pseudo-panel data models constructed from repeated cross sections in which the number of individuals per group is large relative to the number of groups and time periods. First, we show that, when time-invariant group fixed effects are neglected, the OLS estimator does not converge in probability to a constant but rather to a random variable. Second, we show that, while the fixed-effects (FE) estimator is consistent, the usual t statistic is not asymptotically normally distributed, and we propose a new robust t statistic whose asymptotic distribution is standard normal. Third, we propose efficient GMM estimators using the orthogonality conditions implied by grouping and we provide t tests that are valid even in the presence of time-invariant group effects. Our Monte Carlo results show that the proposed GMM estimator is more precise than the FE estimator and that our new t test has good size and is powerful.

JEL classification: C12; C13; C23

Keywords: Fixed effects; Instrumental variables; Time series of cross sections

1. Introduction

Some survey data are not genuine panel data but are repeated cross sections. One problem with using repeated cross section data is that one cannot account for fixed effects (FE) and include a lagged dependent variable because specific individuals are not tracked over time. Since Deaton (1985) this problem has been dealt with by constructing pseudo-panel data from group averages of repeated cross sections of individual data. In this paper we will consider pseudo-panel data models in which the number of groups and time periods are small relative to the number of individuals per group. This situation is relevant for applied work. For example, Browning et al. (1985) use pseudo-panel data with eight cohorts, seven time periods and average cohort size 192 households, and Blundell et al. (1993) use pseudo-panel data with 10 cohorts, four time periods and average cohort size 364. When the number of individuals per group is large, applied researchers tend to treat pseudo-panel data as though they were genuine panels and employ conventional econometric methods for panel data, such as the FE estimator. In this paper, we will show that the FE estimator is not necessarily efficient and that the usual t test is not asymptotically normally distributed in this context. Moreover we will

E-mail address: atsushi_inoue@ncsu.edu

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.08.003

propose an efficient generalized method of moments (GMM) estimator and provide asymptotically valid t tests and a J test of overidentifying restrictions.

There is a rapidly growing literature on pseudo-panel data analysis (see Baltagi, 2001, pp. 189–193; Verbeek, 1995). One branch of the literature analyzes pseudo-panel data models in which the number of individuals per group is small relative to the number of groups or time periods. These papers address the measurement error problem introduced by replacing the population group mean by the sample group average. For example, Deaton (1985) and Verbeek and Nijman (1993) study static pseudo-panel data models and Collado (1997) considers dynamic pseudo-panel data models in this situation.

Another branch of the literature analyzes pseudo-panel data in which the number of individuals per group is large relative to the number of groups and time periods, which is the main focus of our paper. When the number of individuals per group is large, the measurement error will disappear. When dynamic panel data are constructed from repeated cross sections, however, the individuals for which the group mean of lagged dependent variables is computed are different from those for which the group mean of dependent variables is computed. Moffitt (1993) proposes a two-stage least squares estimator to address this issue. Moffitt's (1993) estimator is inconsistent when time-varying exogenous regressors are used (Verbeek and Vella, 2005). Girma (2000) proposes a quasi-differencing approach that takes quasi-difference of the dependent variables of two individuals in the same group and uses explanatory variables of a third individual in the group as instruments. Using time-invariant instruments, Verbeek and Vella (2005) propose an augmented instrumental variables (IV) approach unifying the existing estimators for dynamic pseudo-panel data models, and present identification conditions underlying these estimators. In related work, McKenzie (2004) considers the FE and Arellano-Bond type IV estimators allowing for heterogeneous parameters across groups in multidimensional asymptotics where the number of individuals and the number of groups can be both large. Such approximations would be useful when the number of groups is large relative to the number of individuals per group. None of these papers addresses the issue of efficient estimation and inference based on efficient estimation when group FE are present.[1] As pointed out by Moulton (1986) and more recently by Pepper (2002), neglecting the group effect can have serious effects on inference (see Wooldridge, 2003, for a survey of the more recent literature). In the present paper, we propose efficient estimators and methods of valid inference in the presence of a fixed number of time-invariant group FE. We also allow group sizes to vary across groups and over time. This is not done in many papers in the existing literature but is relevant for applied work because group sizes typically differ in repeated cross sections. This generalization also has implications for efficiency and inference.

Specifically, our paper makes several contributions. First, we study the consequence of neglecting time-invariant group effects in pseudo-panel data models. This is a common problem in applied work.[2] We show that, when the group effect is not taken into account, the OLS estimator does not converge in probability to a constant but rather to a random variable. Second, while the FE estimator is consistent, we show that the usual t statistic is not asymptotically normally distributed. When the number of individuals per group is large, practitioners tend to employ inference methods for genuine panel data. However, as we will show in the paper, that practice could cause asymptotically invalid inference. To ensure valid inference, we introduce a new robust t statistic whose asymptotic null distribution is standard normal. Third, we propose efficient GMM estimators for static and dynamic pseudo-panel data models using the orthogonality conditions implied by grouping and we provide a method of inference for the efficient GMM estimator of dynamic pseudo-panel data models. Our Monte Carlo results show that the proposed efficient GMM estimator is more precise than the FE estimator and that the t test based on the efficient GMM estimator has good size and is more powerful than the t test based on the FE estimator. As a by-product of GMM estimation, one can test the validity of group selection by Hansen's (1982) J test of overidentifying restrictions. Fourth, we consider

[1]Moffitt (1993) and Girma (2000) assume that there is no group effect. McKenzie (2004, Theorem 2) assumes $\mathrm{Var}(\delta_s) = 0$ in our notation when he derives the asymptotic distribution of the OLS and IV estimators and the consistency of the asymptotic covariance matrix estimator. Verbeek and Vella (2005) focus on identification and point estimation and do not have distributional results. The Monte Carlo experiments of Collado (1997), Girma (2000), McKenzie (2004) and Verbeek and Vella (2005) do not assess the accuracy of inference, such as the t test.

[2]In a recent paper Bertrand et al. (2004) list 92 papers that use difference-in-difference estimators and identify 80 papers that have potential clustering problems.

Hausman–Taylor type models for pseudo-panel data. For the expository purpose, we will assume that parameters are constant across groups and over time and focus mainly on the t test which is the leading example of inference in applied work. Extensions to heterogenous parameters and to other tests are straightforward, however.

It should be noted that the group FE are assumed to be time-invariant throughout the paper. Our approach based on the FE transformation is not valid when group effects are time-varying. The random effects or clustering approach (e.g., Moulton, 1986; Pepper, 2002) is not valid neither when the number of groups is fixed. While we cannot allow for arbitrarily time-varying group effects, we will briefly mention how one could deal with a particular time-varying group effect at the end of this paper.

The outline of the rest of the paper is as follows: Section 2 considers linear pseudo-panel data models created from static models with individual-specific and group-specific FE. Section 3 considers linear dynamic pseudo-data models. Section 4 evaluates the quality of the asymptotic approximation in Monte Carlo experiments. Section 5 considers extensions to Hausman–Taylor type models. Section 6 concludes the paper.

2. Static pseudo-panel data models

Consider a model with time-invariant individual-specific and group-specific FE:

$$y_i = \alpha_i + \delta_s + \beta' x_{st} + \gamma' z_i + \varepsilon_i \quad \text{for } i \in I_{N,st},\ \ s = 1, \ldots, S,\ \ t = 1, \ldots, T, \tag{1}$$

where α_i is the individual-specific fixed effect for individual i, δ_s is the time-invariant group-specific fixed effect for group s, x_{st} is a K-dimensional vector of group-time-specific explanatory variables, z_i is an L-dimensional vector of individual-specific explanatory variables, and $I_{N,st}$ is the set of individuals that belong to group s at time t.[3,4] Let $\bar{y}_{st} = (1/N_{st})\sum_{i \in I_{N,st}} y_i$, $\bar{\alpha}_{st} = (1/N_{st})\sum_{i \in I_{N,st}} \alpha_i$, $\bar{z}_{st} = (1/N_{st})\sum_{i \in I_{N,st}} z_i$, $\bar{\varepsilon}_{st} = (1/N_{st})\sum_{i \in I_{N,st}} \varepsilon_i$ where N_{st} is the number of individuals in $I_{N,st}$ and $N = \sum_{s=1}^{S}\sum_{t=1}^{T} N_{st}$. Then taking group-time averages of (1) yields a pseudo-panel data model

$$\bar{y}_{st} = \bar{\alpha}_{st} + \delta_s + \beta' x_{st} + \gamma' \bar{z}_{st} + \bar{\varepsilon}_{st} = \bar{\alpha}_{st} + \delta_s + \theta' \bar{w}_{st} + \bar{\varepsilon}_{st} \quad \text{for } s = 1, \ldots, S,\ \ t = 1, \ldots, T, \tag{2}$$

where $\theta = [\beta'\gamma']'$ and $\bar{w}_{st} = [x'_{st}\bar{z}'_{st}]'$. We consider environments in which the number of groups and time periods are both small relative to the number of individuals in each group. To approximate the finite-sample properties of estimators in such environments, we will use asymptotics in which $N_{st} \to \infty$ for $s = 1, \ldots, S$ and $t = 1, \ldots, T$ with S and T fixed.

There are two commonly used estimation methods. One is to inadvertently neglect the group effect and regress $\bar{y}_{st}$ on $\bar{w}_{st}$. The other is to treat the pseudo-panel data model (2) as a genuine panel data model and estimate θ by the FE estimator. We denote the OLS and FE estimators by $\hat{\theta}_{\text{OLS}}$ and $\hat{\theta}_{\text{FE}}$, respectively. Let

$$\mu_w = \begin{bmatrix} \mu_{w,1} \\ \vdots \\ \mu_{w,S} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1T} & \cdots & x_{S1} & \cdots & x_{ST} \\ v_{z,11} & \cdots & v_{z,1T} & \cdots & v_{z,S1} & \cdots & v_{z,ST} \end{bmatrix}',$$

where $v_{z,st}$ satisfies $\bar{z}_{st} - v_{z,st} \xrightarrow{p} 0$ as $N_{st} \to \infty$, $M = I_S \otimes (I_T - (1/T)\ell_T \ell'_T)$, $\bar{E}(a_i | i \in I_{st}) = \lim_{N_{st}\to\infty}(1/N_{st}) \sum_{i \in I_{N,st}} E(a_i)$, and $w_i = [x'_{st} z'_i]'$ for $i \in I_{st}$.[5] Then under the asymptotics we consider the consistency of the FE

[3] We do not use the subscript it notation that is common for genuine panel data, e.g., y_{it}, because the individual whose y_{it} is observed is different from the individual whose y_{it-1} is observed. In this paper, if two subscripts for individual-specific variables take the same value, then it means that they are the same individual. The set $I_{N,st}$ depends on the total sample size N because we will consider asymptotics in which the number of individuals in each group is increasing.

