Gabriel Montes-Rojas[1] / Walter Sosa-Escudero[2] / Federico Zincenko[3]

Level-Based Estimation of Dynamic Panel Models

[1] Instituto Interdisciplinario de Economía Política, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina, E-mail: gabriel.montes@fce.uba.ar. https://orcid.org/0000-0002-8447-4929.
[2] Universidad de San Andrés-CONICET, Buenos Aires, Argentina
[3] Department of Economics, University of Pittsburgh, Pittsburgh, PA, USA

Abstract:
This paper develops an alternative estimator for linear dynamic panel data models based on parameterizing the covariances between covariates and unobserved time-invariant effects. A GMM framework is used to derive an optimal estimator based on moment conditions in levels, with no efficiency loss compared to the classic alternatives like (Arellano, M., and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–297), (Ahn, S. C., and P. Schmidt. 1995. "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68 (1): 5–27) and (Ahn, S. C., and P. Schmidt. 1997. "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation." *Journal of Econometrics* 76: 309–321). Still, we show analytically and by Monte Carlo simulations that the new procedure leads to efficiency improvements for certain data generating processes. The framework also leads to a very simple test for unobserved effects.

Keywords: asymptotic efficiency, dynamic panel, GMM estimation, individual effects, short panel

JEL classification: C12, C23

DOI: 10.1515/jem-2018-0015

1 Introduction

Dynamic panel models with a fixed and usually small time dimension ("short panels") occupy a substantial body of the applied and theoretical literature. The applied literature has largely favored instrumental variables (IV) and moment conditions estimation strategies to eliminate the so-called Nickel bias (Nickel 1981), which date back to the pioneering work by Anderson and Hsiao (1981, 1982). A successful line of research has relied on the Generalized Method of Moments (GMM) framework to exploit efficiently all the available moment conditions that arise from the dynamic structure of the model. Ahn and Schmidt (1995), Arellano and Bond (1991), and Holtz-Eakin, Newey, and Rosen (1988) (1995, 1997), Arellano and Bover (1995) and Blundell and Bond (1998) are examples of research along this line whose results are widely applied in empirical work. See also Hausman and Pinkovskiy (2017) for a recent development in the literature. Moment based strategies are shown to perform relatively well in practice, are easy to handle and communicate within the relevant IV paradigm in econometrics, and are also computationally convenient, as stressed by Harris, Matyas, and Sevestre (2008) in their survey. Admittedly, IV-GMM methods are not free from limitations, including the problem of weak instruments (Bun and Windmeijer 2010; Stock, Yogo, and Wright 2002) and the related issue of moment multiplicity (Roodman 2009).

In this paper we propose a novel IV-GMM estimator that exploits new moment conditions implied by the commonly used dynamic panel models with unobserved time invariant effects. In the spirit of Chamberlain's (1982, 1984) classic work and following Robertson and Sarafidis (2015), we introduce new parameters to capture the covariances between regressors and time invariant effects, and treat the relationship among such parameters as moment conditions, which are later exploited jointly with those arising from initial conditions and the dynamic structure of the model in a GMM fashion that leads to an optimal estimator. The asymptotic framework corresponds to the large N, finite and small T case. Among other results, our paper shows that estimators like Arellano and Bond (1991) or Ahn and Schmidt (1995, 1997) are particular cases of our framework and, consequently, weakly dominated in efficiency by our strategy. We compare the asymptotic variances and establish under which conditions our estimator is asymptotically more efficient. A Monte Carlo experiment suggests that our proposed GMM estimator performs better (in terms of bias and RMSE) than Ahn and Schmidt (1995)

Gabriel Montes-Rojas is the corresponding author.

and Arellano and Bond (1991) for the case of non-stationary covariates, and that it leads to a simple variance estimator that can be used to test for the presence of unobserved effects, improving upon the performance of alternatives like Wu and Zhu (2012).

The approaches in the line of Chamberlain (1984) are theoretically relevant and simple to communicate and implement in alternative contexts that include incidental parameters. In the case of dynamic panel models, our strategy leads to a model in levels (as opposed to differences) that is simple to implement in practice. This is the key difference with the existing IV-GMM methods because they are based on differences, while the proposed method is IV-GMM based on levels. The methods based on differences eliminate the individual effect before forming the moment conditions, and thus the moment conditions only involve the original model parameters. Our method is based on levels and thus it does not eliminate the individual effects, but it requires the covariance of the individual effects and the IV's in the moment conditions. Those covariances do not increase in number as the sample size (number of individuals) increases, and thus can be treated as parameters. Based on such insights, we prove that the moment conditions in levels can identify all the structural parameters, as well as the implied covariances.

Our analytical framework allows for the construction of an estimator with a closed-form expression of the Jacobian matrix and that avoids the problem of multiple local minima. Note that this alternative implementation does not require restrictions on the initial conditions of the dependent variable process (e.g. as in Blundell and Bond 1998), but we only need to model the covariance of the initial values of the dependent variable and the unobserved individual effects as an additional parameter.

A convenient by-product of our strategy is that it produces a very simple procedure to estimate the variance of the unobserved error term, which is crucial for empirical work where such parameter measures the relative importance of individual-specific time-invariant factors in explaining persistence, as in the classic article by Lillard and Willis (1978). Simple tests are immediate to derive within the proposed framework. For example, our estimation process leads to a very simple alternative to the test by Wu and Zhu (2012) for the presence of unobserved heterogeneity.

The paper is structured as follows. Section 2 presents the model. Section 3 introduces the parametrization together with the corresponding moment conditions in levels and establishes the key identification condition that leads to the proposed estimator, which is derived in Section 4. Section 5 compares the proposed estimator with existing GMM strategies. Section 6 explores the performance of the proposed strategy and other alternatives in finite samples through a Monte Carlo exercise. Section 7 concludes.

2 Model

Consider a dynamic panel data model of the form

$$y_{it} = \alpha_\circ + \gamma_\circ y_{i,t-1} + x'_{it}\beta_\circ + \mu_i + \varepsilon_{it}, \tag{1}$$

where $i = 1, \ldots, N$ and $t = 2, \ldots, T$. In this model, y_{it} is the dependent variable, x_{it} is a $k \times 1$ vector of regressors, μ_i is the individual effect, and ε_{it} is an error term, $\alpha_\circ$ is an intercept, $\gamma_\circ$ is a scalar less than 1 in absolute value, and $\beta_\circ$ is a $k \times 1$ vector of coefficients. The variance of μ_i is denoted by $\sigma^2_{\mu\circ}$. We will assume that $T \geq 3$.

The coefficient $\gamma_\circ$ measures the degree of the state dependence or pure dynamic persistence. The factor μ_i induces an alternative source of persistence, usually referred to as the individual effect or unobserved heterogeneity. The presence of both types of persistence is a key factor in the dynamic panel data literature and, as is well known, standard estimators (OLS, LS dummy variables) are not consistent for $\gamma_\circ$ when $N \to \infty$ and T is fixed (Nickel 1981).

Assume that the researcher observes a random sample $\{[(y_{i1}, y'_i)', (x_{i1}, x_i)'] : i = 1, \ldots, N\}$, where $y_i = (y_{i2}, \ldots, y_{iT})'$ is a $(T-1) \times 1$ random vector and $x_i = (x_{i2}, \ldots, x_{iT})$ is a $k \times (T-1)$ random matrix containing the regressors. The observed initial value of the dependent variable is y_{i1}. The asymptotic properties of our estimator will be derived assuming that N grows to infinity and T is fixed, i.e. short panels, which is standard in microeconometrics.

The following assumptions are imposed. Let $\varepsilon_i = (\varepsilon_{i2}, \ldots, \varepsilon_{iT})'$ be a $(T-1) \times 1$ random vector containing the error terms.

Assumption 1
$\{(y_{i1}, x'_{i1}, \ldots, x'_{iT}, \mu_i, \varepsilon'_i) : i = 1, \ldots, N\}$ *are independent and identically distributed (i.i.d.) random vectors with finite fourth moments.*

This assumption imposes independence across individuals, but different intra-individual structures are allowed. Combined with eq. (1), Assumption 1 implies that $\{(y_{i1}, y_i', x_{i1}', \ldots, x_{iT}') : i = 1, \ldots, N\}$ are i.i.d. random vectors with finite fourth moments.

Assumption 2

The following conditions hold.

1. $\{\mu_i : i = 1, \ldots, N\}$ *have zero mean and variance* $\sigma^2_{\mu\circ} \geq 0$.
2. *For each* i, $\{\varepsilon_{it} : t = 2, \ldots, T\}$ *have zero mean and are uncorrelated, with,* $\mathrm{var}(\varepsilon_{it}) > 0$ *for all* t.
3. **a.** $E(y_{i1}\varepsilon_{it}) = 0$ *for all* $t = 2, \ldots, T$.
   b. $E(x_{is}\varepsilon_{it}) = 0$ *for all* $t = 2, \ldots, T$ *and* $1 \leq s \leq t$.
   c. $E(\mu_i\varepsilon_{it}) = 0$ *for all* $t = 2, \ldots, T$.

The first condition imposes the usual normalizing restriction $E(\mu_i) = 0$, since the model contains a common intercept. The second requires that the errors ε_{it} are uncorrelated across $t = 1, \ldots, T$. This condition can be relaxed to incorporate serial correlation, at the price of modifying the moment conditions in the next section. The zero-mean condition on ε_{it} can also be relaxed as long as $E(\varepsilon_{it})$ is constant across periods; in such case, $E(\varepsilon_{it})$ will be "captured" by $\alpha_\circ$. The variance of ε_{it} is allowed to vary across t. The third part of Assumption 2 implies a set of sequential moment conditions that will be exploited to obtain valid instruments. Assumption 2 together with eq. (1) implies

$$E(y_{is}\varepsilon_{it}) = 0,$$

for all $t = 2, \ldots, T$ and $1 \leq s < t$. Observe that the individual effects μ_i are allowed to be correlated with $(x_{i1}, x_i)'$. No restrictions are imposed on the relationship between y_{i1} and μ_i.

Assumption 1 is equivalent to Ahn and Schmidt's (1995) (SA.1–SA.3) in a model without the covariate x_{it}. In such case Ahn and Schmidt (1995) assume Assumptions M.1–M.5, which are stronger then our Assumption 2.3(b). We remark that zero-mean condition on ε_{it} and Assumption 2.3(c) are not needed in Arellano and Bond's (1991) setting. Specifically, $E(\varepsilon_{it})$ is allowed to vary across t in such a setting.

3 Moment Conditions in Levels

This section presents the moments conditions derived from Assumption 1–Assumption 2 and then establishes an identification result. We start by introducing some notation, similar to Yamagata (2008). Let I_{T-1} denote the identity matrix of dimension $(T-1) \times (T-1)$. Define $\mathbf{Z}_i = [I_{T-1}, \mathbf{Z}_{Yi}, \mathbf{Z}_{Xi}]_{(T-1)\times h}$ with $h = (T-1) + h_y + h_x$, $h_y = T(T-1)/2$, $h_x = k(T+2)(T-1)/2$, where

$$\underset{(T-1)\times h_y}{\mathbf{Z}_{Yi}} = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & y_{i1} & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-1} \end{pmatrix},$$

and

$$\underset{(T-1)\times h_x}{\mathbf{Z}_{Xi}} = \begin{pmatrix} x_{i1}' & x_{i2}' & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & \cdots & \cdots & \cdots & 0_{1\times k} \\ 0_{1\times k} & 0_{1\times k} & x_{i1}' & x_{i2}' & x_{i3}' & 0_{1\times k} & \cdots & \cdots & \cdots & 0_{1\times k} \\ 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & x_{i1}' & \cdots & \cdots & \cdots & 0_{1\times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \vdots \\ 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & \cdots & x_{i1}' & \cdots & x_{iT}' \end{pmatrix}.$$

The nonzero elements of $\mathbf{Z}_{Yi}$ and $\mathbf{Z}_{Xi}$ will play the role of instruments as in Arellano and Bond (1991). This framework could also incorporate strictly exogenous covariates, thus increasing the number of moment conditions that can be used. For simplicity, we focus only on pre-determined covariates only.

By Assumption 2, $\mathbf{Z}_i$ and ε_i satisfy the moment conditions

$$E\left(\mathbf{Z}_i'\varepsilon_i\right) = E\begin{pmatrix} I_{T-1}\varepsilon_i \\ \mathbf{Z}_{Yi}'\varepsilon_i \\ \mathbf{Z}_{Xi}'\varepsilon_i \end{pmatrix} = 0_{h\times 1}.$$

Write $u_{it} = \mu_i + \varepsilon_{it}$ and $u_i = (u_{i2}, \dots, u_{iT})'$. Then

$$E\left(\mathbf{Z}_i'\varepsilon_i\right) = E\left(\mathbf{Z}_i'u_i\right) - E\left(\mathbf{Z}_i'\mu_i\right)\iota_{T-1} = 0_{h\times 1}, \tag{2}$$

where ι_{T-1} stands for a $(T-1)\times 1$ vector of ones.