[4] Because the number of groups and the number of time periods are both fixed in our asymptotics, we could allow for β and γ to vary across groups (e.g., $\beta_1, \ldots, \beta_S$) or over time (e.g., $\gamma_1, \ldots, \gamma_T$) although such generalizations require stronger rank conditions. The same remark also applies the analysis of dynamic pseudo-panel data models in Section 3. To simplify the presentation, however, we will present the homogeneous case only.

[5] We take the limit of sample averages of expectations in the definition of $\bar{E}$ because $I_{N,st}$ depends on the sample size and $\{[y_i, z_i]\}_{i=1}^{N}$ is thus technically a triangular array.

estimator requires that

$$\mu_w' M \begin{bmatrix} \bar{E}(y_i - \theta' w_i | i \in I_{11}) \\ \vdots \\ \bar{E}(y_i - \theta' w_i | i \in I_{1T}) \\ \vdots \\ \bar{E}(y_i - \theta' w_i | i \in I_{S1}) \\ \vdots \\ \bar{E}(y_i - \theta' w_i | i \in I_{ST}) \end{bmatrix} = \mu_w' M \begin{bmatrix} \bar{E}(\alpha_i + \varepsilon_i | i \in I_{11}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{1T}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{S1}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{ST}) \end{bmatrix} = 0, \tag{3}$$

where the first equality follows from the FE transformation. Let $1(\cdot)$ denote the indicator function such that $1(A) = 1$ if expression A is true and $1(A) = 0$ otherwise. Since μ_w is arbitrary, (3) implies that the group selection variable $1(i \in I_{N,st})$ is orthogonal to the disturbance term $\varepsilon_i + \alpha_i$:

$$M \begin{bmatrix} \bar{E}(\alpha_i + \varepsilon_i | i \in I_{11}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{1T}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{S1}) \\ \vdots \\ \bar{E}(\alpha_i + \varepsilon_i | i \in I_{ST}) \end{bmatrix} = 0. \tag{4}$$

Thus the FE estimator can be viewed as a method of moments estimator based on $(K+L)$ linear combinations of the ST moment conditions (4). In this paper we propose an efficient GMM estimator based on (4). Because of the FE transformation, S of these ST moment conditions are redundant and the covariance matrix of the moment conditions (4) is singular. To handle the singular covariance matrix, one can either use the generalized inverse or delete one time period from the T time periods for each group. Whichever approach is taken and whichever time period is dropped in the second approach the resulting estimates are numerically identical as in the case of genuine panel data (see Im et al., 1999). Without loss of generality, we will drop the last time period. Then the proposed GMM estimator can be written as

$$\hat{\theta}_{\mathrm{GMM}} = (\dot{W}' \hat{\dot{\Omega}}^{-1} \dot{W})^{-1} \dot{W}' \hat{\dot{\Omega}}^{-1} \dot{\bar{y}}, \tag{5}$$

where $\dot{W}$ is the $S(T-1) \times (K+L)$ matrix obtained from deleting the Tth, $2T$th, …, STth rows of MW, $W = [\bar{w}_1', \ldots, \bar{w}_S']'$, $\bar{w}_s = [\bar{w}_{s1} \cdots \bar{w}_{sT}]'$, $\dot{\Omega}$ is the $S(T-1) \times S(T-1)$ matrix obtained from deleting the Tth, $2T$th, …, STth columns and rows of $M\hat{\Pi}^{-1}M$, $\hat{\Pi}$ is the $ST \times ST$ diagonal matrix whose diagonal elements are given by $N_{11}/N, \ldots, N_{1T}/N, \ldots, N_{S1}/N, \ldots, N_{ST}/N$, $\dot{\bar{y}}$ is the $S(T-1) \times 1$ vector obtained from deleting the Tth, $2T$th, …, STth rows of $M\bar{y}, \bar{y} = [\bar{y}_1, \ldots, \bar{y}_S]'$, and $\bar{y}_s = [\bar{y}_{s1}, \ldots, \bar{y}_{sT}]'$.

We impose the following conditions.

Assumption 1. (a) $\{[\alpha_i, \varepsilon_i]\}_{i=1}^N$ are iid with finite fourth moments.

(b) $E(\varepsilon_i | \{1(i \in I_{N,st})\}_{s=1,\ldots,S,t=1,\ldots,T}) = 0$ and $E(\alpha_i | \{1(i \in I_{N,st})\}_{s=1,\ldots,S,t=1,\ldots,T}) = E(\alpha_i) = 0$ for $s = 1, \ldots, S$, $t = 1, \ldots, T$ and $N = 1, 2, \ldots$.

(c) $\mathrm{Var}(\alpha_i + \varepsilon_i | \{1(i \in I_{N,st})\}_{s=1,\ldots,S,t=1,\ldots,T}) = \sigma^2$ for $N = 1, 2, \ldots$.

(d) For $s = 1, \ldots, S$ and $t = 1, \ldots, T$, $N_{st}/N \to \pi_{st} \in (0,1)$ as $N \to \infty$ where $N = \sum_{s=1}^S \sum_{t=1}^T N_{st}$.

(e) $v_{z,st}$ satisfies $\bar{z}_{st} - v_{z,st} \xrightarrow{p} 0$ as $N_{st} \to \infty$ for $s = 1, \ldots, S$ and $t = 1, \ldots, T$.

(f) $\mu_w' \mu_w$, $\mu_w' M \mu_w$ and $\dot{\mu}_w' \dot{\Omega}^{-1} \dot{\mu}_w$ are nonsingular with probability one for the OLS, FE and GMM estimators, respectively, where $\dot{\mu}_W$ and $\dot{\Omega}$ are obtained from μ_W and $M\Pi^{-1}M$ in the same way as $\dot{W}$ and $\hat{\dot{\Omega}}$ are from MW

and $M\hat{\Pi}^{-1}M$, respectively, and Π is the $ST \times ST$ diagonal matrix whose diagonal elements are given by $\pi_{11}, \ldots, \pi_{1T}, \ldots, \pi_{S1}, \ldots, \pi_{ST}$.

Remarks. While Assumption 1(a) assumes that $\alpha_i$and ε_i are iid, δ_s, x_{st} and z_i do not have to be and x_{st} and z_i can be even nonstationary, for example. Assumption 1(b) requires that the group selection variable be strictly exogenous. Unlike genuine panel data models, however, the explanatory variables can be endogenous. Assumption 1(b) also requires that the mean of the individual-specific FE in each group is constant across groups and over time. The assumption that $E(\alpha_i) = 0$ is without the loss of generality because the mean of δ_s is arbitrary. Assumption 1(c) is made to simplify the presentation and can be relaxed (see the fifth remark following Theorem 1). Assumption 1(d) postulates that the number of individuals per group goes to infinity. The vector $v_{z,st}$ in Assumption 1(e) can be either deterministic or stochastic. Assumption 1(f) requires that the group-specific explanatory variables and the means of the individual-specific explanatory variables have some variation across groups and time periods. Durbin (1954) interpreted the group selection variable as an instrument (see Angrist, 1991, pp. 249–250, for the further references). Assumptions 1(b) and (f) can be interpreted as the validity and relevance conditions for the group selection variable as an instrument, respectively. Although in this paper we take group selection as given, this interpretation provides some insight into how groups should be selected. If two heterogeneous groups are combined into a group, the FE transformation will not remove the combined group-specific fixed effect. Thus the resulting instrument will be invalid and the FE and GMM estimators will be inconsistent. On the other hand, if one homogeneous group is divided into two groups, the FE and GMM estimators will remain consistent provided that the identification condition holds but there will be no efficiency gain.

Theorem 1. *Suppose that Assumption* 1 *holds and that* $N \to \infty$.

(a)
$$\hat{\theta}_{\mathrm{OLS}} - \theta \xrightarrow{\mathrm{d}} (\mu_w' \mu_w)^{-1} \mu_w' (\delta \otimes \ell_T), \tag{6}$$
$$\sqrt{N}(\hat{\theta}_{\mathrm{FE}} - \theta) \xrightarrow{\mathrm{d}} \sigma(\mu_w' M \mu_w)^{-1} \mu_w M \Pi^{-1/2} \zeta, \tag{7}$$
$$\sqrt{N}(\hat{\theta}_{\mathrm{GMM}} - \theta) \xrightarrow{\mathrm{d}} \sigma[\dot{\mu}_w \dot{\Omega}^{-1} \dot{\mu}_w]^{-1} \dot{\mu}_w \dot{\Omega}^{-1} \dot{\zeta}, \tag{8}$$

where $\delta = [\delta_1, \ldots, \delta_S]'$, ζ *is the ST-dimensional standard normal random vector and* $\dot{\zeta}$ *is obtained from deleting the Tth,* $2T$*th, …, STth rows of* $M\Pi^{-1/2}\zeta$.