The matrix $E\left(\mathbf{Z}_i'\mu_i\right)$ contains the covariances between μ_i and the elements of $\mathbf{Z}_i$, namely, $E(y_{i1}\mu_i), E(y_{i2}\mu_i), \dots, E(y_{iT}\mu_i)$, $E(x_{i1}\mu_i), \dots, E(x_{iT}\mu_i)$. The fact that these covariances involve the levels of the variables is an important feature of the paper. First, Assumption 1 allows for the covariance between x_{it} and μ_i to vary across t, but not across i. This parametrization can be changed to accommodate the same covariance parameter across t or for some covariates to be uncorrelated with μ_i, thus generating a mixed model, i.e. random- and fixed-effects structure together. In both cases, this would imply reducing the number of parameters involved in the estimation procedure. Second, the standard dynamic panel data literature applies first-differences to eliminate μ_i, which is correlated with $(y_{i,t-1}, x_{it}')$ invalidating the previously proposed instruments. In our case, we consider these correlations as free parameters following the method of Robertson and Sarafidis (2015). In turn this is related to the correlated random-effects model of Chamberlain (1982, 1984) where unobservable individual specific components are modeled as linear projections onto the observables plus a disturbance. The intuition behind such strategy is that covariates themselves are able to explain unobserved heterogeneity and what is left is idiosyncratic noise. Third, note that this alternative implementation does not require restrictions on the initial conditions of the dependent variable process (e.g. as in Blundell and Bond 1998), but we only need to model the covariance of the initial values of the dependent variable and the unobserved individual effects as an additional parameter, i.e., $E(y_{i1}\mu_i)$.

Let $\tau_{1\circ}^y = E(y_{i1}\mu_i)$ and $\tau_{t\circ}^x = E(x_{it}\mu_i)$ for $t = 1, \dots, T$. We consider these covariances as parameters of our model. Let $\tau_\circ = (\tau_{1\circ}^y, \tau_{1\circ}^{x\prime}, \dots, \tau_{T\circ}^{x\prime})'$ be a $(kT+1)\times 1$ vector containing these covariances and $\theta_\circ = (\alpha_\circ, \gamma_\circ, \beta_\circ', \sigma_{\mu\circ}^2, \tau_\circ')$ be the true parameters vector, with dimension $1\times[k(T+1)+4]$. Observe that the covariances $E(y_{i2}\mu_i), \dots, E(y_{iT}\mu_i)$ can be completely characterized in terms of $(\gamma_\circ, \beta_\circ', \sigma_{\mu\circ}^2, \tau_\circ')$. By eq. (1) and Assumption 2:

$$E(y_{i2}\mu_i) = \gamma_\circ\tau_{1\circ}^y + \tau_{2\circ}^{x\prime}\beta_\circ + \sigma_{\mu\circ}^2. \tag{3}$$

It can be shown by induction that

$$E\left(y_{it}\mu_i\right) = \gamma_\circ^{t-1}\tau_{1\circ}^y + \sum_{j=2}^{t}\gamma_\circ^{t-j}\tau_{j\circ}^{x\prime}\beta_\circ + \frac{\gamma_\circ^{t-1}-1}{\gamma_\circ-1}\sigma_{\mu\circ}^2, \tag{4}$$

for $2\le t\le T$. Note that $\sigma_{\mu\circ}^2 = 0$ implies $\tau_\circ = 0_{(kT+1)\times 1}$.

In light of the previous discussion, we now parameterize the covariances $E\left(\mathbf{Z}_i'\mu_i\right)\iota_{T-1}$ and $\sigma_{\mu\circ}^2$. Consider the row vector $\theta = (\alpha, \gamma, \beta', \sigma_\mu^2, \tau') \in \mathbb{R}\times(-1,1)\times\mathbb{R}^k\times\mathbb{R}\times\mathbb{R}^{kT+1}$ and let Θ be a compact subset such that $\theta_\circ \in \text{interior}(\Theta)$. Although $\sigma_{\mu\circ}^2 \ge 0$, σ_μ^2 may be allowed to take negative values as the identification result (Lemma 1 below) does not rely on the restriction $\sigma_\mu^2 \ge 0$. In view of expressions (2)–(4), we construct the function $\Psi : \Theta \to \mathbb{R}^{h\times 1}$ as

$$\Psi(\theta) = \begin{pmatrix} 0_{(T-1)\times 1} \\ \Psi_Y(\theta) \\ \Psi_X(\theta) \end{pmatrix}, \tag{5}$$

where

$$\underset{h_y\times 1}{\Psi_Y(\theta)} = (\tau_1^y, \tau_1^y, \psi_2(\theta), \tau_1^y, \psi_2(\theta), \psi_3(\theta), \tau_1^y, \dots, \tau_1^y, \dots, \psi_{T-1}(\theta))',$$

$$\underset{h_x\times 1}{\Psi_X(\theta)} = (\tau_1^{x\prime}, \tau_2^{x\prime}, \tau_1^{x\prime}, \tau_2^{x\prime}, \tau_3^{x\prime}, \tau_1^{x\prime}, \dots, \tau_1^{x\prime}, \dots, \tau_T^{x\prime})',$$

and

$$\psi_t(\theta) = \gamma^{t-1}\tau_1^y + \sum_{l=2}^{t} \gamma^{t-l}\tau_l^{x\prime}\beta + \frac{\gamma^{t-1}-1}{\gamma-1}\sigma_\mu^2, \tag{6}$$

for $2 \le t \le T-1$ and $|\gamma| < 1$. Note that $\Psi(\theta)$ depends only on $(\gamma, \beta', \sigma_\mu^2, \tau')$ and not on the data.

Alternatively, $\Psi_Y(\theta)$ and $\Psi_X(\theta)$ can be constructed as follows. The parameter τ_1^y occupies the positions $\{[t(t-1)/2]+1 : t = 1, \dots, T-1\}$ of $\Psi_Y(\theta)$, $\psi_2(\theta)$ occupies positions $\{[t(t-1)/2]+2 : t = 2, \dots, T-1\}$ of $\Psi_Y(\theta)$, and in general for $2 \le j \le T-1$, $\psi_j(\theta)$ occupies positions $\{[t(t-1)/2]+j : t = j, \dots, T-1\}$ of $\Psi_Y(\theta)$. Regarding $\Psi_X(\theta)$, let $\Psi_X^{(j_1:j_2)}(\theta)$ denote the sub-vector of $\Psi_X(\theta)$ from position j_1 to j_2. For each $t = 1, \dots, T$ and $l = \max\{t-1, 1\}, \dots, T-1$, we set $j_1 = k[t-2+l(l+1)/2]+1$, $j_2 = k[t-1+l(l+1)/2]$, and $\Psi_X^{(j_1:j_2)}(\theta) = \tau_t^x$.

Observe that $\psi_t(\theta_\circ) = E(y_{it}\mu_i)$ for $t \ge 2$ due to eq. (4). The matrices $E\left(\mathbf{Z}'_{Yi}\mu_i\right)$ and $E\left(\mathbf{Z}'_{Xi}\mu_i\right)$ can then be written as

$$E(\mathbf{Z}'_{Yi}\mu_i) = \begin{pmatrix} \tau_{1\circ}^y & 0 & 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \tau_{1\circ}^y & \psi_2(\theta_\circ) & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \tau_{1\circ}^y & \dots & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \tau_{1\circ}^y & \psi_2(\theta_\circ) & \dots & \psi_{T-1}(\theta_\circ) \end{pmatrix}$$

and

$$E(\mathbf{Z}'_{Xi}\mu_i) = \begin{pmatrix} \tau_{1\circ}^{x\prime} & \tau_{2\circ}^{x\prime} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & \dots & \dots & \dots & 0_{1\times k} \\ 0_{1\times k} & 0_{1\times k} & \tau_{1\circ}^{x\prime} & \tau_{2\circ}^{x\prime} & \tau_{3\circ}^{x\prime} & 0_{1\times k} & \dots & \dots & \dots & 0_{1\times k} \\ 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & \tau_{1\circ}^{x\prime} & \dots & \dots & \dots & 0_{1\times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & 0_{1\times k} & \dots & \tau_{1\circ}^{x\prime} & \dots & \tau_{T\circ}^{x\prime} \end{pmatrix},$$

respectively, which implies

$$E\left(\mathbf{Z}'_i u_i\right) - \Psi(\theta_\circ) = 0_{h\times 1}. \tag{7}$$

Define the function $g_i : \Theta \to \mathbb{R}^{h\times 1}$ as

$$\underset{h\times 1}{g_i(\theta)} = \mathbf{Z}'_i(y_i - \mathbf{x}_i\kappa) - \Psi(\theta), \tag{8}$$

where $\theta = (\kappa', \sigma_\varepsilon^2, \sigma_\mu^2, \tau')$, $\kappa = (\alpha, \gamma, \beta')'$, and $\mathbf{x}_i = (\iota_{T-1}, y_{i,-1}, x_i')$. This function satisfies $E[g_i(\theta_\circ)] = 0_{h\times 1}$ since $u_i = y_i - \alpha_\circ\iota_{T-1} - \gamma_\circ y_{i,-1} - x_i'\beta_\circ$ and eq. (7). After imposing a standard rank condition, we will show that $\theta_\circ$ is the unique solution to the (nonlinear) system of equations $E[g_i(\theta)] = 0_{h\times 1}$ with $\theta \in \Theta$.

Write $\tilde{\mathbf{Z}}_i = (\tilde{\mathbf{Z}}_{Yi}, \tilde{\mathbf{Z}}_{Xi})_{(T-2)\times\tilde{h}}$, where $\tilde{\mathbf{Z}}_{Yi}$ is constructed by removing the last of row and the last $T-1$ columns of $\mathbf{Z}_{Yi}$, $\tilde{\mathbf{Z}}_{Xi}$ is constructed in a similar manner by removing the last of row and the last kT columns of $\mathbf{Z}_{Xi}$, and $\tilde{h} = [T-1+k(T+1)](T-2)/2$. Denote further $\Delta y_{i,-1} = (y_{i2}-y_{i1}, \dots, y_{i,T-1}-y_{i,T-2})'$, $\Delta\tilde{x}_i = (x_{i3}-x_{i2}, \dots, x_{iT}-x_{i,T-1})'$, and $\Delta\mathbf{x}_i = (\Delta y_{i,-1}, \Delta\tilde{x}_i)_{(T-2)\times(k+1)}$. Now we are ready to impose the following rank condition.

Assumption 3
$E\left(\tilde{\mathbf{Z}}'_i\Delta\mathbf{x}_i\right)$ has rank $k+1$.

This assumption is standard in the dynamic panel literature, e.g. it coincides with Assumption 3 in Yamagata (2008), and rules out perfect multicollinearity among regressors. It is employed for identification purposes and is verifiable for any given sample.

To complete this section, the next identification lemma characterizes the true parameters vector $\theta_\circ$ as the unique solution of a system of moment conditions based on $g_i(\cdot)$. A detailed proof of this lemma is provided in Appendix A.3.

Lemma 1
Under Assumption 1–Assumption 3, $\theta_\circ \in \Theta$ is the unique solution of $E[g_i(\theta)] = 0_{h\times 1}$.

4 Optimal GMM Estimator

This section proposes a GMM estimator for $\theta_\circ$ and derives its asymptotic properties.

Consider $\Omega \equiv E[g_i(\theta_\circ)g_i(\theta_\circ)']$. Since $g_i(\theta_\circ) = \mathbf{Z}_i'u_i - E(\mathbf{Z}_i'u_i)$ by eqs. (7)–(8), Ω is the variance-covariance matrix of $\mathbf{Z}_i'u_i$: $\Omega = E\{[\mathbf{Z}_i'u_i - E(\mathbf{Z}_i'u_i)][\mathbf{Z}_i'u_i - E(\mathbf{Z}_i'u_i)]'\}$. Assumption 1 (finite fourth moments) implies that Ω is finite, symmetric, and positive semidefinite. Further, we impose the following assumption. Let $\xrightarrow{P}$ denote convergence in probability.

Assumption 4

Ω is positive definite and there exists an estimator $\dot{\Omega}$ that satisfies $\dot{\Omega} \xrightarrow{P} \Omega$ under Assumption 1–Assumption 3.

Assumption 4 is standard is the context of linear IV models (particularly, in dynamic panels). Essentially, the product between any instrument and any residual must be nonconstant (have positive variance) and perfect collinearity among such products is not allowed. Regarding the existence of a consistent estimator of Ω, Appendix A.1 provides a concrete suggestion for $\dot{\Omega}$, as well as a first-step consistent estimator of $\theta_\circ$.

The optimal GMM estimator of $\theta_\circ$ is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \bar{g}(\theta)'\dot{\Omega}^{-1}\bar{g}(\theta),$$

where $\bar{g}(\theta) = (1/N)\sum_{i=1}^{N} g_i(\theta)$ and $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\beta}', \hat{\sigma}_\mu^2, \hat{\tau}')$. Since $\bar{g}(\cdot)$ is continuously differentiable, $\hat{\theta}$ can be characterized as the unique solution of the system of (nonlinear) equations

$$ {}_{\theta}\bar{g}(\theta)'\dot{\Omega}^{-1}\bar{g}(\theta) = 0_{[k(T+1)+4]\times 1}, \tag{9}$$

where ${}_{\theta}\bar{g}(\theta) = (1/N)\sum_{i=1}^{N} {}_{\theta}g_i(\theta)$ and

$$\underset{h\times[k(T+1)+4]}{{}_{\theta}g_i(\theta)} = \frac{\partial g_i(\theta)}{\partial \theta} = \left(\frac{\partial g_i(\theta)}{\partial \alpha} \quad \frac{\partial g_i(\theta)}{\partial \gamma} \quad \frac{\partial g_i(\theta)}{\partial \beta'} \quad \frac{\partial g_i(\theta)}{\partial \sigma_\mu^2} \quad \frac{\partial g_i(\theta)}{\partial \tau'} \right).$$

Two remarks are noteworthy. First, computational differentiation is not required to calculate ${}_{\theta}\bar{g}(\theta)$. A closed-form expression of the Jacobian matrix of $\Psi(\theta)$ is provided in Appendix A.2. Second, the system of equations (9) can be solved by standard numerical methods, e.g. by using the Newton-Raphson algorithm. We suggest using a first-step consistent estimator of θ_0 as initial value (Appendix A.1 provides an example of such an estimator). Following this suggestion, the initial value for solving (9) would be close to the global minimum and therefore multiple local minima (or equivalently, multiple solutions to eq. (9)) would not be a concern.

The next assumption is imposed to derive the asymptotic normality of $\hat{\theta}$.

Assumption 5

$G'\Omega^{-1}G$ is nonsingular where $G = E[{}_{\theta}g_i(\theta_0)]$.

The following theorem contains the asymptotic results. Let $\xrightarrow{D}$ denote convergence in distribution.

Theorem 1

(Consistency and asymptotic normality). *The following conditions hold.*

1. *Under Assumptions* Assumption 1–Assumption 4, *we have* $\hat{\theta} \xrightarrow{P} \theta_\circ$.

2. *Under Assumptions* Assumption 1–Assumption 5, *the asymptotic distribution of our estimator is given by* $\sqrt{N}(\hat{\theta} - \theta_\circ) \xrightarrow{D} N(0, \Sigma)$, *where* $\Sigma = B^{-1}$ *with* $B = G'\Omega^{-1}G$.

Under Assumption 1–Assumption 5, we have $\hat{\theta} \xrightarrow{P} \theta_\circ$ and the asymptotic distribution of our estimator is given by $\sqrt{N}(\hat{\theta} - \theta_\circ) \xrightarrow{D} N(0, \Sigma)$, where $\Sigma = B^{-1}$ with $B = G'\Omega^{-1}G$.

The asymptotic variance Σ can be consistently estimated by $\hat{\Sigma} = \hat{B}^{-1}$, where $\hat{B} = \hat{G}'\hat{\Omega}^{-1}\hat{G}$, $\hat{G} = \bar{g}(\hat{\theta})$, and $\hat{\Omega} = (1/N)\sum_{i=1}^{N} g_i(\hat{\theta})g_i(\hat{\theta})'$.

Lemma 2

Under Assumption 1–Assumption 5, $\hat{\Sigma} \xrightarrow{P} \Sigma$.

Among other hypotheses of interests, the results in Theorem 1 and Lemma 2 can be used to derive a simple test for the absence of individuals effects, which is an economically relevant issue in the literature that distinguished true dependence vs. that derived from unobserved heterogeneity, as in the classic paper by Lillard and Willis (1978) mentioned in the Introduction. Under the null $H_0 : (\sigma^2_{\mu\circ}, \tau'_\circ) = 0_{1\times(kT+2)}$, we have

$$(\hat{\sigma}^2_\mu, \hat{\tau}')\hat{\Sigma}^{-1}_{\sigma^2_\mu \tau}(\hat{\sigma}^2_\mu, \hat{\tau}')' \xrightarrow{D} \chi^2_{kT+2},$$

where $\hat{\Sigma}_{\sigma^2_\mu \tau}$ is the sub-matrix of $\hat{\Sigma}$ associated with $(\hat{\sigma}^2_\mu, \hat{\tau}')$ and χ^2_{kT+2} denotes a (central) chi-square distribution with $kT + 2$ degrees of freedom. This results compares to tests derived by Harris, Matyas, and Sevestre (2008, sec. 8.6.2) and Wu and Zhu (2012) in a different context.

5 Comparison with Existing GMM Estimators and Efficiency Gains

This section relates our estimator to the ones proposed by Arellano and Bond (1991) and by Ahn and Schmidt (1995). We compare the asymptotic variances and show that our estimator is at least as efficient as these classic alternatives. We also establish under which conditions our estimator is more efficient or, in other words, when our asymptotic variance is strictly smaller than the one from its competitors.

The optimal Arellano-Bond GMM estimator can be obtained by

$$(\hat{\gamma}^{AB}, \hat{\beta}^{AB\prime}) = \underset{(\gamma,\beta')}{\text{argmin}}\, \bar{g}^{AB}(\gamma, \beta')'(\dot{\Omega}^{AB})^{-1}\bar{g}^{AB}(\gamma, \beta'), \tag{10}$$

where $\bar{g}^{AB}(\gamma, \beta') = (1/N)\sum_{i=1}^N g_i^{AB}(\gamma, \beta')$, $g_i^{AB}(\gamma, \beta') = \mathrm{D}^{AB} g_i(\theta)$, D^{AB} is a $\tilde{h} \times h$ nonstochastic matrix that is characterized in Appendix A.3.1, and $\dot{\Omega}^{AB}$ is a consistent estimator of

$$\Omega^{AB} \equiv E\left[g_i^{AB}(\gamma_\circ, \beta'_\circ) g_i^{AB}(\gamma_\circ, \beta'_\circ)'\right] = \mathrm{D}^{AB}\Omega\mathrm{D}^{AB\prime}.$$

Specifically, we have that

$$\underset{\tilde{h}\times 1}{g_i^{AB}(\gamma, \beta')} = \tilde{\mathbf{Z}}_i' \left(\Delta y_i - \gamma\Delta y_{i,-1} - \Delta\tilde{x}_i\beta\right),$$

so the linear transformation D^{AB} removes not only the individual effects, but also the function $\Psi(\cdot)$ from $g_i(\cdot)$. We highlight that the specific form of Ω^{AB} depends on the assumptions about ε_i –such as homoskedasticity–.

Example 1

$(T = 4\ \&\ k = 0)$.

We have

$$g_i^{AB}(\gamma) = \begin{pmatrix} y_{i1}(\Delta y_{i3} - \gamma\Delta y_{i2}) \\ y_{i1}(\Delta y_{i4} - \gamma\Delta y_{i3}) \\ y_{i2}(\Delta y_{i4} - \gamma\Delta y_{i3}) \end{pmatrix}$$

and

$$\underset{3\times 9}{\mathrm{D}^{AB}} = \begin{pmatrix} 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}.$$

The function g_i^{AB} is related to the moment conditions

$$E\begin{pmatrix} y_{i1}\Delta u_{i3} \\ y_{i1}\Delta u_{i4} \\ y_{i2}\Delta u_{i4} \end{pmatrix} = 0_{3\times 1}.$$

The asymptotic distribution of $(\hat{\gamma}^{AB}, \hat{\beta}^{AB\prime})$ is given by

$$\sqrt{N}\left[\begin{pmatrix}\hat{\gamma}^{AB}\\ \hat{\beta}^{AB\prime}\end{pmatrix}-\begin{pmatrix}\gamma_\circ\\ \beta'_\circ\end{pmatrix}\right]\xrightarrow{D} N\left(0,\Sigma^{AB}_{\gamma\beta}\right),$$

where $\Sigma^{AB}_{\gamma\beta}=[G^{AB\prime}(\Omega^{AB})^{-1}G^{AB}]^{-1}$,

$$G^{AB}=E[\ _{(\gamma,\beta')}g_i^{AB}(\gamma_\circ,\beta'_\circ)]=\mathrm{D}^{AB}E[\ _{(\gamma,\beta')}g_i(\theta_\circ)],$$

and

$$\underset{\tilde{h}\times(1+k)}{_{(\gamma,\beta')}g_i^{AB}(\gamma,\beta')}=\begin{pmatrix}\frac{\partial g_i^{AB}(\gamma,\beta')}{\partial\gamma} & \frac{\partial g_i^{AB}(\gamma,\beta')}{\partial\beta'}\end{pmatrix}.$$

Ahn and Schmidt (1995, 1997) have incorporated to the Arellano-Bond estimator the quadratic moment conditions that exploit the absence of serial correlation in ε_{it} (Assumption 2.2). The optimal Ahn-Schmidt GMM estimator can be obtained by

$$(\hat{\gamma}^{AS},\hat{\beta}^{AS\prime})=\underset{(\gamma,\beta')}{\operatorname{argmin}}\,\bar{g}^{AS}(\gamma,\beta')'(\dot{\Omega}^{AS})^{-1}\bar{g}^{AS}(\gamma,\beta'),\tag{11}$$

where $\bar{g}^{AS}(\gamma,\beta')=(1/N)\sum_{i=1}^N g_i^{AS}(\gamma,\beta')$, $g_i^{AS}(\gamma,\beta')=\mathrm{D}^{AS}(\gamma,\beta')g_i(\theta)$, $\mathrm{D}^{AS}(\gamma,\beta')$ is a $\tilde{h}'\times h$ nonstochastic matrix described in eq. (A.7) of Ahn and Schmidt (1995), $\tilde{h}'=\tilde{h}+T-3$, and $\dot{\Omega}^{AS}$ is a consistent estimator of

$$\Omega^{AS}\equiv E\left[g_i^{AS}(\gamma_\circ,\beta'_\circ)g_i^{AS}(\gamma_\circ,\beta'_\circ)'\right]=\mathrm{D}^{AS}(\gamma_\circ,\beta'_\circ)\Omega\mathrm{D}^{AS\prime}(\gamma_\circ,\beta'_\circ).$$

Specifically, we have that

$$\underset{\tilde{h}\times 1}{g_i^{AS}(\gamma,\beta')}=\begin{pmatrix} g_i^{AB}(\gamma)\\ (y_{iT}-\gamma y_{i,T-1}-x'_{it}\beta)\left(\Delta y_{i3}-\gamma\Delta y_{i,2}-\Delta x'_{i3}\beta\right)\\ \vdots\\ (y_{iT}-\gamma y_{i,T-1}-x'_{it}\beta)\left(\Delta y_{i,T-1}-\gamma\Delta y_{i,T-2}-\Delta x'_{i,T-1}\beta\right)\end{pmatrix};$$

when $T=3$, Ahn-Schmidt estimator coincides with Arellano-Bond's.

Example 2
($T=4$ & $k=0$, cont.). We have

$$g_i^{AS}(\gamma)=\begin{pmatrix} y_{i1}(\Delta y_{i3}-\gamma\Delta y_{i2})\\ y_{i1}(\Delta y_{i4}-\gamma\Delta y_{i3})\\ y_{i2}(\Delta y_{i4}-\gamma\Delta y_{i3})\\ (y_{i4}-\gamma y_{i3})(\Delta y_{i3}-\gamma\Delta y_{i2})\end{pmatrix}=\begin{pmatrix} g_i^{AB}(\gamma)\\ (y_{i4}-\gamma y_{i3})(\Delta y_{i3}-\gamma\Delta y_{i2})\end{pmatrix}$$

and

$$\underset{4\times 9}{\mathrm{D}^{AS}(\gamma)}=\begin{pmatrix}0&0&0&-1&1&0&0&0&0\\0&0&0&0&-1&0&1&0&0\\0&0&0&0&0&-1&0&1&0\\0&0&0&0&0&0&\gamma&-(\gamma+1)&1\end{pmatrix}.$$

The function g_i^{AS} is related to the moment conditions

$$E\begin{pmatrix}y_{i1}\Delta u_{i3}\\ y_{i1}\Delta u_{i4}\\ y_{i2}\Delta u_{i4}\\ u_{i4}\Delta u_{i3}\end{pmatrix}=0_{4\times 1}.$$

The asymptotic distribution of $(\hat{\gamma}^{AS}, \hat{\beta}^{AS\prime})$ is given by

$$\sqrt{N}\left[\begin{pmatrix}\hat{\gamma}^{AS}\\ \hat{\beta}^{AS\prime}\end{pmatrix} - \begin{pmatrix}\gamma_\circ\\ \beta'_\circ\end{pmatrix}\right] \xrightarrow{D} N\left(0, \Sigma^{AS}_{\gamma\beta}\right),$$

where $\Sigma^{AS}_{\gamma\beta} = [G^{AS\prime}(\Omega^{AS})^{-1}G^{AS}]^{-1}$, $G^{AS} = E[\ {}_{(\gamma,\beta')}g_i^{AS}(\gamma_\circ, \beta'_\circ)]$, and

$$\underset{\tilde{h}'\times(1+k)}{{}_{(\gamma,\beta')}g_i^{AS}(\gamma,\beta')} = \begin{pmatrix}\dfrac{\partial g_i^{AS}(\gamma,\beta')}{\partial\gamma} & \dfrac{\partial g_i^{AS}(\gamma,\beta')}{\partial\beta'}\end{pmatrix}.$$

To compare our estimator $(\hat{\gamma}, \hat{\beta}')$ with the previous ones, we consider the function

$$\underset{h\times 1}{g_i^D(\theta)} = \begin{pmatrix} g_{i,1}^D(\gamma,\beta') \\ g_{i,2}^D(\gamma,\beta';\theta_{\setminus\gamma\beta})\end{pmatrix},$$

where

$$\underset{(\tilde{h}'+T-2)\times 1}{g_{i,1}^D(\gamma,\beta')} = \begin{pmatrix} g_i^{AS}(\gamma,\beta') \\ \Delta y_i - \gamma\Delta y_{i,-1} - \Delta\tilde{x}_i\beta\end{pmatrix},$$

$\theta_{\setminus\gamma\beta} = (\alpha, \sigma^2_\mu, \tau')$, and

$$\underset{(kT+3)\times 1}{g_{i,2}^D(\gamma,\beta';\theta_{\setminus\gamma\beta})} = \begin{pmatrix} y_{iT} - \gamma y_{i,T-1} - x'_{iT}\beta \\ y_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ \vdots \\ x_{iT}(y_{iT} - \alpha - \gamma y_{i,T-1} - \beta x_{iT}) \\ y_{i2}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \end{pmatrix} - \begin{pmatrix} \alpha \\ \tau_1^y \\ \tau_1^x \\ \vdots \\ \tau_T^x \\ \gamma\tau_1^y + \tau_2^x\beta + \sigma^2_\mu \end{pmatrix}.$$