(b) *Under the null hypothesis that* $\theta_j = \theta_{0,j}$,

$$t_{\hat{\theta}_{\mathrm{OLS},j}} = \frac{\hat{\theta}_{\mathrm{OLS},j} - \theta_{0,j}}{\sqrt{\hat{\sigma}^2_{\mathrm{OLS}}[(W'W)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \frac{\eta_j}{\sqrt{\frac{1}{ST-K-L}\sum_{s=1}^{S}\sum_{t=1}^{T}(\delta_s + \eta'\mu_{w,st})^2[(\mu_w'\mu_w)^{-1}]_{jj}}},$$

$$t_{\hat{\theta}_{\mathrm{FE},j}} = \frac{\hat{\theta}_{\mathrm{FE},j} - \theta_{0,j}}{\sqrt{\hat{\sigma}^2_{\mathrm{FE}}[(W'MW)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \frac{[(\mu_w' M \mu_w)^{-1}\mu_w' M \Pi^{-1/2}\zeta]_j / \sqrt{[(\mu_w' M\mu_w)^{-1}]_{jj}}}{\sqrt{\frac{1}{S(T-1)-K-L}\zeta'\Pi^{-1/2}(M - M\mu_w(\mu_w' M\mu_w)^{-1}\mu_w' M)\Pi^{-1/2}\zeta}},$$

$$t^{\mathrm{robust}}_{\hat{\theta}_{\mathrm{FE},j}} = \frac{\sqrt{N}(\hat{\theta}_{\mathrm{FE},j} - \theta_{0,j})}{\sqrt{\hat{\sigma}^2_{\mathrm{GMM}}[(W'MW)^{-1}W'M\hat{\Pi}^{-1}MW(W'MW)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1),$$

$$t_{\hat{\theta}_{\mathrm{GMM},j}} = \frac{\sqrt{N}(\hat{\theta}_{\mathrm{GMM},j} - \theta_{0,j})}{\sqrt{\hat{\sigma}^2_{\mathrm{GMM}}[(\dot{W}'\hat{\dot{\Omega}}^{-1}\dot{W})^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1),$$

where $\hat{\sigma}^2_{\mathrm{OLS}} = (1/(ST-K-L))\sum_{s=1}^{S}(\bar{y}_s - \bar{w}_s\hat{\theta}_{\mathrm{OLS}})'(\bar{y}_s - \bar{w}_s\hat{\theta}_{\mathrm{OLS}})$, $\hat{\sigma}^2_{\mathrm{FE}} = (1/(S(T-1)-K-L))\sum_{s=1}^{S}(\bar{y}_s - \bar{w}_s\hat{\theta}_{\mathrm{FE}})'M(\bar{y}_s - \bar{w}_s\hat{\theta}_{\mathrm{FE}})$, $\hat{\sigma}^2_{\mathrm{GMM}} = (1/S)\sum_{s=1}^{S}\{(1/N_s)\sum_{i\in I_s}(y_i - \hat{\theta}'_{\mathrm{FE}}w_i)^2 - [(1/N_s)\sum_{i\in I_s}(y_i - \hat{\theta}'_{\mathrm{FE}}w_i)]^2\}$, $I_s = \bigcup_{t=1}^{T} I_{st}$, $N_s = \sum_{t=1}^{T} N_{st}$, η *is the* $(K+L)$*-dimensional random vectors on the RHS of* (6), *and* $\mu_{w,s} = [(1/T)\sum_{t=1}^{T}x_{st}'(1/T)\sum_{t=1}^{T}\mu_{z,ts}']'$.

(c) *If* $S(T-1)-K-L>0$ *then*

$$J = N(\dot{\bar{y}} - \dot{W}\hat{\theta}_{\mathrm{GMM}})'\hat{\tilde{\Omega}}^{-1}(\dot{\bar{y}} - \dot{W}\hat{\theta}_{\mathrm{GMM}})/\hat{\sigma}^2_{\mathrm{GMM}} \xrightarrow{\mathrm{d}} \chi^2_{S(T-1)-K-L}.$$

Remarks. (1) Theorem 1(a) shows that the OLS estimator is inconsistent. In a recent paper, Andrews (2003, Corollary 4) provides necessary and sufficient conditions for the OLS estimator to be consistent allowing for general cross-sectional dependence. Even if Andrews' (2003) necessary and sufficient conditions (Assumptions CU and CMZ) are satisfied, however, the OLS estimator is inconsistent in our case. This is because Andrews' results require $S \to \infty$, not $N \to \infty$.

(2) While the FE and GMM estimators are consistent, the unconditional asymptotic distributions of the these estimators are not normal because μ_w is random even in the limit. We will illustrate this point by simulation in Section 4.

(3) $t_{\hat{\theta}_{\mathrm{FE},j}}$ is the t statistic when the pseudo-panel data are treated as genuine panel data. Theorem 1(b) shows that its asymptotic distribution is not standard normal and thus the resulting inference can be misleading. When each group has the same size, i.e., $\Pi = (1/ST)I_{ST}$, one can show that the limit distribution is a t distribution with $S(T-1)-K-L$ degrees of freedom. Otherwise, the limit distribution is nonstandard. When $\hat{\sigma}^2_{\mathrm{FE}}$ is replaced by a consistent estimator for σ^2, such as $\hat{\sigma}^2_{\mathrm{GMM}}$, the asymptotic distribution becomes standard normal.

(4) $\hat{\theta}_{\mathrm{FE}}$ and $\hat{\theta}_{\mathrm{GMM}}$ are GMM estimators based on the same moment conditions (4). Because $\sigma^2 M\Pi^{-1}M$ is the covariance matrix of the moment conditions (4), $\hat{\theta}_{\mathrm{GMM}}$ is the GMM estimator with optimal weighting matrix. That is why our GMM estimator is more efficient than the FE estimator conditional on μ_w.

(5) Theorem 1(c) shows the asymptotic distribution of Hansen's (1982) J test of overidentifying restrictions when (4) holds. The test will be useful when one is interested in the validity of group selection. When two heterogenous groups are combined in a group, the test is expected to reject the null of overidentifying restrictions with probability approaching one as the sample size increases.

(6) When the variance of $\alpha_i + \varepsilon_i$ varies across groups and over time, the GMM estimator can be modified as follows:

$$\hat{\beta}_{\mathrm{GMM}} = (\dot{W}'\hat{\tilde{\Omega}}^{-1}\dot{W})^{-1}\dot{W}'\hat{\tilde{\Omega}}^{-1}\dot{\bar{y}}, \tag{9}$$

where $\hat{\tilde{\Omega}}$ is the $S(T-1)\times S(T-1)$ matrix obtained from deleting the Tth, $2T$th, ..., STth columns and rows of $M\tilde{\Sigma}^{-1}M$, $\tilde{\Sigma}$ is the $ST \times ST$ diagonal matrix whose $((s-1)T+t,(s-1)T+t)$th diagonal element is given by

$$\frac{N}{N_{st}}\left\{\frac{1}{N_{st}}\sum_{i\in N_{st}}(y_i - \hat{\theta}'_{\mathrm{FE}}w_i)^2 - \left[\frac{1}{N_{st}}\sum_{i\in N_{st}}(y_i - \hat{\theta}'_{\mathrm{FE}}w_i)\right]^2\right\}.$$

The GMM estimator (9) is consistent, asymptotically normally distributed and is efficient among the class of GMM estimators based on (4) even allowing for varying variances of $\alpha_i + \varepsilon_i$.

(7) In applied work it is common to consider time-specific FE. Our results will carry through with $M = I_S \otimes (I_T - (1/T)\ell_T\ell'_T)$ replaced by $(I_S - (1/S)\ell_S\ell'_S)\otimes(I_T - (1/T)\ell_T\ell'_T)$.

3. Dynamic pseudo-panel data models

Consider a model with time-invariant individual-specific and group-specific FE:

$$y_i = \alpha_i + \delta_s + \beta' x_{st} + \rho y_i^* + \gamma' z_i + \varepsilon_i \quad \text{for } i \in I_{N,st},\ s = 1,\ldots,S,\ t = 1,\ldots,T, \tag{10}$$

where y_i^* is the lagged dependent variable of individual i. In the pseudo-panel data, y_i^* is not available to the econometrician. Taking group-time averages of (10) yields

$$\bar{y}_{st} = \bar{\alpha}_{st} + \delta_s + \beta' x_{st} + \rho \bar{y}^*_{st-1} + \gamma' \bar{z}_{st} + \bar{\varepsilon}_{st} \quad \text{for } s = 1,\ldots,S,\ t = 1,\ldots,T. \tag{11}$$

Replacing $\bar{y}^*_{st-1}$ by $\bar{y}_{st-1}$, we obtain a dynamic pseudo-panel data model:

$$\bar{y}_{st} = \bar{\alpha}_{st} + \delta_s + \beta' x_{st} + \rho \bar{y}_{st-1} + \gamma' \bar{z}_{st} + \bar{u}_{st} \quad \text{for } s = 1, \ldots, S, \;\; t = 1, \ldots, T, \tag{12}$$

where $\bar{u}_{st} = \bar{\varepsilon}_{st} + \rho(\bar{y}^*_{st-1} - \bar{y}_{st-1})$.

In the standard dynamic panel data model, it is known that the FE estimator is inconsistent (Nickel, 1981) and IV estimators have been used (e.g., Anderson and Hsiao, 1981; Arellano and Bond, 1991). In the dynamic pseudo-panel data model, however, the FE estimator is consistent when $N \to \infty$ (McKenzie, 2004). The consistency of the FE estimator follows from $K + L + 1$ linear combinations of the following orthogonality conditions:

$$M \begin{bmatrix} \bar{E}(y_i - \beta' x_{11} - \gamma' z_i | i \in I_{11}) - \rho \bar{E}(y_j | j \in I_{10}) \\ \vdots \\ \bar{E}(y_i - \beta' x_{1T} - \gamma' z_i | i \in I_{1T}) - \rho \bar{E}(y_j | j \in I_{1,T-1}) \\ \vdots \\ \bar{E}(y_i - \beta' x_{S1} - \gamma' z_i | i \in I_{S1}) - \rho \bar{E}(y_j | j \in I_{S0}) \\ \vdots \\ \bar{E}(y_i - \beta' x_{ST} - \gamma' z_i | i \in I_{ST}) - \rho \bar{E}(y_j | j \in I_{S,T-1}) \end{bmatrix} = 0, \tag{13}$$

where the population weight is given by the transpose of the $ST \times (K + L + 1)$ matrix

$$\mu_w = \begin{bmatrix} x'_{11} & v_{y,10} & v'_{z,11} \\ \vdots & \vdots & \vdots \\ x'_{1T} & v_{y,1,T-1} & v'_{z,1T} \\ \vdots & \vdots & \vdots \\ x'_{S1} & v_{y,S0} & v'_{z,S1} \\ \vdots & \vdots & \vdots \\ x'_{ST} & v_{y,S,T-1} & v'_{z,ST} \end{bmatrix},$$

and $v_{y,st}$ and $v_{z,st}$ satisfy $\bar{y}_{st} - v_{y,st} \xrightarrow{p} 0$ and $\bar{z}_{st} - v_{z,st} \xrightarrow{p} 0$, respectively. If the group selection variable $1(i \in I_{st})$ is orthogonal to the individual-specific fixed effect α_i and to the disturbance ε_i, and if the means of y^*_i and y_j are the same for $i \in I_{N,st}$ and $j \in I_{N,s,t-1}$, then (13) holds even though the lagged dependent variable is not strictly exogenous in the individual-level FE regression model. Thus the FE estimator can be interpreted as a method-of-moments estimator based on $K + L + 1$ linear combinations of (13) with sample weight given by the transpose of

$$W = \begin{bmatrix} x'_{11} & \bar{y}_{10} & \bar{z}'_{11} \\ \vdots & \vdots & \vdots \\ x'_{1T} & \bar{y}_{1,T-1} & \bar{z}'_{1T} \\ \vdots & \vdots & \vdots \\ x'_{S1} & \bar{y}_{S0} & \bar{z}'_{S1} \\ \vdots & \vdots & \vdots \\ x'_{ST} & \bar{y}_{S,T-1} & \bar{z}'_{ST} \end{bmatrix}.$$