Several remarks are noteworthy. First, $g_{i,1}^D(\gamma,\beta')$ incorporates the zero-mean condition of ε_{it} (Assumption 2.2) through $\Delta y_i - \gamma\Delta y_{i,-1} - \Delta\tilde{x}_i\beta$, so in this regard there might be a risk of misspecification bias only if $E(\Delta u_{it}) \neq 0$. Second, $g_{i,1}^D(\gamma,\beta')$ does not depend on $\theta_{\setminus\gamma\beta}$. Third, $g_i^D(\theta)$ can be obtained by applying a transformation to $g_i(\theta)$ along the lines of eqs. (10)–(11): there is a $h \times h$ matrix $\mathrm{D}(\gamma,\beta')$ such that $g_i^D(\theta) = \mathrm{D}(\gamma,\beta')g_i(\theta)$. Fourth, this matrix is nonsingular because none of the rows of $g_i^D(\theta)$ can be written as a linear combination of others. Hence, our estimator $\hat{\theta}$ can be expressed as

$$\begin{aligned}\hat{\theta} &= \underset{\theta\in\Theta}{\operatorname{argmin}}\ \bar{g}(\theta)'\mathrm{D}(\gamma,\beta')'\left[\mathrm{D}(\gamma,\beta')\dot{\Omega}\mathrm{D}(\gamma,\beta')'\right]^{-1}\mathrm{D}(\gamma,\beta')\bar{g}(\theta) \\ &= \underset{\theta\in\Theta}{\operatorname{argmin}}\ \bar{g}^D(\theta)'\left[\dot{\Omega}^D(\gamma,\beta')\right]^{-1}\bar{g}^D(\theta),\end{aligned} \tag{12}$$

where $\bar{g}^D(\theta) = (1/N)\sum_{i=1}^N g_i^D(\theta)$ and $\dot{\Omega}^D(\gamma,\beta') = \mathrm{D}(\gamma,\beta')\dot{\Omega}\mathrm{D}(\gamma,\beta')'$; note that $\dot{\Omega}^D(\gamma,\beta')$ is a consistent estimator of $\Omega^D(\gamma,\beta') = \mathrm{D}(\gamma,\beta')\Omega\mathrm{D}(\gamma,\beta')'$. From expression (12), our estimator $\hat{\theta}$ can be regarded as a *continuous updating* GMM estimator based on $g_i^D(\theta)$ using $\dot{\Omega}^D(\gamma,\beta')^{-1}$ as weighting matrix. We refer to Hall (2005, Sec. 3.7) for further discussion on these types of GMM estimators. Fifth, the asymptotic distribution of $\hat{\theta}$, as well as its asymptotic variance, has been obtained in Theorem 1.

To compare the asymptotic variances, let $\Sigma_{\gamma\beta}$ denote the $(k+1)\times(k+1)$ sub-matrix of Σ associated with $(\gamma_\circ, \beta'_\circ)$, i.e. the asymptotic variance-covariance matrix of $(\hat{\gamma}, \hat{\beta}')$. Theorem 2 below compares the asymptotic variances $\Sigma_{\gamma\beta}$ and $\Sigma^{AS}_{\gamma\beta}$; $\Sigma^{AB}_{\gamma\beta}$ is omitted from the comparison because we already know that the Ahn-Schmidt estimator is asymptotically as efficient as Arellano-Bond's, i.e. the difference $\Sigma^{AB}_{\gamma\beta} - \Sigma^{AS}_{\gamma\beta}$ is positive semi-definite. Denote $G^D_{1,2} = E(\Delta\tilde{x}_i)$, $\Omega^D_{11} = E[g_{i,1}^D(\gamma_\circ,\beta'_\circ)g_{i,1}^D(\gamma_\circ,\beta'_\circ)']$, and partition

$$\Omega^D_{11} = \begin{pmatrix} \underset{\tilde{h}'\times\tilde{h}'}{\Omega^{AS}} & \underset{\tilde{h}'\times(T-2)}{\Omega^D_{11,12}} \\ \underset{(T-2)\times\tilde{h}'}{\Omega^{D\prime}_{11,12}} & \underset{(T-2)\times(T-2)}{\Omega^D_{11,22}} \end{pmatrix}.$$

Theorem 2
Under Assumption 1–Assumption 5,

$$\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AS})^{-1} = [G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{D} - G_{1,2}^{D\prime}]\Upsilon[G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{D} - G_{1,2}^{D\prime}]',$$

being $\Upsilon \equiv \left[\Omega_{11,22}^{D} - \Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{D}\right]^{-1}$ *positive definite.*

Observe that our estimator has smaller (asymptotic) variance than Arellano-Bond estimator, i.e. the difference $\Sigma_{\gamma\beta}^{AS} - \Sigma_{\gamma\beta}$ is positive definite, if and only if $\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AS})^{-1}$ is positive definite. So from Theorem 2, the efficiency gain of using moment conditions in levels depends crucially on the terms $\Omega_{11,12}^{D}$ and $G_{1,2}^{D}$: our estimator of is asymptotically equivalent to Ahn-Schmidt's if and only if $G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{D} = G_{1,2}^{D\prime}$. As an illustration, we provide the following example.

Example 3
($T = 4$ & $k = 0$, cont.).
We have $G_{1,2}^{D\prime} = E[\Delta y_i]$ and

$$\Omega_{11,12}^{D} = E\left[\begin{pmatrix} y_{i1}\Delta u_{i3} \\ y_{i1}\Delta u_{i4} \\ y_{i2}\Delta u_{i4} \\ (\alpha_\circ + u_{i4})\Delta u_{i3} \end{pmatrix}(\Delta u_{i3}\ \Delta u_{i4})\right].$$

In this example, the efficiency gain of our estimator comes from two sources: (i) the variation in the unconditional mean of the dependent variable, which is captured by $G_{1,2}^{D}$; (ii) the interactions between $(\Delta u_{i3}, \Delta u_{i4}) = (\Delta\varepsilon_{i3}, \Delta\varepsilon_{i4})$ and (u_{i4}, y_{i1}, y_{i2}), which is captured by $\Omega_{11,12}^{D}$. There is no efficiency gain, e.g. when $\Omega_{11,12}^{D} = 0_{4\times 2}$ and $G_{1,2}^{D} = 0_{2\times 1}$. These two conditions hold under very particular circumstances, e.g. if $\{y_{i1}, \mu_i, \varepsilon_{i2}, \varepsilon_{i3}, \varepsilon_{i4}\}$ are independent of each other and $E(y_{i1}) = \alpha_\circ = 0$.

6 Monte Carlo Experiments

In this section we study the finite sample performance of the proposed estimator. To facilitate comparisons and replicability, we use the design in Yamagata (2008) and Wu and Zhu (2012).

Consider the dynamic panel data model

$$\begin{aligned} y_{it} &= \alpha + \gamma y_{it-1} + x_{it}\beta + \mu_i + \varepsilon_{it}, \\ x_{it} &= \delta_t + 0.5x_{it-1} + 0.5\varepsilon_{it-1} + \rho\mu_i + v_{it}, \end{aligned}$$

with $i = 1, 2, \ldots, N$ and $t = -48, -47, \ldots, T$. We set the initial conditions to $y_{i,-49} = x_{i,-49} = 0$ and discard the initial 50 observations.

We let $\mu_i \overset{i.i.d.}{\sim} N(0, \sigma_\mu^2)$, $\varepsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$, $v_{it} \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$. We fix $\alpha = 0$, $\beta = 1$, $\sigma_\mu^2 = 53/108$, $\sigma_\varepsilon^2 = 55/20$ and $\sigma_v^2 = 1$ as in Yamagata (2008, 141) and different parameter values specifications: $\gamma \in \{0, 0.25, 0.5, 0.75\}$, $\rho \in \{0, 0.25\}$. The sample sizes considered are $N \in \{200, 300\}$ and $T \in \{4, 8\}$.[1] We consider two different scenarios. First, we consider x_{it} stationary using $\delta_t = 0$. Then, we use $\delta_t = t$ for which x_{it} is trend non-stationary. The parameters (α, γ, β) are estimated using three different GMM estimators: (i) our proposed GMM model, (ii) Ahn and Schmidt (1995) (AS) estimator, and (iii) Arellano and Bond (1991) (AB) estimator. The number of Monte Carlo repetitions is 2000. In all cases we report the empirical bias and root-mean square error (RMSE).

Consider first the model where x_{it} is stationary and has zero mean. Table 1 and Table 2 report bias and RMSE for estimating γ and β, respectively. For this case, the three procedures under comparison perform similarly as expected from previous section's discussion. Still, it is interesting to remark that throughout the different parameter configurations, the proposed estimator is always ranked above AS and below AB in terms of bias and RMSE. Results are presented graphically in Figure 1. The figure shows RMSE for the three estimators and each point corresponds to each row of the Table 1 and Table 2. The top graph corresponds to the estimation of γ (Table 1) and the bottom graph to the estimation of β (Table 2). The graph shows that the three estimators behave similarly for this case.

Table 1: Bias and RMSE for estimating γ, x_{it} stationary.

N	T	γ	ρ	Bias			RMSE		
				New	AS	AB	New	AS	AB
200	4	0	0	−0.006	−0.0037	−0.0088	0.0575	0.0534	0.0616
200	4	0	0.25	−0.005	−0.0024	−0.008	0.0579	0.0537	0.0639
200	4	0.25	0	−0.0106	−0.0072	−0.0143	0.0659	0.0606	0.0707
200	4	0.25	0.25	−0.0087	−0.0056	−0.0136	0.0674	0.0625	0.0743
200	4	0.5	0	−0.0118	−0.0077	−0.0171	0.0708	0.0656	0.0787
200	4	0.5	0.25	−0.0136	−0.0089	−0.0203	0.0729	0.0679	0.0844
200	4	0.75	0	−0.0136	−0.0094	−0.0199	0.0692	0.0635	0.0766
200	4	0.75	0.25	−0.0101	−0.0063	−0.0201	0.0744	0.0661	0.0837
200	8	0	0	−0.0053	−0.0016	−0.0094	0.0318	0.0229	0.0353
200	8	0	0.25	−0.0053	−0.0014	−0.0098	0.0299	0.0215	0.0334
200	8	0.25	0	−0.007	−0.0026	−0.0117	0.0302	0.0214	0.0347
200	8	0.25	0.25	−0.0065	−0.002	−0.0118	0.0302	0.0215	0.0347
200	8	0.5	0	−0.0092	−0.0036	−0.0151	0.0293	0.0202	0.0347
200	8	0.5	0.25	−0.0095	−0.0029	−0.017	0.0306	0.0209	0.0378
200	8	0.75	0	−0.0105	−0.0042	−0.0171	0.0274	0.0183	0.034
200	8	0.75	0.25	−0.0104	−0.0033	−0.0193	0.0272	0.0181	0.036
300	4	0	0	−0.0041	−0.0027	−0.005	0.0454	0.0434	0.0496
300	4	0	0.25	−0.0041	−0.0023	−0.0064	0.0475	0.0451	0.0518
300	4	0.25	0	−0.0055	−0.0036	−0.0069	0.0545	0.0518	0.0585
300	4	0.25	0.25	−0.0064	−0.0044	−0.0094	0.0542	0.0515	0.0589
300	4	0.5	0	−0.0074	−0.005	−0.0109	0.0585	0.0561	0.0632
300	4	0.5	0.25	−0.0054	−0.0027	−0.0093	0.0601	0.0573	0.0673
300	4	0.75	0	−0.0068	−0.0051	−0.0111	0.0555	0.0517	0.0604
300	4	0.75	0.25	−0.0081	−0.0063	−0.0148	0.0601	0.0538	0.0657
300	8	0	0	−0.003	−0.0013	−0.0063	0.0244	0.0194	0.0273
300	8	0	0.25	−0.0032	−0.0012	−0.0067	0.0248	0.0197	0.028
300	8	0.25	0	−0.0045	−0.0022	−0.0084	0.0241	0.0194	0.0276
300	8	0.25	0.25	−0.0051	−0.0023	−0.0101	0.0247	0.0195	0.0292
300	8	0.5	0	−0.0055	−0.0027	−0.0101	0.0238	0.0185	0.0277
300	8	0.5	0.25	−0.0054	−0.0024	−0.0111	0.0236	0.0184	0.0293
300	8	0.75	0	−0.0066	−0.0035	−0.0118	0.0208	0.0158	0.026
300	8	0.75	0.25	−0.0055	−0.0023	−0.0131	0.0216	0.0166	0.0283

Table 2: Bias and RMSE for estimating β, x_{it} stationary.

N	T	γ	ρ	Bias			RMSE		
				New	AS	AB	New	AS	AB
200	4	0	0	−0.0015	−0.001	−0.0019	0.0575	0.054	0.0583
200	4	0	0.25	−0.0043	−0.0038	−0.0061	0.0576	0.0535	0.059
200	4	0.25	0	−0.0053	−0.0047	−0.0066	0.0589	0.0548	0.0611
200	4	0.25	0.25	−0.0038	−0.0032	−0.0069	0.0579	0.0543	0.0608
200	4	0.5	0	−0.0071	−0.0053	−0.01	0.0614	0.0581	0.0661
200	4	0.5	0.25	−0.0075	−0.0058	−0.0127	0.064	0.06	0.0707
200	4	0.75	0	−0.0093	−0.0061	−0.0141	0.068	0.0648	0.0767
200	4	0.75	0.25	−0.0086	−0.0052	−0.0174	0.0757	0.0697	0.0866
200	8	0	0	−0.0019	−0.0012	−0.002	0.0254	0.0189	0.0254
200	8	0	0.25	−0.0017	−0.0008	−0.0025	0.0259	0.0191	0.026
200	8	0.25	0	−0.0018	−0.0009	−0.0021	0.025	0.0185	0.025
200	8	0.25	0.25	−0.0006	0.0001	−0.0017	0.0251	0.0187	0.0253
200	8	0.5	0	−0.0024	−0.0013	−0.0035	0.0251	0.0186	0.0254
200	8	0.5	0.25	−0.0029	−0.0008	−0.0053	0.0249	0.0184	0.026
200	8	0.75	0	−0.006	−0.0027	−0.0092	0.0257	0.0186	0.0279
200	8	0.75	0.25	−0.0055	−0.0014	−0.0107	0.0265	0.019	0.0304
300	4	0	0	−0.0026	−0.0025	−0.0025	0.0465	0.0445	0.0465

300	4	0	0.25	−0.0023	−0.0018	−0.0034	0.0468	0.0447	0.0486
300	4	0.25	0	−0.0025	−0.0021	−0.0028	0.0472	0.0454	0.0482
300	4	0.25	0.25	−0.0023	−0.0018	−0.0037	0.0477	0.0457	0.0489
300	4	0.5	0	−0.0027	−0.002	−0.0047	0.0511	0.0492	0.0534
300	4	0.5	0.25	−0.0037	−0.0026	−0.0067	0.0525	0.0502	0.0569
300	4	0.75	0	−0.0055	−0.0044	−0.0089	0.0569	0.0537	0.0612
300	4	0.75	0.25	−0.0067	−0.0052	−0.0128	0.0592	0.0547	0.0661
300	8	0	0	−0.0014	−0.0011	−0.0014	0.0212	0.0172	0.0212
300	8	0	0.25	−0.0005	−0.0003	−0.001	0.0208	0.0168	0.0209
300	8	0.25	0	−0.001	−0.0007	−0.001	0.0201	0.0164	0.0201
300	8	0.25	0.25	−0.0003	0.0001	−0.0013	0.0202	0.0165	0.0205
300	8	0.5	0	−0.0011	−0.0005	−0.002	0.0203	0.0164	0.0205
300	8	0.5	0.25	−0.0015	−0.0005	−0.0031	0.0203	0.0165	0.0208
300	8	0.75	0	−0.0037	−0.0023	−0.0064	0.0208	0.0166	0.0224
300	8	0.75	0.25	−0.0035	−0.0015	−0.0081	0.021	0.0165	0.0238

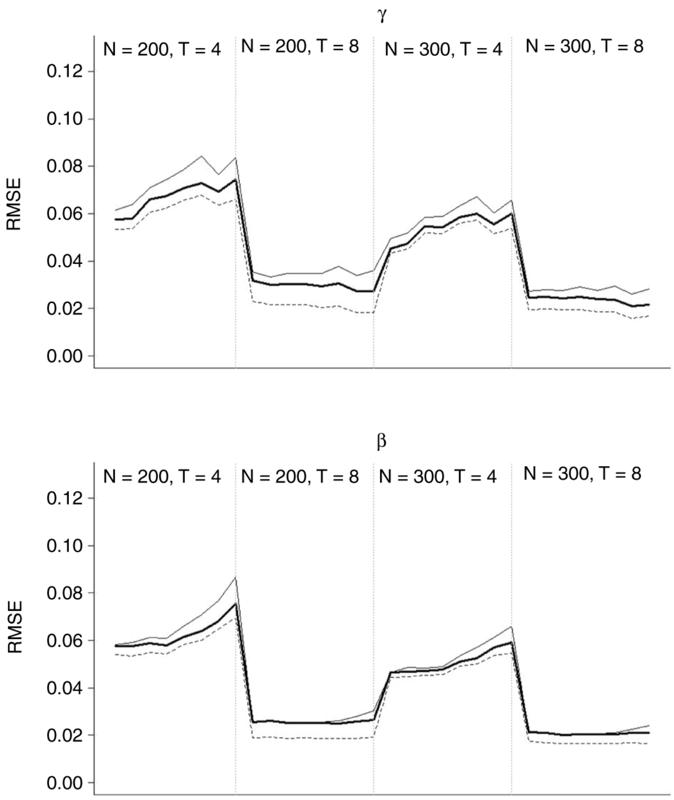

Figure 1: RMSE for the stationary x_{it} case. Notes: RMSE for the estimation of γ (top panel) and β (bottom panel) in the x_{it} stationary case. Thick line corresponds to our estimator, dashed line is Ahn-Schmidt and solid line is Arellano-Bond. Horizontal axis corresponds to rows in Table 1 and Table 2.

When x_{it} is non-stationary (Table 3 and Table 4 for γ and β, respectively), our proposed estimator systematically over-performs both AS and AB in terms of RMSE and bias for the "small T" case ($T = 4$). This result can be explained by the previous section's discussion. For example, when $\gamma = 0.25$ and $\rho = 0.25$, the RMSE of our procedure when estimating β is 0.0358, compared to 0.0801 of AS and 0.1157 of AB. Though qualitatively these differences persist for the $T = 8$ case, they are quantitatively smaller. Figure 2 presents the results of Table 3 and Table 4 graphically.

Table 3: Bias and RMSE for estimating γ, x_{it} non-stationary.

N	T	γ	ρ		Bias			RMSE	
				New	**AS**	**AB**	**New**	**AS**	**AB**
200	4	0	0	−0.0053	0.0029	−0.0062	0.0525	0.0696	0.0852
200	4	0	0.25	−0.0032	0.0035	−0.0049	0.0514	0.0706	0.0873
200	4	0.25	0	−0.0052	−0.005	−0.0091	0.0497	0.0759	0.1035
200	4	0.25	0.25	−0.0047	−0.0082	−0.0105	0.0507	0.0782	0.1062

200	4	0.5	0	−0.0021	−0.0069	−0.0051	0.0408	0.0577	0.0651
200	4	0.5	0.25	−0.0038	−0.0088	−0.0058	0.0386	0.0586	0.0677
200	4	0.75	0	−0.0013	−0.0008	−0.001	0.0233	0.0252	0.0272
200	4	0.75	0.25	0.0002	−0.0007	−0.0001	0.0228	0.0261	0.0286
200	8	0	0	−0.0038	−0.0016	−0.0074	0.0301	0.0235	0.0348
200	8	0	0.25	−0.0035	−0.0009	−0.0071	0.0285	0.0224	0.0333
200	8	0.25	0	−0.0038	−0.0018	−0.0068	0.026	0.0228	0.0329
200	8	0.25	0.25	−0.0041	−0.0018	−0.0074	0.026	0.0225	0.0326
200	8	0.5	0	−0.0032	−0.003	−0.0058	0.0205	0.0193	0.0286
200	8	0.5	0.25	−0.003	−0.003	−0.006	0.0202	0.0194	0.028
200	8	0.75	0	−0.0007	−0.0009	−0.0015	0.0107	0.0094	0.0142
200	8	0.75	0.25	−0.0009	−0.0014	−0.0014	0.0109	0.0097	0.0148
300	4	0	0	−0.0032	0.0026	−0.0042	0.0423	0.0598	0.0693
300	4	0	0.25	−0.003	0.0025	−0.006	0.0426	0.0608	0.0707
300	4	0.25	0	−0.0032	−0.0055	−0.0102	0.0413	0.0672	0.087
300	4	0.25	0.25	−0.0037	−0.0049	−0.0085	0.0394	0.0657	0.0853
300	4	0.5	0	−0.0027	−0.0061	−0.0048	0.0314	0.0482	0.052
300	4	0.5	0.25	−0.0009	−0.0052	−0.0021	0.0313	0.0488	0.054
300	4	0.75	0	−0.0005	−0.0007	−0.0005	0.0192	0.0211	0.0224
300	4	0.75	0.25	−0.0005	−0.001	−0.0004	0.0182	0.021	0.0223
300	8	0	0	−0.0022	−0.0009	−0.0051	0.0232	0.0204	0.0273
300	8	0	0.25	−0.0025	−0.0011	−0.0055	0.0234	0.0204	0.0278
300	8	0.25	0	−0.0027	−0.0017	−0.0053	0.0211	0.0206	0.0267
300	8	0.25	0.25	−0.0033	−0.0021	−0.0064	0.0209	0.0203	0.0271
300	8	0.5	0	−0.0021	−0.0026	−0.0042	0.0161	0.0178	0.0229
300	8	0.5	0.25	−0.0019	−0.0027	−0.0044	0.0159	0.0177	0.023
300	8	0.75	0	−0.0004	−0.0008	−0.0008	0.0085	0.0086	0.0114
300	8	0.75	0.25	−0.0002	−0.0009	−0.0007	0.0087	0.0088	0.012

Table 4: Bias and RMSE for estimating β, x_{it} non-stationary.

N	T	γ	ρ	Bias			RMSE		
				New	AS	AB	New	AS	AB
200	4	0	0	0.0015	−0.0046	0.0018	0.0329	0.0663	0.0839
200	4	0	0.25	0.0005	−0.0034	0.002	0.0329	0.0686	0.0881
200	4	0.25	0	0.0024	0.0048	0.0066	0.0368	0.0794	0.1138
200	4	0.25	0.25	0.0025	0.0094	0.009	0.0358	0.0801	0.1157
200	4	0.5	0	0.0011	0.0093	0.0051	0.04	0.0732	0.0844
200	4	0.5	0.25	0.0004	0.0093	0.0031	0.0377	0.0754	0.0895
200	4	0.75	0	0.0015	0.0011	0.0008	0.0435	0.0512	0.0561
200	4	0.75	0.25	0.0002	0.0028	0.0009	0.0429	0.0549	0.061
200	8	0	0	0.0018	0.0011	0.0037	0.0162	0.0144	0.0208
200	8	0	0.25	0.0017	0.0005	0.0035	0.0159	0.0142	0.0206
200	8	0.25	0	0.0022	0.0013	0.0043	0.0183	0.018	0.0257
200	8	0.25	0.25	0.0026	0.0013	0.0047	0.0181	0.0174	0.0253
200	8	0.5	0	0.0031	0.0035	0.0057	0.021	0.0216	0.0322
200	8	0.5	0.25	0.0027	0.0032	0.0058	0.0206	0.0214	0.0314
200	8	0.75	0	0.001	0.0018	0.0025	0.021	0.019	0.0296
200	8	0.75	0.25	0.0014	0.0028	0.0024	0.0216	0.0197	0.0311
300	4	0	0	0.0005	−0.0033	0.0017	0.0266	0.0597	0.0698
300	4	0	0.25	0.0013	−0.0023	0.0046	0.0254	0.0593	0.0702
300	4	0.25	0	0.0015	0.0065	0.0099	0.029	0.0705	0.0956
300	4	0.25	0.25	0.0015	0.0052	0.0071	0.0284	0.0698	0.0949
300	4	0.5	0	0.0013	0.0072	0.0043	0.0314	0.0636	0.0692
300	4	0.5	0.25	0	0.0073	0.0017	0.031	0.0641	0.0721
300	4	0.75	0	0.0003	0.0013	0.0005	0.0366	0.0434	0.0464
300	4	0.75	0.25	0.0004	0.0019	0.0002	0.0339	0.0432	0.0467
300	8	0	0	0.0008	0.0004	0.0024	0.0125	0.0129	0.0167
300	8	0	0.25	0.001	0.0005	0.0026	0.0127	0.0129	0.0171
300	8	0.25	0	0.0014	0.0011	0.0032	0.0145	0.0163	0.0209
300	8	0.25	0.25	0.0021	0.0016	0.0043	0.0145	0.0161	0.0213
300	8	0.5	0	0.0019	0.0028	0.0041	0.0163	0.0197	0.0257

300	8	0.5	0.25	0.0019	0.0031	0.0047	0.016	0.0198	0.0259
300	8	0.75	0	0.0007	0.0017	0.0015	0.0165	0.0174	0.0237
300	8	0.75	0.25	0.0004	0.0019	0.001	0.0171	0.0179	0.0253

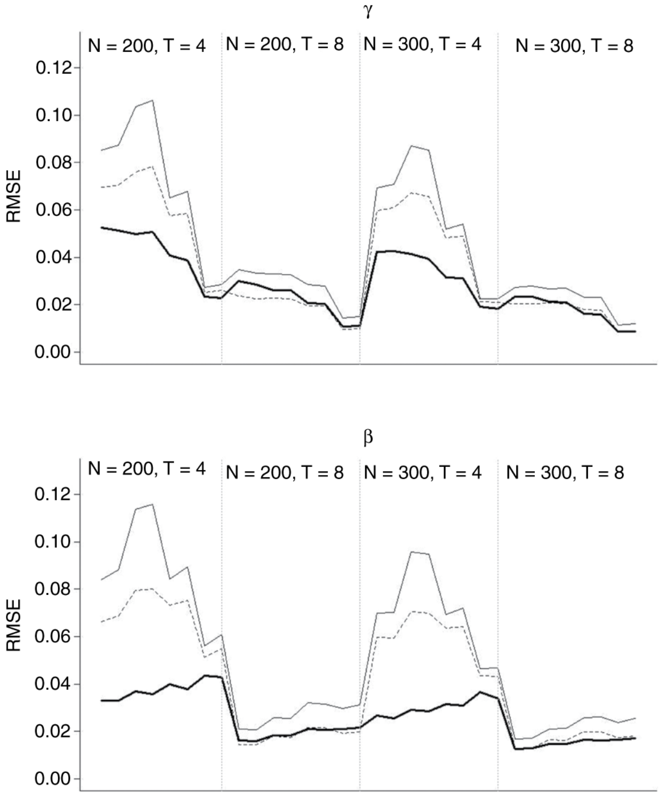

Figure 2: RMSE for the non-stationary x_{it} case. Notes: RMSE for the estimation of γ (top panel) and β (bottom panel) in the x_{it} non-stationary case. Thick line corresponds to our estimator, dashed line is Ahn-Schmidt and solid line is Arellano-Bond. Horizontal axis corresponds to rows in Table 3 and Table 4.