We propose a GMM estimator based on (13) with optimal weighting matrix that is the inverse of the covariance matrix of (13). Without loss of generality we drop the last time period. Then we obtain

$$\hat{\theta}_{\mathrm{GMM}} = (\ddot{W}'\hat{\ddot{\Omega}}^{-1}\ddot{W})^{-1}\ddot{W}'\hat{\ddot{\Omega}}^{-1}\ddot{\bar{y}}, \tag{14}$$

where $\ddot{W}$ and $\ddot{\bar{y}}$ are defined as $\dot{W}$ and $\dot{\bar{y}}$ in Section 2, $\hat{\ddot{\Omega}}$ is the $S(T-1)\times S(T-1)$ matrix obtained from deleting the Tth row, $2T$th, ..., STth rows of $M\hat{\Sigma}M$, $\hat{\Sigma}$ is the $ST \times ST$ tridiagonal matrix with $((s-1)T+t,(s-1)T+t)$th diagonal element given by

$$\frac{N}{N_{st}}\left[\frac{1}{N_{st}}\sum_{i\in I_{st}}(y_i-\hat{\theta}'w_i)^2-\left(\frac{1}{N_{st}}\sum_{i\in I_{st}}(y_i-\hat{\theta}'w_i)\right)^2\right]+\frac{\hat{\rho}^2 N}{N_{st-1}}\left[\frac{1}{N_{st-1}}\sum_{j\in N_{st-1}}y_j^2-\left(\frac{1}{N_{st-1}}\sum_{j\in N_{st-1}}y_j\right)^2\right] \tag{15}$$

and $((s-1)T+t,(s-1)T+t-1)$th and $((s-1)T+t-1,(s-1)T+t)$th elements given by

$$-\frac{\hat{\rho}N}{N_{st-1}}\left[\frac{1}{N_{st-1}}\sum_{i\in N_{st}}(y_i-\hat{\theta}'w_i)y_i-\frac{1}{N_{st-1}}\sum_{i\in N_{st}}(y_i-\hat{\theta}'w_i)\frac{1}{N_{st-1}}\sum_{i\in N_{st}}y_i\right], \tag{16}$$

where $\hat{\theta}=[\hat{\beta},\hat{\gamma},\hat{\rho}]'$ is some consistent estimator of θ, such as the FE estimator. We assume the following conditions.

Assumption 2. (a) $\{[\alpha_i,\varepsilon_i]\}_{i=1}^N$ are iid with finite fourth moments.

(b) $E(\varepsilon_i|\{1(i\in I_{st})\}_{s=1,\dots,S,\ t=1,\dots,T})=0$ and $E(\alpha_i|\{1(i\in I_{st})\}_{s=1,\dots,S,\ t=1,\dots,T})=E(\alpha_i)=0$.

(c) $\sqrt{N}(\bar{\alpha}+\rho(\bar{y}^*-\bar{y}_{-1})+\bar{\varepsilon})\xrightarrow{d}\mathrm{N}(0,\Sigma)$ where Σ is an $ST\times ST$ tridiagonal matrix such that $((s-1)T+t,(s-1)T+t)$th diagonal element given by

$$\frac{1}{\pi_{st}}\overline{Var}(y_i-\theta'w_i|i\in I_{st})+\frac{\rho^2}{\pi_{st-1}}\overline{Var}(y_j|j\in I_{st-1}) \tag{17}$$

and $((s-1)T+t,(s-1)T+t-1)$th and $((s-1)T+t-1,(s-1)T+t)$th elements given by

$$-\frac{\rho}{\pi_{st-1}}\overline{Cov}(y_i-\theta'w_i,y_i|i\in I_{st-1}), \tag{18}$$

where $\bar{\alpha}=[\bar{\alpha}_{11},\dots,\bar{\alpha}_{1T},\dots,\bar{\alpha}_{S1},\dots,\bar{\alpha}_{ST}]'$, $\bar{y}^*=[\bar{y}^*_{10},\dots,\bar{y}^*_{1,T-1},\dots,\bar{y}^*_{S0},\dots,\bar{y}^*_{S,T-1}]'$, $\bar{y}_{-1}=[\bar{y}_{10},\dots,\bar{y}_{1,T-1},\dots,\bar{y}_{S0},\dots,\bar{y}_{S,T-1}]'$, $\overline{Var}(a_i|i\in I_{st})=\lim_{N_{st}\to\infty}(1/N_{st})\sum_{i\in I_{N,st}}a_i^2-((1/N_{st})\sum_{i\in I_{N,st}}a_i)^2$ and

$$\overline{Cov}(a_i,b_i|i\in I_{st})=\lim_{N_{st}\to\infty}\frac{1}{N_{st}}\sum_{i\in I_{N,st}}a_ib_i-\frac{1}{N_{st}^2}\left(\sum_{i\in I_{N,st}}a_i\right)\left(\sum_{i\in I_{N,st}}b_i\right).$$

(d) $N_{st}/N\to\pi_{st}\in(0,1)$ as $N\to\infty$ for $s=1,\dots,S$ and $t=0,\dots,T$, where $N=\sum_{s=1}^S\sum_{t=0}^T N_{st}$.

(e) $v_{y,s,t-1}$ and $v_{x,st}$ satisfy $\bar{y}_{st-1}-v_{y,st-1}\xrightarrow{p}0$ $\bar{y}^*_{st}-v_{y,st-1}\xrightarrow{p}0$, and $\bar{z}_{st}-v_{z,st}\xrightarrow{p}0$ as $N_{st},N_{s,t-1}\to\infty$.

(f) $\mu_w'M\mu_w$, $\mu_w'M\Sigma M\mu_w$, $\ddot{\mu}_w'\ddot{\Omega}^{-1}\ddot{\mu}_w$ are nonsingular with probability one, where $\ddot{\mu}_W$ and $\ddot{\Omega}$ are obtained from μ_W and $M\Sigma^{-1}M$ in the same way as $\dot{\mu}_w$ and $\dot{\Omega}$ in Section 2.

(g) The initial estimator $\hat{\theta}$ used in $\hat{\Sigma}$ is consistent.

Remarks. (1) In the appendix we prove that Assumption 2(c) follows from the central limit theorem under Assumptions 1(c), 2(a), (b), (d) and (e) and more primitive assumptions:

(h) Conditional on $1(i\in I_{N,st})$, $\{y_i^*,z_i\}_{i\in I_{N,st}}$ is iid with finite fourth moments for $s=1,\dots,S$, $t=1,\dots,T$ and $N=1,2,\dots$.

(i) Conditional on $1(i\in I_{N,st})$ and $1(j\in I_{N,s,t-k})$, $\{y_i^*\}_{i\in I_{N,st}}$ and $\{y_j\}_{j\in I_{N,s,t-1}}$ are independent for $k\neq 0$, $s=1,\dots,S$, $t=1,\dots,T$ and $N=1,2,\dots$.

Assumption 2(h) allows that y_i^* and z_i are unconditionally correlated but requires that they are uncorrelated once conditioned on the group effect. Assumption 2(i) postulates that the data consists of independent cross sections.

(2) Assumption 2(f) is an identification condition and requires that the group-specific explanatory variables and the means of the lagged dependent variable and the individual-specific explanatory variables vary across groups and over time. Although it is likely to be satisfied in many applications, Assumption 2(f) can fail in some situations. For example, if $\beta = \gamma = 0$, $|\rho| < 1$ and the mean of the initial condition of y_i conditional on α_i and δ_s is set to $(\alpha_i + \delta_s)/(1-\rho)$, then the $(K+1)$st column of μ_w is $\ell_T \otimes \delta/(1-\rho)$. Because the FE transformation will replace this column by zeros, $\mu_w' M \mu_w$ becomes singular. The identification condition also fails when $\rho = 0$. When $\rho = 0$, the $(K+1)$st column of μ_w is $\ell_T \otimes \delta + \mu_w \theta$. After the FE transformation, this column is a linear combination of the other $K+L$ columns and thus $\mu_w' M \mu_w$ becomes singular.

(3) Some existing estimators require different identification conditions. Girma (2000) proposes a quasi-differencing approach that takes quasi-difference of the dependent variables of two individuals in a group and uses variables of a third individual in the same group as instruments. The identification condition of Girma (2000) requires that the individual-specific explanatory variables are correlated in each group, which appears to be a strong requirement. For example, if individual income is an explanatory variable and states are groups, the identification condition of Girma's (2000) estimator requires that individual incomes are correlated within a state whereas the identification condition of McKenzie's (2004) and ours requires that the state average incomes have some variation across states and over time.

(4) Verbeek and Vella (2005) require individual-specific time-invariant instruments that are uncorrelated with α_i and ε_i but are correlated with x_{ts}, y_i^* and z_i. When the group selection variable is the only instrument, their estimator simplifies to the FE estimator (see Verbeek and Vella, 2005, p. 88). This version of their estimator and our GMM estimator is based on the same moment conditions (13). Because our GMM estimator is a GLS version of the FE estimator, our GMM estimator is more efficient than the FE estimator.

Theorem 2. *Suppose that Assumption 2 holds.*

(a)
$$\sqrt{N}(\hat{\theta}_{\mathrm{FE}} - \theta) \xrightarrow{\mathrm{d}} (\mu_w' M \mu_w)^{-1} \mu_w' M \Sigma^{1/2} \zeta, \tag{19}$$

$$\sqrt{N}(\hat{\theta}_{\mathrm{GMM}} - \theta) \xrightarrow{\mathrm{d}} (\ddot{\mu}_w' \ddot{\Omega}^{-1} \ddot{\mu}_w)^{-1} \ddot{\mu}_w' \ddot{\Omega}^{-1} \ddot{\zeta}. \tag{20}$$

where $\ddot{\zeta}$ is obtained from deleting the Tth, 2Tth, …, STth rows of $M\Sigma^{1/2}\zeta$.