As a by-product, the proposed set of moment conditions leads to a simple procedure to estimate σ_μ^2 explicitly and conduct hypothesis tests. As mentioned before this is empirically relevant in cases where the interest lies in measuring the relative contribution of pure dynamic persistence vs. that derived from the presence of unobserved heterogeneity, as in the classic paper by Lillard and Willis (1978). See Arias, Marchionni, and Sosa-Escudero (2011) for a recent study along these lines.

In the context of this paper, the null hypothesis of no unobserved heterogeneity ($H_0 : \sigma_\mu^2 = 0$) implies that the τ parameters are themselves 0. The test would proceed by directly estimating σ_μ^2 and then testing down the corresponding null through a simple Wald-type test. Monte Carlo results for bias and RMSE for the estimation of σ_μ^2 are reported in Table 5 and Table 6 for the stationary and non-stationary case, respectively. We compare these results with those of Wu and Zhu (2012), who derive tests for random-effects for dynamic panel data models. In particular, our experimental design allows us to compare with the results of Wu and Zhu (2012, Table 1, Case (i)) for γ_2^μ with $(N, T) = (200, 8)$ and $(N, T) = (300, 8)$. This corresponds to $\rho = 0$ and $\gamma = 0.5$ in our experiments, for x_{it} stationary.

Table 5: Bias and RMSE for estimating σ_μ^2, x_{it} stationary.

N	*T*	γ	ρ	Bias	RMSE
200	4	0	0	0.026	0.0195
200	4	0	0.25	0.0297	0.024
200	4	0.25	0	0.0461	0.0332
200	4	0.25	0.25	0.0493	0.0447
200	4	0.5	0	0.0811	0.0632
200	4	0.5	0.25	0.1037	0.0951
200	4	0.75	0	0.1815	0.1919
200	4	0.75	0.25	0.2385	0.5908
200	8	0	0	−0.0035	0.0064
200	8	0	0.25	−0.0002	0.0071

200	8	0.25	0	0	0.0077
200	8	0.25	0.25	0.0051	0.0093
200	8	0.5	0	0.0138	0.0104
200	8	0.5	0.25	0.0246	0.0151
200	8	0.75	0	0.0504	0.0253
200	8	0.75	0.25	0.0644	0.0351
300	4	0	0	0.0156	0.0116
300	4	0	0.25	0.0235	0.0167
300	4	0.25	0	0.0277	0.0203
300	4	0.25	0.25	0.0365	0.0296
300	4	0.5	0	0.0516	0.0397
300	4	0.5	0.25	0.0594	0.0591
300	4	0.75	0	0.1093	0.1019
300	4	0.75	0.25	0.1643	0.2188
300	8	0	0	−0.007	0.0042
300	8	0	0.25	−0.0024	0.0046
300	8	0.25	0	−0.0002	0.0051
300	8	0.25	0.25	0.0026	0.0058
300	8	0.5	0	0.0066	0.007
300	8	0.5	0.25	0.0123	0.0089
300	8	0.75	0	0.0302	0.0139
300	8	0.75	0.25	0.0345	0.021

Table 6: Bias and RMSE for estimating σ^2_μ, x_{it} non-stationary.

N	T	γ	ρ	Bias	RMSE
200	4	0	0	0.015	0.0151
200	4	0	0.25	0.011	0.0139
200	4	0.25	0	0.0152	0.018
200	4	0.25	0.25	0.0142	0.0182
200	4	0.5	0	0.0191	0.0195
200	4	0.5	0.25	0.0196	0.019
200	4	0.75	0	0.0163	0.0181
200	4	0.75	0.25	0.0142	0.0191
200	8	0	0	−0.008	0.006
200	8	0	0.25	−0.0073	0.0062
200	8	0.25	0	−0.0092	0.0067
200	8	0.25	0.25	−0.0054	0.0069
200	8	0.5	0	−0.0071	0.007
200	8	0.5	0.25	−0.0055	0.0072
200	8	0.75	0	−0.0097	0.007
200	8	0.75	0.25	−0.0103	0.0067
300	4	0	0	0.0077	0.0092
300	4	0	0.25	0.0111	0.0102
300	4	0.25	0	0.0104	0.0112
300	4	0.25	0.25	0.0128	0.0116
300	4	0.5	0	0.0131	0.0115
300	4	0.5	0.25	0.0105	0.0122
300	4	0.75	0	0.0099	0.0118
300	4	0.75	0.25	0.0093	0.0124
300	8	0	0	−0.0099	0.004
300	8	0	0.25	−0.0055	0.0039
300	8	0.25	0	−0.0052	0.0044
300	8	0.25	0.25	−0.0048	0.0042
300	8	0.5	0	−0.0056	0.0047
300	8	0.5	0.25	−0.0042	0.0045
300	8	0.75	0	−0.0068	0.0044
300	8	0.75	0.25	−0.0081	0.0046

Our estimator has a bias of 0.0138 and 0.0066 for $(N, T) = (200, 8)$ and $(N, T) = (300, 8)$, respectively, significantly smaller than the bias of 0.044 and 0.035 for Wu and Zhu. In terms of RMSE, our estimator achieves 0.0104 and 0.007 for $(N, T) = (200, 8)$ and $(N, T) = (300, 8)$, respectively, again smaller than their 0.0143 and 0.0095. Note that although both bias and RMSE decrease with N and T, they increase when γ increases, pointing out that, as in Zincenko, Sosa-Escudero, and Montes-Rojas (2014), dynamic persistence and unobserved heterogeneity are confounding factors. In fact, our model can be seen as an extension of Zincenko, Sosa-Escudero, and Montes-Rojas (2014) to allow for an arbitrary covariance between the individual effects and the exogenous covariates.

7 Concluding Remarks

This paper proposes a simple framework for the estimation of dynamic panel models based on parameterizing the relationship between covariates and unobserved time invariant effects, in the spirit of Chamberlain's (1980, 1982) approach.

Such a perspective has been already adopted in many panel structures (in particular those involving qualitative data) and in some dynamic models. Our approach leads to a set of moment conditions that are embedded in a GMM framework to derive an asymptotically optimal estimator of the parameters of interest. The paper explicitly compares the proposed moment conditions and resulting estimator with those in classic papers like Arellano and Bond (1991) and Ahn and Schmidt (1995) (1995, 1997), implying no efficiency loss.

Though mostly of theoretical and modeling interest, Monte Carlo results suggest that the new estimator performs better (in terms of bias and RMSE) for the case of non-stationary covariates. Also, the framework leads to a simple variance estimator that can be used to test for the presence of unobserved effects. The derived procedure performs better than available alternatives like Wu and Zhu (2012).

There are several contexts in which our approach can be useful. For instance, the unequally-spaced dynamic panels considered by Sasaki and Xin (2017), among others. In this context, moments in levels may provide necessary and sufficient identification conditions, as well as an efficient estimator for the parameters of interest.

Acknowledgements

We thank Chengying Luo for excellent research assistance. We are also grateful to Editor Prof. Raffaella Giacomini, the Associate Editor and two anonymous referees for constructive comments that have improved the paper. Financial support from the Central Research Development Fund (University of Pittsburgh) and Programa RAICES (CONICET, Argentina) is gratefully acknowledge.

A Appendix

A.1 Consistent Estimator of Ω

To construct a consistent estimator of Ω, first, we propose a consistent (first-step) estimator of $\theta_\circ$. We suggest using $\dot{\theta} = (\dot{\alpha}, \dot{\gamma}, \dot{\beta}', \dot{\sigma}^2_\mu, \dot{\tau}')$, where $(\dot{\gamma}, \dot{\beta}')$ is the first-step Arellano-Bond estimator of $(\gamma_\circ, \beta'_\circ)$:

$$\begin{pmatrix} \dot{\gamma} \\ \dot{\beta} \end{pmatrix} = \left[\left(\frac{1}{N} \sum_{i=1}^{N} \Delta \mathbf{x}_i' \tilde{\mathbf{Z}}_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \Delta \mathbf{x}_i \right) \right]^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \Delta \mathbf{x}_i' \tilde{\mathbf{Z}}_i \right) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \Delta y_i \right).$$

Further, we propose

$$\dot{\alpha} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{it} - \dot{\gamma} y_{i,t-1} - x'_{it}\dot{\beta}),$$
$$\dot{\sigma}_{\mu}^2 = \frac{1}{N(T-2)} \sum_{i=1}^{N} \sum_{t=3}^{T} \dot{u}_{it}\dot{u}_{i,t-1},$$
$$\dot{\tau}_1^y = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i1}\dot{u}_{it},$$
$$\dot{\tau}_1^x = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=2}^{T} x_{i1}\dot{u}_{it},$$
$$\dot{\tau}_j^x = \frac{1}{N(T-j+1)} \sum_{i=1}^{N} \sum_{t=j}^{T} x_{ij}\dot{u}_{it} \text{ for } j \geq 2,$$

$\dot{\tau} = (\dot{\tau}_1^y, \dot{\tau}_1^{x\prime}, \ldots, \dot{\tau}_T^{x\prime})'$, and $\dot{u}_{it} = y_{it} - \dot{\alpha} - \dot{\gamma} y_{i,t-1} - x'_{it}\dot{\beta}$. The natural estimator of Ω then becomes

$$\dot{\Omega} = \frac{1}{N} \sum_{i=1}^{N} g_i(\dot{\theta}) g_i(\dot{\theta})'.$$

The next lemma establishes consistency.

Lemma 3

Under Assumption 1–Assumption 3, $\dot{\theta} \xrightarrow{P} \theta_\circ$ *and* $\dot{\Omega} \xrightarrow{P} \Omega$.

Proof.

Consistency of $(\gamma_\circ, \beta'_\circ)$ follows by standard arguments. First, note that

$$\frac{1}{N} \sum_{i=1}^{N} \Delta \mathbf{x}'_i \tilde{\mathbf{Z}}_i$$

and $(1/N) \sum_{i=1}^{N} \tilde{\mathbf{Z}}'_i \Delta y_i$ converge in probability to $E(\Delta \mathbf{x}'_i \tilde{\mathbf{Z}}_i)$ and $E(\tilde{\mathbf{Z}}'_i \Delta y_i)$, respectively; as a result,

$$\begin{pmatrix} \dot{\gamma} \\ \dot{\beta} \end{pmatrix} \xrightarrow{P} \left[E\left(\Delta \mathbf{x}'_i \tilde{\mathbf{Z}}_i\right) E\left(\tilde{\mathbf{Z}}'_i \Delta \mathbf{x}_i\right)\right]^{-1} E\left(\Delta \mathbf{x}'_i \tilde{\mathbf{Z}}_i\right) E\left(\tilde{\mathbf{Z}}'_i \Delta y_i\right). \quad (13)$$

Then, the desired result follows from

$$E\left(\tilde{\mathbf{Z}}'_i \Delta y_i\right) = E\left(\tilde{\mathbf{Z}}'_i \Delta \mathbf{x}_i\right) \begin{pmatrix} \gamma_\circ \\ \beta_\circ \end{pmatrix}. \quad (14)$$

To show that the other estimators are consistent, we use Lemma 4.3 in Newey and McFadden (1994). Define the function

$$f_\alpha([(y_{i1}, y'_i)', (x_{i1}, x_i)']; \gamma, \beta') = \frac{1}{T-1} \sum_{t=2}^{T} (y_{it} - \gamma y_{i,t-1} - x'_{it}\beta)$$

and write

$$\dot{\alpha} = \frac{1}{N} \sum_{i=1}^{N} f_\alpha([(y_{i1}, y'_i)', (x_{i1}, x_i)']; \dot{\gamma}, \dot{\beta}').$$

Then, observe that $E\left\{f_\alpha([(y_{i1}, y'_i)', (x_{i1}, x_i)']; \gamma_\circ, \beta'_\circ)\right\} = \alpha_0$ and, by Lemma 4.3 of Newey and McFadden (1994), we have that

$$\frac{1}{N}\sum_{i=1}^{N} f_\alpha([(y_{i1}, y_i')', (x_{i1}, x_i)']; \dot{\gamma}, \dot{\beta}') \xrightarrow{P} E\{f_\alpha([(y_{i1}, y_i')', (x_{i1}, x_i)']; \gamma_\circ, \beta_\circ)\}$$

if the following conditions hold: (i) $f_\alpha([(y_{i1}, y_i')', (x_{i1}, x_i)']; \gamma, \beta')$ is continuous at $(\gamma_\circ, \beta_\circ')$ with probability one; (ii) there is neighborhood $\mathcal{B}$ of $(\gamma_\circ, \beta_\circ')$ such that

$$E\left[\sup_{(\gamma,\beta')\in\mathcal{B}} |f_\alpha([(y_{i1}, y_i')', (x_{i1}, x_i)']; \gamma, \beta')|\right] < \infty;$$

(iii) $(\dot{\gamma}, \dot{\beta}') \xrightarrow{P} (\gamma_\circ, \beta_\circ')$. Clearly, (i) holds because $f_\alpha([(y_{i1}, y_i')', (x_{i1}, x_i)']; \gamma, \beta')$ is linear in (γ, β') for any realization of $[(y_{i1}, y_i')', (x_{i1}, x_i)']$. Condition (ii) holds for any such a neighborhood because $E\left|[(y_{i1}, y_i')', (x_{i1}, x_i)']\right|$ is a finite matrix and T is also finite. Condition (iii) holds from eqs. (13)–(14).

Proceeding in a similar manner with the rest of the estimators, we obtain the desired results. In particular, being $\|\cdot\|_\infty$ the elementwise sup-norm of a matrix or vector, we highlight that $E[\sup_{\theta\in\mathcal{N}} \|g_i(\theta)g_i(\theta)'\|_\infty] < \infty$ for any neighborhood $\mathcal{N}$ containing $\theta_\circ$ because $(y_{i1}, y_i', x_{i1}', \dots, x_{iT}')$ has finite second fourth moment. This is an immediate implication of Assumption 1, which is discussed in sec. 2. ⊡

A.2 Closed-Form Expression for $\nabla_\theta\Psi(\theta)$

Consider any $\theta \in \Theta$. The Jacobian matrix of $\Psi(\theta)$, denoted by $\nabla_\theta\Psi(\theta)$, can partitioned in 5 blocks:

$$\underset{h\times[k(T+1)+4]}{{}_{\theta}\Psi(\theta)} = \begin{pmatrix} \underset{h\times 1}{{}_{\alpha}\Psi(\theta)} & \underset{h\times 1}{{}_{\gamma}\Psi(\theta)} & \underset{h\times k}{{}_{\beta}\Psi(\theta)} & \underset{h\times 1}{{}_{\sigma_\mu^2}\Psi(\theta)} & \underset{h\times(kT+1)}{{}_{\tau}\Psi(\theta)} \end{pmatrix}$$
$$\equiv \begin{pmatrix} \frac{\partial\Psi(\theta)}{\partial\alpha} & \frac{\partial\Psi(\theta)}{\partial\gamma} & \frac{\partial\Psi(\theta)}{\partial\beta'} & \frac{\partial\Psi(\theta)}{\partial\sigma_\mu^2} & \frac{\partial\Psi(\theta)}{\partial\tau'} \end{pmatrix}.$$

It follows immediately that ${}_{\alpha}\Psi(\theta) = 0_{h\times 1}$. We provide below closed-form expressions for ${}_{\gamma}\Psi(\theta)$, ${}_{\beta}\Psi(\theta)$, ${}_{\sigma_\mu^2}\Psi(\theta)$, and ${}_{\tau}\Psi(\theta)$. Such expressions will be employed to compute ${}_{\theta}g_i(\theta)$.

First, observe that

$$\frac{\partial\psi_t(\theta)}{\partial\gamma} = (t-1)\gamma^{t-2}\tau_1^y + \sum_{l=2}^{t}(t-l)\gamma^{t-l-1}\tau_{x,l}'\beta + \left\{(t-1)\frac{\gamma^{t-2}}{\gamma-1} - \frac{\gamma^{t-1}-1}{(\gamma-1)^2}\right\}\sigma_\mu^2.$$

Then, we can write

$${}_{\gamma}\Psi(\theta) = \begin{pmatrix} 0_{(T-1)\times 1} \\ {}_{\gamma}\Psi_Y(\theta) \\ 0_{h_x\times 1} \end{pmatrix},$$

where ${}_{\gamma}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\gamma$ is a $h_y \times 1$ vector that has the following form: 0 occupies the positions $\{[t(t-1)/2] + 1 : t = 1, \dots, T-1\}$, $\partial\psi_2(\theta)/\partial\gamma$ occupies positions $\{[t(t-1)/2] + 2 : t = 2, \dots, T-1\}$, and in general $\partial\psi_j(\theta)/\partial\gamma$ occupies positions $\{[t(t-1)/2] + j : t = j, \dots, T-1\}$ for $2 \le j \le T-1$.

Second, note that

$$\underset{1\times k}{\frac{\partial\psi_t(\theta)}{\partial\beta'}} = \sum_{l=2}^{t}\gamma^{t-l}\tau_l^{x\prime}.$$

Then,

$${}_{\beta}\Psi(\theta) = \begin{pmatrix} 0_{(T-1)\times k} \\ {}_{\beta}\Psi_Y(\theta) \\ 0_{h_x\times k} \end{pmatrix},$$

where ${}_{\beta}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\beta'$ is a $h_y \times k$ matrix whose rows can be constructed as in ${}_{\gamma}\Psi_Y(\theta)$. Proceeding in a similar manner, we can also construct ${}_{\sigma^2_\mu}\Psi(\theta)$.

Next consider ${}_{\tau}\Psi(\theta)$. We write

$$ {}_{\tau}\Psi(\theta) = \begin{pmatrix} 0_{(T-1)\times 1} & 0_{(T-1)\times k} & \dots & 0_{(T-1)\times k} & \dots & 0_{(T-1)\times k} \\ {}_{\tau^y_1}\Psi_Y(\theta) & {}_{\tau_{x,1}}\Psi_Y(\theta) & \dots & {}_{\tau_{x,t}}\Psi_Y(\theta) & \dots & {}_{\tau_{x,T}}\Psi_Y(\theta) \\ 0_{h_x\times 1} & {}_{\tau_{x,1}}\Psi_X(\theta) & \dots & {}_{\tau_{x,t}}\Psi_X(\theta) & \dots & {}_{\tau_{x,T}}\Psi_X(\theta) \end{pmatrix}, $$

where ${}_{\tau^y_1}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\tau^y_1$, ${}_{\tau_{x,t}}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\tau_{x,t}$ and

$$ {}_{\tau_{x,t}}\Psi_X(\theta) \equiv \partial\Psi_X(\theta)/\partial\tau_{x,t}. $$

The dimensions of these sub-matrices are $h_y \times 1$, $h_y \times k$ and $h_x \times k$, respectively. They can be constructed following previous steps. In particular, if ${}_{\tau_{x,t}}\Psi_X^{(j_1:j_2,:)}(\theta)$ denote the sub-matrix of ${}_{\tau_{x,t}}\Psi_Y(\theta)$ from row j_1 to j_2 and containing all columns, then

$$ {}_{\tau_{x,t}}\Psi_X^{(j_1:j_2,:)}(\theta) = I_{k\times k} $$

for each $(j_1, j_2) \in \{(k[t-2+l(l+1)/2]+1, k[t-1+l(l+1)/2]) : l = \max\{t-1,1\},\dots,T-1\}$, whereas the remaining elements of ${}_{\tau_{x,t}}\Psi_Y(\theta)$ are all equal to 0.

A.3 Proofs

The section of the Appendix contains the proof of the lemmas and theorems stated in the body of the text.

A.3.1 Proof of Lemma Lemma 1

From eqs. (7)–(8), $\theta_\circ$ is a solution of $E[g_i(\theta)] = 0_{h\times 1}$. We show that $\theta_\circ$ is indeed the unique solution. Let $\tilde\theta = (\tilde\alpha, \tilde\gamma, \tilde\beta', \tilde\sigma^2_\mu, \tilde\tau')$ satisfy $E[g_i(\tilde\theta)] = 0_{h\times 1}$.

We prove first that $(\tilde\gamma, \tilde\beta') = (\gamma_\circ, \beta'_\circ)$. Let $\mathbf{Z}_i^{(t,l)}$ and $\tilde{\mathbf{Z}}_i^{(t,l)}$ denote the (t,l)-coefficient of $\mathbf{Z}_i$ and $\tilde{\mathbf{Z}}_i$, respectively. Define the mappings $\mathscr{J} : \{1,\dots,h\} \to \{1,\dots,T-1\}$ and $\tilde{\mathscr{J}} : \{1,\dots,\tilde h\} \to \{1,\dots,T-2\}$ such that

$$ \mathbf{Z}_i^{(\mathscr{J}(l),l)} \neq 0 \text{ and } \tilde{\mathbf{Z}}_i^{(\tilde{\mathscr{J}}(l),l)} \neq 0. $$

Essentially, $\mathscr{J}(l)$ (or $\tilde{\mathscr{J}}(l)$) provides the number of the row that contains the nonzero element of column l of $\mathbf{Z}_i$ (or $\tilde{\mathbf{Z}}_i$). Note that both $\mathscr{J}(l)$ and $\tilde{\mathscr{J}}(l)$ are well-defined as each column of $\mathbf{Z}_i$ and $\tilde{\mathbf{Z}}_i$ contains only one nonzero coefficient. Now, for a given $l = 1,\dots,\tilde h$, define further $(\mathscr{L}_1(l), \mathscr{L}_2(l)) \in \{1,\dots,h\}^2$ such that

$$ \tilde{\mathbf{Z}}_i^{(\tilde{\mathscr{J}}(l),l)} = \mathbf{Z}_i^{(\tilde{\mathscr{J}}(l),\mathscr{L}_1(l))} = \mathbf{Z}_i^{(\tilde{\mathscr{J}}(l)+1,\mathscr{L}_2(l))}. $$

Observe that both $\mathscr{L}_1(l)$ and $\mathscr{L}_2(l)$ are well-defined as $\tilde{\mathbf{Z}}_i$ is a submatrix of $\mathbf{Z}_i$ and also each row of $\mathbf{Z}_i$ and $\tilde{\mathbf{Z}}_i$ does not contain nonzero repeated elements. Moreover, we must have $\mathscr{L}_1(l) < \mathscr{L}_2(l)$ by construction of $\mathbf{Z}_i$. Then, let D^{AB} be nonstochastic $\tilde h \times h$ matrix whose components are given by

$$ \left[\mathrm{D}^{AB}\right]^{(\tilde l,l)} = \begin{cases} -1 & \text{if } l = \mathscr{L}_1(\tilde l), \\ 1 & \text{if } l = \mathscr{L}_2(\tilde l), \\ 0 & \text{otherwise.} \end{cases} $$

For instance, when $T = 3$ and $k = 1$, we have

$$ \underset{3\times 10}{\mathrm{D}^{AB}} = \begin{pmatrix} 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}. $$

By construction, D^{AB} must satisfy

$$\mathrm{D}^{AB}\mathbf{Z}_i'(y_i - \mathbf{x}_i\kappa) = \tilde{\mathbf{Z}}_i'\Delta y_i - \tilde{\mathbf{Z}}_i'\Delta\mathbf{x}_i\begin{pmatrix}\gamma\\ \beta\end{pmatrix} \tag{15}$$

as well as $\mathrm{D}^{AB}\Psi(\theta) = 0_{\tilde{h}\times 1}$ (see eqs. (5)–(7)). We refer to Ahn and Schmidt (1995, Appendix A.2) for further discussion and examples about these results. Observe that $E[\mathrm{D}^{AB}g_i(\tilde{\theta})] = \mathrm{D}^{AB}E[g_i(\tilde{\theta})] = 0_{\tilde{h}\times 1}$, so eq. (15) yields Arellano and Bond's (1991) system of linear equations:

$$E\left(\tilde{\mathbf{Z}}_i'\Delta\mathbf{x}_i\right)\begin{pmatrix}\tilde{\gamma}\\ \tilde{\beta}\end{pmatrix} = E\left(\tilde{\mathbf{Z}}_i'\Delta y_i\right). \tag{16}$$

Since $E(\tilde{\mathbf{Z}}_i\Delta\mathbf{x}_i)$ has full rank (Assumption 3), there is a unique solution to this system of (linear) equations and, as a result, we must have $(\tilde{\gamma}, \tilde{\beta}') = (\gamma_\circ, \beta_\circ')$.

Regarding the other parameters, using the first equation of the system $E[g_i(\tilde{\theta})] = 0_{h\times 1}$, we obtain

$$E(y_{i2}) - \tilde{\alpha} - \tilde{\gamma}E(y_{i1}) - E(x_{i2}')\tilde{\beta} = 0$$

and therefore $\tilde{\alpha} = E(y_{i2}) - \tilde{\gamma}E(y_{i1}) - E(x_{i2}')\tilde{\beta} = E(y_{i2}) - \gamma_\circ E(y_{i1}) - E(x_{i2}')\beta_\circ = \alpha_\circ$. Using also the T-th equation of $E[g_i(\tilde{\theta})] = 0$, it follows that

$$\begin{aligned}\tilde{\tau}_1^y &= E\left[y_{i1}\left(y_{i2} - \tilde{\alpha} - \tilde{\gamma}y_{i1} - x_{i1}'\tilde{\beta}\right)\right]\\ &= E\left[y_{i1}\left(y_{i2} - \alpha_\circ - \gamma_\circ y_{i1} - x_{i1}'\beta_\circ\right)\right] = E(y_{i1}u_{i2}) = E(y_{i1}\mu_i) = \tau_{1\circ}^y.\end{aligned}$$

Proceeding in a similar manner, we can prove that $\tilde{\tau}_t^x = \tau_{t\circ}^x$ for every $t = 1, \dots, T$. Finally, exploiting the $(T+2)$th equation of $E[g_i(\tilde{\theta})] = 0_{h\times 1}$ related to expression (3), we obtain

$$E\left[y_{i2}\left(y_{i2} - \tilde{\alpha} - \tilde{\gamma}y_{i1} - x_{i1}'\tilde{\beta}\right)\right] - \left(\tilde{\gamma}\tilde{\tau}_1^y + \tilde{\tau}_2^{x\prime}\tilde{\beta} + \tilde{\sigma}_\mu^2\right) = 0$$

and therefore

$$\begin{aligned}\tilde{\sigma}_\mu^2 &= E\left[y_{i2}\left(y_{i2} - \tilde{\alpha} - \tilde{\gamma}y_{i1} - x_{i1}'\tilde{\beta}\right)\right] - \left(\tilde{\gamma}\tilde{\tau}_1^y + \tilde{\tau}_2^{x\prime}\tilde{\beta}\right)\\ &= E\left[y_{i2}\left(y_{i2} - \alpha_\circ - \gamma_\circ y_{i1} - x_{i1}'\beta_\circ\right)\right] - \left(\gamma_\circ\tilde{\tau}_{1\circ}^y + \tau_{2\circ}^{x\prime}\beta_\circ\right)\\ &= \sigma_{\mu\circ}^2.\end{aligned}$$

□

A.3.2 Proof of Theorem Theorem 1

1. By Theorem 2.6 in Newey and McFadden (1994), to establish consistency it suffices to check that the following conditions are satisfied:[2]

   (i) $\dot{\Omega} \xrightarrow{P} \Omega$ and Ω is positive definite;
   (ii) $E[g_i(\theta)] = 0$ if and only if $\theta = \theta_\circ$;
   (iii) $\theta_\circ \in$ interior(Θ) for some compact set Θ;
   (iv) $g_i(\cdot)$ is a continuously differentiable on interior(Θ) with probability one;
   (v) $E[\sup_{\theta\in\Theta}\|g_i(\theta)\|_\infty^2]$ is finite.