(b) *Under the null hypothesis that $\theta_j = \theta_{0,j}$,*

$$t_{\hat{\theta}_{\mathrm{FE},j}} = \frac{\sqrt{N}(\hat{\theta}_{\mathrm{FE},j} - \theta_{0,j})}{\sqrt{[(W'MW)^{-1} W' M \hat{\Sigma}^{-1} MW(W'MW)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1), \tag{21}$$

$$t_{\hat{\theta}_{\mathrm{GMM},j}} = \frac{\sqrt{N}(\hat{\theta}_{\mathrm{GMM},j} - \theta_{0,j})}{\sqrt{[(\ddot{\mu}_w' \hat{\ddot{\Omega}}^{-1} \ddot{\mu}_w)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1). \tag{22}$$

(c) *If $S(T-1) - K - L - 1 > 0$ then*

$$J = N(\ddot{y} - \ddot{W}\hat{\theta}_{\mathrm{GMM}})'(\hat{\ddot{\Omega}})^{-1}(\ddot{y} - \ddot{W}\hat{\theta}_{\mathrm{GMM}}) \xrightarrow{\mathrm{d}} \chi^2_{S(T-1)-K-L-1}.$$

4. Monte Carlo experiments

In this section, we will assess the quality of the asymptotic approximations proposed in the previous sections in Monte Carlo experiments. First, consider static pseudo-panel data models. We set $\{\pi_{st}\}$ to ST uniform random variables that sum to one and set N_{st} to be the smallest integer that is larger than or equal to $\pi_{st}\bar{N}ST$ where $\bar{N} \in \{128, 256\}$ is a given average group size. We generate pseudo-panel data from

$$y_i = \alpha_i + \delta_s + \beta x_{st} + \gamma z_i + \varepsilon_i, \tag{23}$$

$$z_i = v_{z,st} + \varepsilon_{z,i}, \tag{24}$$

for $i \in I_{N,st},\ s = 1, \ldots, S, t = 1, \ldots, T$, where $\{\alpha_i\}$, $\{\varepsilon_i\}$ and $\{\varepsilon_{z,i}\}$ are independent zero-mean iid normal random variables, $K = L = 1,\ \beta = \gamma = 0,\ S = T = 8$. δ_s, $v_{z,st}$ and x_{st} have zero mean and are distributed as (i) iid normal (ii) iid log-normal and (iii) a Gaussian AR(1) process with slope coefficient 0.9 and initial condition drawn from the unconditional distribution. The variances of α_i, ε_i, δ_s, $v_{z,st}$ and $\varepsilon_{z,i}$ are set so that $Var(\alpha_i + \varepsilon_i + \delta_s) = 1,\ Var(z_i) = 1$ and

$$Var(\delta_s)/Var(\alpha_i + \varepsilon_i + \delta_s) \in \{0.25, 0.5\}, \tag{25}$$

$$Var(v_{z,st})/Var(z_i) \in \{0.25, 0.5\}. \tag{26}$$

The variance of x_{st} is set to one.

In Table 1 we compute the OLS, FE and proposed GMM estimators of β with the number of Monte Carlo iterations set to 2000 and report their medians (Med), median absolute deviations (MAD), root mean squared error (RMSE) and RMSE ratios and the rejection frequencies of the t tests for testing $\beta = 0$. As expected, the median bias of the FE and GMM estimators is close to zero, and the GMM estimator exhibits smaller variabilities than the FE estimator in terms of MAD and RMSE. The usual t test for the FE estimator is oversized especially when the group components follow the persistent AR(1) process. The proposed t test for the FE estimator and the t test based on the GMM estimator have good size regardless of the distribution and contribution of the group components. In terms of RMSE ratio the GMM estimator has efficiency gain over the FE estimator ranging from 26% to 72%. Perhaps it is surprising that the OLS estimator, which has been shown to be inconsistent in Section 2, has very small median bias and the t test appears to have good size in case (i). This is an artifact due to the normality of $\{\delta_s\}$, $\{x_{st}\}$ and $\{\mu_{z,st}\}$. When they are not iid normal, the size of the t test is significantly distorted. The results for γ are qualitatively similar and are not reported to save space.

Next, we consider dynamic pseudo-panel data models:

$$y_i = \alpha_i + \delta_s + \rho y_i^* + \beta x_{st} + \gamma z_i + \varepsilon_i, \tag{27}$$

for $i \in I_{st},\ s = 1, \ldots, S,\ t = 1, \ldots, T$ where $\rho = 0.9$ and the initial condition of y_i is set to zero. All the other aspects of the data-generating process are the same as the those for the static pseudo-panel data model.

In Table 2 we compute an Arellano–Bond type IV estimator with instrument $\bar{y}_{st-2}$, the FE estimator and the proposed GMM estimator of β with the number of Monte Carlo iterations set to 2000 and report their medians, MAD, RMSE, RMSE ratios and the rejection frequencies of the t tests for testing $\beta = 0$. The three estimators have relatively small bias, which is expected because they are all consistent estimators. The t test for the IV estimator is computed assuming that the data are genuine dynamic panel and is severely undersized. Although the conventional IV estimator for genuine dynamic panel is consistent, it has large variance. The proposed GMM estimator has smaller variance than the IV and FE estimators and the efficiency gain of the GMM estimator over the FE estimator ranges from 37% to 46%. The proposed t tests based on the FE and GMM estimators performs well. While the contributions of the group components (25) and (26) affect the RMSE of the estimators, they do not affect the size of the t tests much. The results for γ and ρ are similar and thus are not reported.

The new t tests based on the FE and GMM estimators have good size in the static and dynamic pseudo-panel data models. How about the power of these t tests? Fig. 1 shows the rejection frequencies of the new t tests based on the FE and GMM estimators for testing $\beta = 0$ when $\beta = 0, 0.01, 0.02, \ldots, 0.1$, $\gamma = 0$, $\bar{N} = 128$, δ_s, $v_{z,st}$ and x_{st} have zero mean and are distributed as the Gaussian AR(1) process with slope coefficient 0.9, $K = L = 1$, $S = T = 8$, the two variance ratios (25) and (26) are set to 0.5. The t test based on the GMM estimator is much more powerful than the t test based on the FE estimator. While we report the power results for this particular parametrization only to save space, the results are similar in all the other cases considered in Tables 1 and 2.

Finally we evaluate our theoretical observation that these estimators are not normally distributed even asymptotically. We generated 2000 realizations from (23) with average group size $\bar{N} = 1000$, $K = L = 1$, $S = T = 8$ and $\{\delta_s\}$, $\{x_{st}\}$, $\{\mu_{z,st}\}$ distributed as the Gaussian AR(1) processes. Fig. 2 shows the histograms the estimators of β. Table 3 reports the summary statistics of the empirical distribution of the three estimators of β.

Table 1
Static pseudo-panel data model estimators

$\bar{N}$	y	z	OLS estimator of β				FE estimator of β					GMM estimator of β				
			Med	MAD	RMSE	Standard t test	Med	MAD	RMSE	Standard t test	New t test	Med	MAD	RMSE	RMSE ratio	New t test
(i) IID normal																
128	0.25	0.25	0.002	0.039	2.885	0.050	0.000	0.011	0.801	0.050	0.044	0.000	0.007	0.482	0.602	0.043
128	0.25	0.50	0.002	0.039	2.885	0.050	0.000	0.011	0.804	0.050	0.045	0.000	0.007	0.482	0.600	0.044
128	0.50	0.25	0.001	0.055	3.989	0.050	0.000	0.009	0.654	0.050	0.044	0.000	0.006	0.394	0.602	0.043
128	0.50	0.50	0.001	0.055	3.988	0.049	0.000	0.009	0.657	0.050	0.045	0.000	0.006	0.394	0.600	0.044
256	0.25	0.25	−0.000	0.040	2.918	0.055	0.001	0.008	0.615	0.055	0.049	−0.000	0.005	0.337	0.549	0.044
256	0.25	0.50	−0.000	0.040	2.918	0.055	0.001	0.008	0.617	0.055	0.050	−0.000	0.005	0.337	0.547	0.044
256	0.50	0.25	−0.001	0.056	4.075	0.054	0.000	0.006	0.502	0.055	0.049	−0.000	0.004	0.276	0.549	0.044
256	0.50	0.50	−0.001	0.056	4.075	0.053	0.000	0.006	0.504	0.055	0.050	−0.000	0.004	0.276	0.547	0.044
(ii) IID log-normal																
128	0.25	0.25	0.138	0.138	8.941	0.802	0.001	0.011	0.989	0.059	0.040	0.000	0.008	0.604	0.611	0.044
128	0.25	0.50	0.135	0.135	8.790	0.784	0.001	0.012	0.991	0.059	0.041	0.000	0.008	0.604	0.610	0.045
128	0.50	0.25	0.194	0.194	12.613	0.834	0.000	0.009	0.807	0.059	0.040	0.000	0.006	0.493	0.611	0.044
128	0.50	0.50	0.191	0.191	12.399	0.820	0.000	0.009	0.809	0.059	0.041	0.000	0.006	0.494	0.610	0.045
256	0.25	0.25	0.139	0.139	9.020	0.816	0.000	0.008	0.759	0.059	0.051	−0.000	0.006	0.439	0.579	0.051
256	0.25	0.50	0.138	0.138	8.936	0.801	0.000	0.008	0.760	0.059	0.051	−0.000	0.006	0.439	0.578	0.051
256	0.50	0.25	0.196	0.196	12.733	0.833	0.000	0.007	0.620	0.059	0.051	−0.000	0.005	0.359	0.579	0.051
256	0.50	0.50	0.193	0.193	12.616	0.826	0.000	0.007	0.621	0.059	0.051	−0.000	0.005	0.359	0.578	0.051
(iii) Gaussian AR(1)																
128	0.25	0.25	0.001	0.067	5.644	0.372	0.000	0.014	1.385	0.106	0.040	−0.000	0.008	0.726	0.524	0.048
128	0.25	0.50	0.001	0.069	5.746	0.380	0.000	0.013	1.313	0.076	0.041	−0.000	0.009	0.734	0.559	0.050
128	0.50	0.25	0.001	0.095	7.908	0.382	0.000	0.011	1.131	0.106	0.040	−0.000	0.007	0.592	0.524	0.048
128	0.50	0.50	0.001	0.097	8.059	0.391	0.000	0.011	1.072	0.076	0.041	−0.000	0.007	0.600	0.559	0.050
256	0.25	0.25	−0.000	0.066	5.887	0.388	0.000	0.010	1.087	0.109	0.044	0.000	0.006	0.498	0.459	0.049
256	0.25	0.50	−0.000	0.068	5.950	0.396	0.000	0.009	1.007	0.081	0.045	0.000	0.006	0.501	0.497	0.051
256	0.50	0.25	−0.001	0.093	8.301	0.397	0.000	0.008	0.887	0.109	0.044	0.000	0.005	0.407	0.459	0.049
256	0.50	0.50	−0.001	0.094	8.393	0.402	0.000	0.008	0.822	0.081	0.045	0.000	0.005	0.409	0.497	0.051

Notes: The numbers in the columns labeled "Median", "MAD" and "RMSE" are the medians, median absolute deviations and root mean squared errors of the estimates, respectively. The numbers under "RMSE ratio" are the RMSE of the GMM estimator over the RMSE of the FE estimator. The numbers in the columns labeled "standard t test" and "new t test" are the rejection frequencies of the usual and proposed t tests at the 5% significance level. "(i) IID normal", "(ii) IID log-normal" and "(iii) Gaussian AR(1)" refer to the distribution of the group components. $\bar{N}$ is the average number of individuals per cell. The numbers in the columns labeled "y" and "z" are the contributions of the group components in (25) and (26), respectively.