   Conditions (i) and (ii) follows immediately from Assumption 4 and Lemma 1, respectively. For condition (iii), we can take any compact set of $\mathbb{R}\times(-1,1)\times\mathbb{R}^k\times\mathbb{R}\times\mathbb{R}^{kT+1}$ containing $\theta_\circ$. Note that the functions $g_i(\cdot)$ and $\Psi(\cdot)$ are well-defined on $\mathbb{R}\times(-1,1)\times\mathbb{R}^k\times\mathbb{R}\times\mathbb{R}^{kT+1}$. Condition (iv) holds because $g_i(\cdot)$ is continuously differentiable on $\mathbb{R}\times(-1,1)\times\mathbb{R}^k\times\mathbb{R}\times\mathbb{R}^{kT+1}$ for any realization of $[(y_{i1}, y_i')', (x_{i1}, x_i)']$. In particular, closed-forms expression for the Jacobian matrix of $\Psi(\theta)$ are provided in Appendix A.2. Condition (v) holds for any compact set Θ because $(y_{i1}, y_i', x_{i1}', \dots, x_{iT}')$ has finite fourth moment.

2. By Theorem 3.4 in Newey and McFadden (1994), in addition to conditions (i)-(v), to establish asymptotic normality it suffices to verify:

(vi) $G'\Omega^{-1}G$ is nonsingular.

(vii) $E[\sup_{\theta\in\Theta} \| {}_{\theta}g_i(\theta)\|_\infty^2]$ is finite.

Condition (vi) follows immediately from Assumption 5, while (vii) follows by construction of $g_i(\theta)$ (see eq. (8)) and the characterization provided in Appendix A.2.

□

A.3.3 Proof of Lemma Lemma 2

By Theorem 4.5 in Newey and McFadden (1994), conditions (i)–(vii) are sufficient to establish the consistency of the asymptotic variance estimator.

□

A.3.4 Proof of Theorem Theorem 2

Denote $\Omega^D = \Omega^D(\gamma_\circ, \beta_\circ') = E[g_i^D(\theta_\circ)g_i^D(\theta_\circ)']$ and consider the unfeasible estimator

$$\tilde{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, \bar{g}^D(\theta)'(\Omega^D)^{-1}\bar{g}^D(\theta).$$

The reason for considering such an estimator is that, by expression (12), it has the same asymptotic variance as $\hat{\theta}$; see Hall (2005, Sec. 3.7). After partitioning

$$\Omega^D = \begin{pmatrix} \Omega_{11}^D & \underset{(\tilde{h}'+T-2)\times(kT+3)}{\Omega_{12}^D} \\ \underset{(kT+3)\times(\tilde{h}'+T-2)}{\Omega_{12}^{D\prime}} & \underset{(kT+3)\times(kT+3)}{\Omega_{22}^D} \end{pmatrix},$$

$\tilde{\theta}$ can be computed by solving the following linear system of equations:

$$\begin{pmatrix} {}_{(\gamma,\beta')}\bar{g}_1^D(\gamma,\beta')' & {}_{(\gamma,\beta')}\bar{g}_2^D(\gamma,\beta';\theta_{\setminus\gamma\beta})' \\ {}_{\theta_{\setminus\gamma\beta}}\bar{g}_1^D(\gamma,\beta')' & {}_{\theta_{\setminus\gamma\beta}}\bar{g}_2^D(\gamma,\beta';\theta_{\setminus\gamma\beta})' \end{pmatrix} \begin{pmatrix} \Omega_{11}^D & \Omega_{12}^D \\ \Omega_{12}^{D\prime} & \Omega_{22}^D \end{pmatrix}^{-1} \begin{pmatrix} \bar{g}_1^D(\gamma,\beta') \\ \bar{g}_2^D(\gamma,\beta';\theta_{\setminus\gamma\beta}) \end{pmatrix}$$
$$= 0_{[k(T+1)+4]\times 1},$$

where ${}_{(\gamma,\beta')}\bar{g}_1^D(\gamma,\beta') = (1/N)\sum_{i=1}^N {}_{(\gamma,\beta')}g_{i,1}^D(\gamma,\beta')$,

$${}_{(\gamma,\beta')}g_{i,1}^D(\gamma,\beta') = \left(\frac{\partial g_{i,1}^D(\gamma,\beta')}{\partial\gamma} \quad \frac{\partial g_{i,1}^D(\gamma,\beta')}{\partial\beta'}\right),$$

and the remaining terms are defined in a similar manner. From the inverse formula for symmetric partitioned matrices (Theil 1983, eq. 3.2) and since $g_{i,1}^D(\gamma,\beta')$ does not depends on $\theta_{\setminus\gamma\beta}$, i.e. ${}_{\theta_{\setminus\gamma\beta}}\bar{g}_1^D(\gamma,\beta') = 0_{(\tilde{h}'+T-2)\times(kT+3)}$, our unfeasible estimator $(\tilde{\gamma},\tilde{\beta}')$ can be obtained by solving the (linear) system of equations

$${}_{(\gamma,\beta')}\bar{g}_1^D(\gamma,\beta')'(\Omega_{11}^D)^{-1}\bar{g}_1^D(\gamma,\beta') = 0_{(k+1)\times 1}$$

or, equivalently, by solving the following optimization problem:

$$(\tilde{\gamma},\tilde{\beta}') = \underset{(\gamma,\beta')}{\operatorname{argmin}}\, \bar{g}_1^D(\gamma,\beta')'(\Omega_{11}^D)^{-1}\bar{g}_1^D(\gamma,\beta'). \tag{17}$$

From these expressions, it follows that we can ignore the presence of $\theta_{\backslash\gamma\beta}$, as well as Ω_{12}^D and Ω_{22}^D, when computing $(\tilde{\gamma}, \tilde{\beta}')$.

Following the arguments in the Proof of Theorem 1.2, it can be shown that the asymptotic variance of $(\tilde{\gamma}, \tilde{\beta}')$ is given by $\left[G_1^{D\prime}(\Omega_{11}^D)^{-1}G_1^{D\prime}\right]^{-1}$, where

$$G_1^D = E[\ _{(\gamma,\beta')}g_{i,1}^D(\gamma_\circ, \beta'_\circ)].$$

Since $(\tilde{\gamma}, \tilde{\beta}')$ and $(\hat{\gamma}, \hat{\beta}')$ are asymptotically equivalent, it follows immediately that $\Sigma_{\gamma\beta} = \left[G_1^{D\prime}(\Omega_{11}^D)^{-1}G_1^{D\prime}\right]^{-1}$. Partition

$$\underset{(\tilde{h}'+T-2)\times(k+1)}{G_1^D} = \begin{pmatrix} \underset{\tilde{h}'\times(k+1)}{G^{AS}} \\ \underset{(T-2)\times(k+1)}{G_{1,2}^D} \end{pmatrix},$$

note that $G_{1,2}^D = E(\Delta\tilde{x}_i)$, and write

$$(\Omega_{11}^D)^{-1} = \begin{pmatrix} (\Omega^{AS})^{-1} + (\Omega^{AS})^{-1}\Omega_{11,12}^D \Upsilon \Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1} & -(\Omega^{AS})^{-1}\Omega_{11,12}^D\Upsilon \\ -\Upsilon\Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1} & \Upsilon \end{pmatrix}.$$

This inverse is obtained by applying eq. (3.2) of Theil (1983). We have that Υ is positive definite because $\Omega^D = D\Omega D'$ is positive definite (Assumption 4) and so is its inverse. Finally, it follows that

$$\begin{aligned}
\Sigma_{\gamma\beta}^{-1} &= \begin{pmatrix} G^{AS\prime} & G_{1,2}^{D\prime} \end{pmatrix} (\Omega_{11}^D)^{-1} \begin{pmatrix} G^{AS} \\ G_{1,2}^D \end{pmatrix} \\
&= G^{AS\prime}\left[(\Omega^{AS})^{-1} + (\Omega^{AS})^{-1}\Omega_{11,12}^D\Upsilon\Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1}\right]G^{AS} \\
&\quad - G_{1,2}^{D\prime}\Upsilon\Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1}G^{AS} - G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^D\Upsilon G_{1,2}^D \\
&\quad + G_{1,2}^{D\prime}\Upsilon\tilde{G}_{1,2}^D \\
&= (\Sigma_{\gamma\beta}^{AS})^{-1} + G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^D\Upsilon\Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1}G^{AS} \\
&\quad - G_{1,2}^{D\prime}\Upsilon\Omega_{11,12}^{D\prime}(\Omega^{AS})^{-1}G^{AS} - G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^D\Upsilon G_{1,2}^D \\
&\quad + G_{1,2}^{D\prime}\Upsilon\tilde{G}_{1,2}^D
\end{aligned}$$

and therefore

$$\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AS})^{-1} = [G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^D - G_{1,2}^{D\prime}]\Upsilon[G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^D - G_{1,2}^{D\prime}]'.$$

□

Notes

1 Although not reported, we also considered additional sample sizes and distributions as in Wu and Zhu (2012).
2 Theorem 2.6 in Newey and McFadden (1994) indeed requires weaker conditions than (i)–(v).

References

Ahn, S. C., and P. Schmidt. 1995. "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68 (1): 5–27.
Ahn, S. C., and P. Schmidt. 1997. "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation." *Journal of Econometrics* 76: 309–321.
Anderson, T. W., and C. Hsiao. 1981. "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association* 76: 598–606.
Anderson, T. W., and C. Hsiao. 1982. "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics* 18: 47–82.

Arellano, M., and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2): 277–297.
Arellano, M., and O. Bover. 1995. "Another Look at the Instrumental Variable Estimation of Error-Components Models." *Journal of Econometrics* 68 (1): 29–51.
Arias, O., M. Marchionni, and W. Sosa-Escudero. 2011. Sources of Income Persistence: Evidence from Rural El Salvador." *Journal of Income Distribution* 20: 3–28.
Blundell, S., and S. Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87 (1): 115–143.
Bun, M., and F. Windmeijer. 2010. "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models." *Econometrics Journal* 13: 95–126.
Chamberlain, G. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47 (1): 225–238.
Chamberlain, G. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18: 5–46.
Chamberlain, G. 1984. "Panel Data." In *Handbook of Econometrics*, edited by Z. Griliches and M. Intriligator, vol. 2, ch. 22, 1247–1318. North-Holland.
Crepon, B., and J. Mairesse. 2008. "The Chamberlain Approach to Panel Data: An Overview and Some Simulations." In *The Econometrics of Panel Data*, (3rd ed.), edited by L. Matyas and P. Sevestre, ch. 5, 113–183. Springer.
Hall, A. R. 2005. *Generalized Method of Moments*. Oxford University Press.
Harris, M. N., and L. Matyas. 2004. "A Comparative Analysis of Different IV and GMM Estimators of Dynamic Panel Data Models." *International Statistical Review* 72: 397–408.
Harris, M., L. Matyas, and P. Sevestre. 2008. "Dynamic Models for Short Panels." In *The Econometrics of Panel Data* (3rd ed.), edited by L. Matyas and P. Sevestre, ch. 8, 249–278. Springer.
Hausman, J. N., and M. Pinkovskiy. 2017. *Estimating Dynamic Panel Models: Backing out the Nickell Bias*. Federal Reserve Bank of New York Staff Reports 824.
Holtz-Eakin, D., W. K. Newey, and H. S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56 (6): 1371–1395.
Judson, R., and A. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economics Letters* 65: 9–15.
Lillard, L. A., and R. J. Willis. 1978. "Dynamic Aspects of Earning Mobility." *Econometrica* 46 (5): 985–1012.
Newey, W. K., and D. L. McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, edited by R. F. Engle and D. L. McFadden, vol. 4, ch. 36, 2111–2245. North-Holland.
Nickel, S. 1981. "Bias in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417–1426.
Robertson, D., and V. Sarafidis. 2015. "IV Estimation of Panel with Factor Residuals." *Journal of Econometrics* 185 (2): 526–541.
Roodman, D. 2009. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics* 71 (1): 135–158.
Sasaki, Y., and Y. Xin. 2017. "Unequal Spacing in Dynamic Panel Data: Identification and Estimation." *Journal of Econometrics* 196 (2): 320–330.
Stock, J., M. Yogo, and J. Wright. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20: 518–529.
Theil, H. 1983. "Linear Algebra and Matrix Methods in Econometrics." In *Handbook of Econometrics*, edited by Z. Griliches and M. D. Intriligator, vol. 1, ch. 1, 3–65. North-Holland.
Wu, J., and L. Zhu. 2012. "Estimation of and Testing for Random Effects in Dynamic Panel Data Models." *Test* 21 (3): 477–497.
Yamagata, T. 2008. "A Joint Serial Correlation Test for Linear Panel Data Models." *Journal of Econometrics* 146 (1): 135–145.
Zincenko, F., W. Sosa-Escudero, and G. Montes-Rojas. 2014. "Robust Tests for Time-Invariant Individual Heterogeneity versus Dynamic State Dependence." *Empirical Economics* 47 (4): 1365–1387.