Table 2
Dynamic pseudo-panel data model estimators

$\bar{N}$	y	z	IV estimator of β				FE estimator of β				GMM estimator of β				
			Med	MAD	RMSE	Standard t test	Med	MAD	RMSE	New t test	Med	MAD	RMSE	RMSE ratio	New t test
(i) IID normal															
128	0.250	0.250	−0.003	0.041	2.811	0.000	−0.000	0.027	1.808	0.055	0.000	0.016	1.085	0.600	0.053
128	0.250	0.500	−0.003	0.038	2.689	0.000	−0.000	0.025	1.690	0.056	0.000	0.015	0.979	0.579	0.055
128	0.500	0.250	−0.003	0.052	3.491	0.000	0.001	0.034	2.303	0.056	0.001	0.019	1.300	0.565	0.053
128	0.500	0.500	−0.003	0.049	3.311	0.000	0.001	0.033	2.211	0.053	0.001	0.018	1.207	0.546	0.052
256	0.250	0.250	0.001	0.026	1.983	0.000	−0.000	0.019	1.280	0.058	0.001	0.011	0.770	0.602	0.051
256	0.250	0.500	0.001	0.025	1.764	0.000	0.000	0.018	1.198	0.055	0.001	0.010	0.697	0.582	0.048
256	0.500	0.250	0.002	0.034	2.382	0.000	0.000	0.025	1.634	0.054	0.001	0.014	0.930	0.569	0.047
256	0.500	0.500	0.002	0.034	2.276	0.000	−0.000	0.024	1.571	0.056	0.001	0.013	0.867	0.552	0.043
(ii) IID log-normal															
128	0.250	0.250	−0.004	0.045	5.666	0.000	−0.002	0.028	2.207	0.062	0.000	0.018	1.372	0.622	0.057
128	0.250	0.500	−0.003	0.041	4.635	0.000	−0.002	0.026	2.063	0.054	0.001	0.016	1.246	0.604	0.057
128	0.500	0.250	−0.004	0.055	7.534	0.000	−0.001	0.036	2.810	0.057	0.001	0.022	1.663	0.592	0.056
128	0.500	0.500	−0.003	0.053	98.536	0.000	−0.002	0.035	2.696	0.053	0.001	0.020	1.555	0.577	0.054
256	0.250	0.250	−0.000	0.030	2.396	0.000	−0.000	0.021	1.580	0.050	0.001	0.013	0.989	0.626	0.052
256	0.250	0.500	0.001	0.028	2.204	0.000	−0.001	0.020	1.475	0.050	0.001	0.011	0.896	0.607	0.048
256	0.500	0.250	0.001	0.039	2.978	0.000	−0.001	0.027	2.008	0.049	0.002	0.015	1.197	0.596	0.048
256	0.500	0.500	0.001	0.037	2.835	0.000	−0.001	0.026	1.926	0.049	0.001	0.014	1.116	0.580	0.047
(iii) Gaussian AR(1)															
128	0.250	0.250	−0.007	0.102	7.079	0.000	−0.001	0.044	3.069	0.059	0.001	0.026	1.791	0.583	0.051
128	0.250	0.500	−0.005	0.094	6.659	0.000	−0.001	0.041	2.876	0.056	0.002	0.023	1.630	0.567	0.055
128	0.500	0.250	−0.004	0.129	8.921	0.000	−0.001	0.056	3.912	0.059	0.001	0.031	2.175	0.556	0.056
128	0.500	0.500	−0.005	0.122	8.425	0.000	−0.001	0.054	3.772	0.059	0.000	0.029	2.040	0.541	0.056
256	0.250	0.250	0.001	0.065	18.186	0.000	−0.001	0.030	2.127	0.063	0.000	0.018	1.237	0.582	0.050
256	0.250	0.500	0.002	0.062	4.522	0.000	−0.001	0.028	1.989	0.062	0.000	0.016	1.133	0.570	0.049
256	0.500	0.250	0.003	0.084	6.009	0.000	−0.000	0.039	2.713	0.061	−0.000	0.022	1.523	0.561	0.050
256	0.500	0.500	0.002	0.082	5.699	0.000	0.001	0.037	2.607	0.060	−0.000	0.021	1.433	0.550	0.052

Notes: See the notes for Table 1.

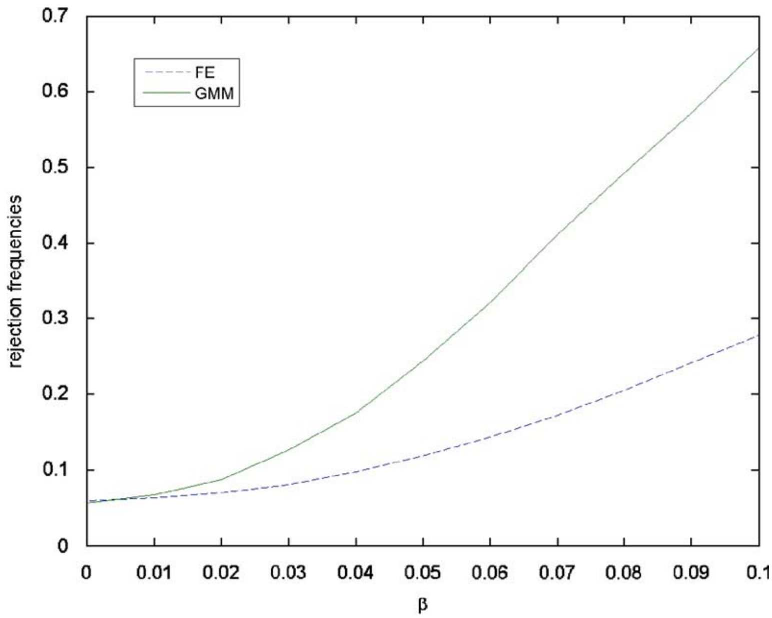

Fig. 1. Power of t tests based on the FE and GMM estimators. *Notes*: The rejection frequencies are based on the dynamic pseudo panel data model analyzed in Table 2 with (iii) Gaussian AR(1) group components, $\bar{N} = 128$, the variance ratios in (25) and (26) set to 0.5.

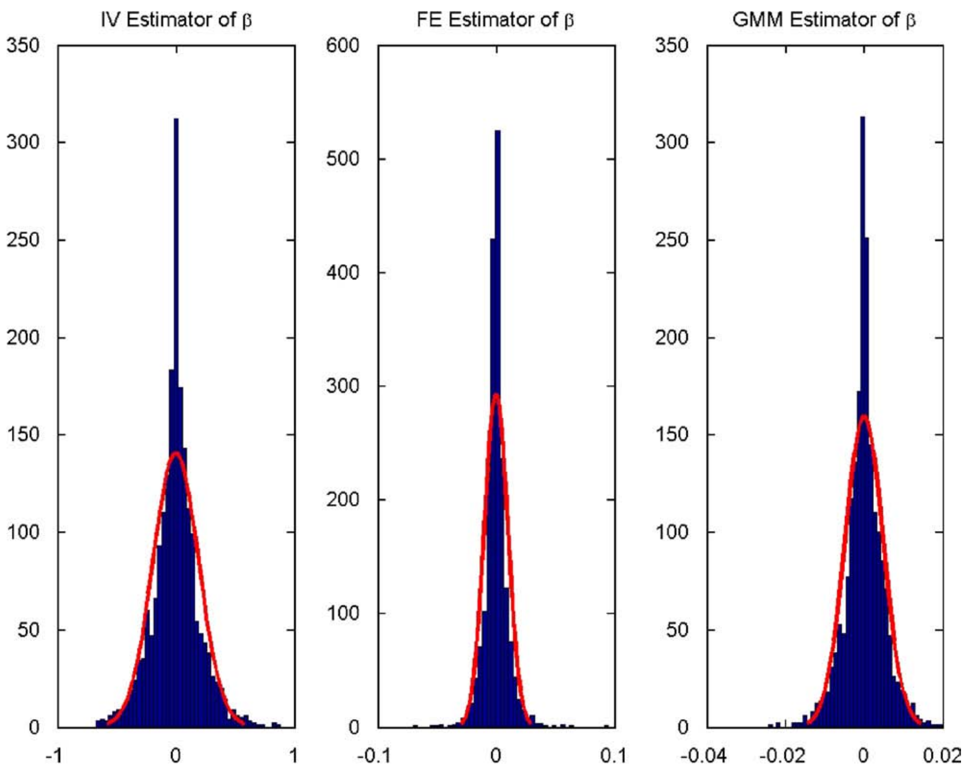

Fig. 2. Distributions of estimators of β.

All of these results indicate that, even for the large group size, these estimators are not normally distributed and their unconditional distributions are heavy-tailed. The results for the other parameters are qualitatively similar and are not reported.

Table 3
Summary statistics for distributions of estimators of β

	OLS	FE	GMM
Skewness	0.133	0.569	−0.150
Kurtosis	4.697	13.414	4.970
Jarque–Bera test	0.000	0.000	0.000
Kolmogorov–Smirnov test	0.000	0.000	0.000

Notes: The results are based on the static pseudo-panel data model analyzed in Table 1 with (iii) Gaussian AR(1) group components and the variance ratios in (25) and (26) set to 0.5 except that $\bar{N} = 1000$.

5. Pseudo-panel data models with time-invariant explanatory variables

Consider static and dynamic pseudo-panel data versions of the Hausman and Taylor type model:

$$y_i = \alpha_i + \delta_s + \beta' x_s + \gamma' z_i + \varepsilon_i \quad \text{for } i \in I_{N,st},\ \ s = 1, \ldots, S,\ \ t = 1, \ldots, T, \tag{28}$$

$$y_i = \alpha_i + \delta_s + \beta' x_s + \rho y_i^* + \gamma' z_i + \varepsilon_i \quad \text{for } i \in I_{N,st},\ \ s = 1, \ldots, S,\ \ t = 1, \ldots, T, \tag{29}$$

where x_s is a K-dimensional vector of time-invariant group-specific explanatory variables, such as birth year, education, gender and race. Because our results in Sections 2 and 3 carry through to the analysis of γ in (28) and γ and ρ in (29), respectively, we will focus on estimation of and inference about β.

We will consider the following two-step estimator that is a version of the estimators of Hausman and Taylor (1981) and Donald and Lang (2001). First, estimate γ and ρ by some consistent estimator, such as the FE and GMM estimators discussed in Section 2. Second, estimating β by regressing $\bar{y}_{st} - \hat{\gamma}'\bar{z}_{st}$ on x_s for (28) and $\bar{y}_{st} - \hat{\rho}\bar{y}_{s,t-1} - \hat{\gamma}'\bar{z}_{st}$ on x_s for (29). Suppose that the following assumptions hold:

Assumption 3. (a) The first-step estimator $\hat{\gamma}$ is consistent for (28). $\hat{\gamma}$ and $\hat{\rho}$ are consistent for (29).
(b) $X'X$ is nonsingular with probability one where $X = [x_1, x_2, \ldots, x_s]'$.

Theorem 3. *Suppose that Assumptions* 1(a), (b), (d) *and* 3 *hold and that* $N \to \infty$ *for* (28). *In addition suppose that Assumption* 2(e) *holds for* (29).

(a) $$\hat{\beta} - \beta \xrightarrow{\mathrm{d}} (X'X)^{-1}X'\delta. \tag{30}$$

(b) *Under the null hypothesis that* $\beta_j = \beta_{0,j}$,

$$t_{\beta_j} = \frac{\hat{\beta}_j - \beta_{0,j}}{\sqrt{\hat{\sigma}^2[(X'X)^{-1}]_{jj}}} \xrightarrow{\mathrm{d}} \frac{\eta_j}{\sqrt{(1/(ST-K))(\delta + X\eta)'(\delta + X\eta)[(X'X)^{-1}]_{jj}}}, \tag{31}$$

where $\delta = [\delta_1, \ldots, \delta_S]'$ *and* η *is the K-dimensional random vector on the RHS of* (30).

Remarks. (1) Theorem 3(a) shows that the second-step estimator is inconsistent and converges to a random variable. This is because the group selection variable is not orthogonal to the group effects and thus is an invalid instrument.

(2) Assuming the group effect is normally distributed, Donald and Lang (2001) show that the asymptotic distribution of $t_{\hat{\beta}_j}$ is a t distribution with $S-2$ degrees of freedom in the cross sectional case. When the group effect is not normal, however, our Theorem 3(b) shows that the asymptotic distribution is not even a t distribution and is nonstandard in general. Because $\hat{\beta}$ is inconsistent, the usual t test is invalid.

Because the asymptotic distribution of the t statistic depends nuisance parameters that cannot be consistently estimated, the standard bootstrap and simulation methods do not solve the inference problem. Any transformation that removes the nuisance parameter δ_s, such as the FE transformation and the first-difference transformation, will remove x_s (and thus β) and render inference on β infeasible. If we impose further assumptions, however, it is still possible to construct valid confidence regions of β. Note that, when the null hypothesis that $\beta = \beta_0$ is true, δ_s can be estimated by the first terms on the RHS of

$$\delta_s = (1/T)\sum_{t=1}^{T}(\bar{y}_{st} - \beta_0' x_s - \hat{\gamma}'\bar{z}_{st}) + o_p(1), \tag{32}$$

$$\delta_s = (1/T)\sum_{t=1}^{T}(\bar{y}_{st} - \beta_0' x_s - \hat{\rho}\bar{y}_{s,t-1} - \hat{\gamma}'\bar{z}_{st}) + o_p(1), \tag{33}$$

respectively, in (28) and (29).

Proposition 1. *Suppose that Assumptions* 1(a), (b), (d) *and* 3 *holds, that* $N \to \infty$ *and that* $\{\delta_s\}_{s=1}^{S}$ *is iid for* (28). *In addition suppose that Assumption* 2(e) *holds for* (29)

(a) *If the distribution of* δ_s *is symmetric around zero, then the sign statistics based on the first terms on the RHS of* (32) *and* (33) *are asymptotically pivotal under the null hypothesis that* $\beta = \beta_0$.
(b) *If* δ_s *and* x_s *are independent, then Spearman's rank correlation and Kendall's rank correlation between the first term on the RHS of* (32) [*or* (33)] *and* $\{x_s\}_{s=1}^{S}$ *are asymptotically pivotal under the null hypothesis that* $\beta = \beta_0$.

Remarks. These testing procedures test the joint hypothesis that $\beta = \beta_0$ and the symmetry of the distribution of δ_s or the joint hypothesis that $\beta = \beta_0$ and the independence between δ_s and x_s. When the null is rejected, it does not necessarily mean that $\beta \neq \beta_0$. The same comment applies to the procedure of Donald and Lang (2001), and our semiparametric assumptions are much weaker than their normality assumption, however.

6. Concluding remarks

In this paper we considered pseudo-panel data models constructed from repeated cross sections in which the number of individuals per group is large relative to the number of groups and time periods. There are several contributions of this paper. First, we showed that, when group effects are neglected, the OLS estimator does not converge in probability to a constant but rather to a random variable. Second, we showed that, while the FE estimator is consistent, the usual t statistic is not asymptotically normally distributed and we propose a new t statistic whose asymptotic distribution is standard normal. Third, we proposed efficient GMM estimators using the orthogonality conditions implied by grouping and we provide t tests that are valid even in the presence of time-invariant group effects and the J test of overidentifying restrictions. Lastly, we considered Hausman–Taylor type models for pseudo-panel data.

Our results suggest that uncritical application of the inference methods for genuine panels to pseudo-panels is potentially misleading. Our Monte Carlo results show that our proposed GMM estimator is more precise than the FE estimator and that the proposed t test constructed from the GMM estimator have good size and is more powerful than the t test from the FE estimator. The proposed GMM estimator does not involve nonlinear optimization and is easy to implement.

Because our proposed estimator is a GMM estimator, the usual identification conditions, such as orthogonality conditions and rank conditions, must be satisfied. Our orthogonality conditions can fail if there is sample selection in grouping or if two heterogenous groups are mixed together. As Verbeek and Vella (2005) point out, the identification condition required for pseudo-panel data models are demanding. If there is no significant difference in explanatory variables across groups, our identification condition is likely to fail and we will suffer from weak instruments (Staiger and Stock, 1997). Practitioners must be aware of these strong identification conditions when applying the proposed estimator. The author is currently investigating some of these important issues.

Throughout the paper the group fixed effect is assumed to be time-invariant. Nonetheless time-varying group effects are important in applied work. While our framework cannot allow for arbitrary time-varying group effects, one could consider time-invariant group effects whose coefficients are time-varying. For such group effects, one would need to use a transformation that is more general than the FE transformation to remove the group effects along the line of the GMM estimator of Ahn et al. (2001) for genuine panel data. Formally developing such GMM estimators for pseudo-panel data is beyond the scope of the paper and is left for future research.

Acknowledgments

I thank three anonymous referees, the co-editor, Lutz Kilian and Gary Solon for helpful conversations and seminar participants at the University of British Columbia, Indiana University, Bloomington, North Carolina State University, Ohio State University, Texas A&M University, the University of Tokyo, Yokohama National University, and the University of Virginia for useful comments.

Appendix

Proof of Theorem 1(a). Because $\bar{\alpha}_{st} + \bar{\varepsilon}_{st} = o_p(1)$ and $\bar{z}_{st} - v_{z,st} = o_p(1)$, we have

$$\hat{\theta}_{\mathrm{OLS}} = (W'W)^{-1}W'\bar{y} = \theta + (\mu_w'\mu_w)^{-1}\mu_w'(\delta \otimes \ell_T) + o_p(1), \tag{34}$$

from which the asymptotic distribution of $\hat{\theta}_{\mathrm{OLS}}$ follows. Let $\dot{\bar{\alpha}}$ and $\dot{\bar{\varepsilon}}$ be the $S(T-1) \times 1$ vectors obtained from deleting the Tth, $2T$th,…,STth rows of $M\bar{\alpha}$ and $M\bar{\varepsilon}$, respectively. Since

$$\begin{aligned}&\sqrt{N}(\hat{\theta}_{\mathrm{FE},j} - \theta) = (W'MW)^{-1}W'M(\bar{\alpha} + \bar{\varepsilon}), \sqrt{N}(\hat{\theta}_{\mathrm{GMM}} - \theta) = (\dot{W}'\hat{\dot{\Omega}}^{-1}\dot{W})^{-1}\dot{W}'\hat{\dot{\Omega}}^{-1}(\dot{\bar{\alpha}} + \dot{\bar{\varepsilon}}),\\ &W - \mu_w \xrightarrow{\mathrm{p}} 0, \sqrt{N}(\dot{\bar{\alpha}} + \dot{\bar{\varepsilon}}) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma^2\dot{\Omega}),\end{aligned} \tag{35}$$

the asymptotic distributions of $\hat{\theta}_{\mathrm{FE}}$ and $\hat{\theta}_{\mathrm{GMM}}$ follow. $\square$

Proof of Theorem 1(b). It follows from Theorem 1(a) that

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{OLS}} &= \frac{1}{ST-K-L}\sum_{s=1}^{S}\sum_{t=1}^{T}(\bar{y}_{st} - \hat{\theta}'_{\mathrm{OLS}}\bar{w}_{st})^2 = \frac{1}{ST-K-L}\sum_{s=1}^{S}\sum_{t=1}^{T}(\bar{\alpha}_{st} + \delta_s + \bar{\varepsilon}_s - (\hat{\theta}_{\mathrm{OLS}} - \theta)'\bar{w}_{st})^2\\ &= \frac{1}{ST-K-L}\sum_{s=1}^{S}\sum_{t=1}^{T}(\delta_s + \eta' v_{w,st})^2 + o_p(1).\end{aligned} \tag{36}$$

Thus the asymptotic null distribution of $t_{\hat{\theta}_{\mathrm{OLS},j}}$ follows. The asymptotic null distribution of $t_{\hat{\theta}_{\mathrm{FE},j}}$ follows from

$$\frac{\sqrt{N}(\hat{\theta}_{\mathrm{FE},j} - \theta)}{\sigma} \xrightarrow{\mathrm{d}} (W'MW)^{-1}W'M\Pi^{-1/2}\eta \tag{37}$$

and

$$\begin{aligned}&\frac{(S(T-1)-L)N\hat{\sigma}^2_{\mathrm{FE}}}{\sigma^2}\\ &= \sqrt{N}\frac{(\bar{\alpha} + \bar{\varepsilon} - W(\hat{\theta}_{\mathrm{FE}} - \theta))'}{\sigma}M(M\Pi^{-1}M)^{\dagger}M\sqrt{N}\frac{(\bar{\alpha} + \bar{\varepsilon} - W(\hat{\theta}_{\mathrm{FE}} - \theta))}{\sigma}\\ &= \frac{\sqrt{N}(\bar{\alpha} + \bar{\varepsilon})'}{\sigma}(M - MW(W'MW)^{-1}W'M)(M - MW(W'MW)^{-1}W'M)\frac{\sqrt{N}(\bar{\alpha} + \bar{\varepsilon})}{\sigma}\\ &\xrightarrow{\mathrm{d}} \zeta'\Pi^{-1/2}(M - M\mu_w(\mu_w'M\mu_w)^{-1}\mu_w'M)\Pi^{-1/2}\zeta,\end{aligned} \tag{38}$$

where ζ is the ST-dimensional standard normal random vector. The asymptotic null distribution of $t_{\hat{\theta}_{\mathrm{FE},j}}$ follows from (38) and

$$\frac{\sqrt{N}(\hat{\theta}_{\mathrm{FE},j} - \theta)}{\sigma} \xrightarrow{\mathrm{d}} (\mu_w'M\mu_w)^{-1}\mu_w'M\Pi^{-1/2}\eta. \tag{39}$$

The asymptotic null distribution of $t_{\hat{\theta}_{\mathrm{GMM},j}}$ follows from Theorem 1(a) and

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{GMM}} &= \frac{1}{N_s}\left\{\sum_{i\in N_s}(y_i-\hat{\theta}'_{\mathrm{FE}}w_i)^2-\left[(1/N_s)\sum_{i\in N_s}(y_i-\hat{\theta}'_{\mathrm{FE}}w_i)\right]^2\right\}\\ &= \frac{1}{N_s}\left\{\sum_{i\in N_s}(y_i-\theta' w_i)^2-\left[(1/N_s)\sum_{i\in N_s}(y_i-\theta' w_i)\right]^2\right\}+\mathrm{o_p}(1)\\ &= \frac{1}{N_s}\left\{\sum_{i\in N_s}(\alpha_i+\delta_s+\varepsilon_i)^2-\left[(1/N_s)\sum_{i\in N_s}(\alpha_i+\delta_s+\varepsilon_i)\right]^2\right\}+\mathrm{o_p}(1)=\sigma^2+\mathrm{o_p}(1). \quad \square \qquad (40)\end{aligned}$$

Proof of Theorem 1(c). Note that the limiting expression of the J statistic can be written as

$$J=(\dot{\Omega}^{-1/2}\dot{\zeta})'\{I_{S(T-1)}-\dot{\Omega}^{-1/2}\dot{\mu}_w[\dot{\mu}'_w\dot{\Omega}^{-1}\dot{\mu}_w]^{-1}\dot{\mu}'_w\dot{\Omega}^{-1/2}\}(\dot{\Omega}^{-1/2}\dot{\zeta})+\mathrm{o_p}(1), \qquad (41)$$

where ζ is the ST-dimensional standard normal random vector. Since $\dot{\Omega}^{-1/2}\dot{\zeta}\sim\mathrm{N}(0,I_{S(T-1)})$,

$$\{I_{S(T-1)}-\dot{\Omega}^{-1/2}\dot{\mu}_w[\dot{\mu}'_w\dot{\Omega}^{-1}\dot{\mu}_w]^{-1}\}$$

is idempotent and its trace is $S(T-1)-K-L$, we obtain Theorem 1(c). $\square$

Proof of Theorem 2(a)–(c). The proof for the asymptotic distribution of the FE estimator follows from Assumption 2 in a straightforward manner and thus is omitted. It follows from the consistency of the initial estimator $\hat{\theta}$ and the law of large numbers that (15) and (16) converge in probability to (17) and (18), respectively. Thus $\hat{\Sigma}\xrightarrow{p}\Sigma$. The rest of the proof for the GMM estimator is analogous to the proof of Theorem 1 and thus is omitted. $\square$

Proof of Remark 1 on Theorem 2. The asymptotic normality follows from Assumptions 2(a), (b), (d), (e) and (h) and the central limit theorem. It remains to show that the asymptotic covariance matrix is given by Σ. First consider the asymptotic variance of $\sqrt{N}(\bar{\alpha}_{(s-1)T+t}+\bar{\varepsilon}_{(s-1)T+t}+\rho(\bar{y}^*_{(s-1)T+t}-\bar{y}_{(s-1)T+t-1}))$. Since

$$\begin{aligned}&\sqrt{N}(\bar{\alpha}_{(s-1)T+t}+\bar{\varepsilon}_{(s-1)T+t}+\rho(\bar{y}^*_{(s-1)T+t}-\bar{y}_{(s-1)T+t-1}))\\ &\quad=\sqrt{N}(\bar{\alpha}_{(s-1)T+t}+\bar{\varepsilon}_{(s-1)T+t}+\rho(\bar{y}^*_{(s-1)T+t}-\mu_{y,(s-1)T+t-1}))-\sqrt{N}\rho(\bar{y}_{(s-1)T+t-1}-\mu_{y,(s-1)T+t-1})\end{aligned}$$

and the first and second terms on the RHS are independent by Assumption (i), its asymptotic variance is given by (17). Next consider the asymptotic covariance between $\sqrt{N}(\bar{\alpha}_{(s-1)T+t-1}+\bar{\varepsilon}_{(s-1)T+t-1}+\rho(\bar{y}^*_{(s-1)T+t-1}-\bar{y}_{(s-1)T+t-2}))$ and $\sqrt{N}(\bar{\alpha}_{(s-1)T+t}+\bar{\varepsilon}_{(s-1)T+t}+\rho(\bar{y}^*_{(s-1)T+t}-\bar{y}_{(s-1)T+t-1}))$. By Assumptions 2(a), (b) and (i), the only relevant terms are the covariance between $\bar{\alpha}_{(s-1)T+t-1}+\bar{\varepsilon}_{(s-1)T+t-1}+\rho\bar{y}^*_{(s-1)T+t-1}$ and $-\rho\bar{y}_{(s-1)T+t-1}$ and its asymptotic covariance is given by (18). Finally, for the off-diagonal element of Σ that is not adjacent to the diagonal elements, the asymptotic covariance is zero by Assumption (i). $\square$

Proof of Theorem 3. Because the proofs are analogous, we will prove the theorem for the dynamic case only. Because $\bar{\varepsilon}_{st}=\mathrm{o_p}(1)$, $\hat{\rho}-\rho=\mathrm{o_p}(1)$ and $\hat{\gamma}-\gamma=\mathrm{o_p}(1)$, we have

$$\hat{\beta}=\beta+(X'X)^{-1}X'(\bar{\alpha}+\delta+\bar{\varepsilon})+\mathrm{o_p}(1)=\beta+(X'X)^{-1}X'\delta+\mathrm{o_p}(1), \qquad (42)$$

from which Theorem 3(a) follows. Theorem 3(b) follows immediately from this and

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{ST-K}\sum_{s=1}^{S}\sum_{t=1}^{T}(\bar{\alpha}_{st}+\delta_s-(\hat{\rho}-\rho)\bar{y}_{s,t-1}-(\hat{\beta}-\beta)'x_{st}-(\hat{\gamma}-\gamma)'\bar{z}_{st})^2\\ &= \frac{1}{ST-K}\sum_{s=1}^{S}\sum_{t=1}^{T}(\delta_s-\eta' x_{st})^2+\mathrm{o_p}(1). \quad \square \qquad (43)\end{aligned}$$

References

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.

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76, 598–606.

Andrews, D.W.K., 2003. Cross section regression with common shocks. Cowles Foundation Discussion Paper 1428, Yale University.

Angrist, J.D., 1991. Grouped-data estimation and testing in simple labor-supply models. Journal of Econometrics 47, 243–266.

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.

Baltagi, B.H., 2001. Econometric Analysis of Panel Data, second ed. Wiley, Chichester, UK.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates. Quarterly Journal of Economics, 249–275.

Blundell, R., Meghir, C., Neves, P., 1993. Labour supply and intertemporal substitutions. Journal of Econometrics 59, 137–160.

Browning, M., Deaton, A., Irish, M., 1985. A profitable approach to labor supply and commodity demands over life-cycle. Econometrica 53, 503–544.

Collado, M.D., 1997. Estimating dynamic models from time series of independent cross-sections. Journal of Econometrics 83, 37–62.

Deaton, A., 1985. Panel data from time series of cross-sections. Journal of Econometrics 30, 109–126.

Donald, S.G., Lang, K., 2001. Inference with difference in differences and other panel data. Unpublished manuscript, University of Texas at Austin and Boston University.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Girma, S., 2000. A quasi-differencing approach to dynamic modelling from a time series of independent cross sections. Journal of Econometrics 98, 365–383.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobserved individual effects. Econometrica 49, 1377–1398.

Im, K.S., Ahn, S.C., Schmidt, P., Wooldridge, J.M., 1999. Efficient estimation of panel data models with strictly exogenous explanatory variables. Journal of Econometrics 93, 177–201.

McKenzie, D.J., 2004. Asymptotic theory for heterogeneous dynamic pseudo-panels. Journal of Econometrics 120, 235–262.

Moffitt, R., 1993. Identification and estimation of dynamic models with a time series of repeated cross-sections. Journal of Econometrics 59, 99–123.

Moulton, B.R., 1986. Random group effects and the precision of regression estimates. Journal of Econometrics 32, 385–397.

Nickel, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.

Pepper, J.V., 2002. Robust inferences from random clustered samples: an application using data from the panel study of income dynamics. Economics Letters 75, 341–345.

Staiger, D., Stock, J.H., 1997. Instrumental variables regressions with weak instruments. Econometrica 65, 557–586.

Verbeek, M., 1995. Pseudo-panel data. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: A Handbook of the Theory with Applications. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 280–291.

Verbeek, M., Nijman, T., 1993. Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. Journal of Econometrics 59, 125–136.

Verbeek, M., Vella, F., 2005. Estimating dynamic models from repeated cross-sections. Journal of Econometrics 127, 83–102.

Wooldridge, J.M., 2003. Cluster-sample methods in applied econometrics. American Economic Review 93, 133–138.