



***WORKING PAPER SERIES***

17/003

Alternative moment conditions and an efficient GMM estimator for dynamic panel  
data models

Gabriel Montes-Rojas, Walter Sosa-Escudero, and Federico Zincenko

January, 2017

# Alternative moment conditions and an efficient GMM estimator for dynamic panel data models

Gabriel Montes-Rojas\*      Walter Sosa-Escudero<sup>†</sup>  
Federico Zincenko<sup>‡</sup>

January 30, 2017

## Abstract

This paper proposes a set of moment conditions for the estimation of linear dynamic panel data models. In the spirit of Chamberlain's (1982, 1984) approach, these conditions arise from parameterizing the relationship between covariates and unobserved time invariant effects. A GMM framework is used to derive an optimal estimator, with no efficiency loss compared to classic alternatives like Arellano and Bond (1991) and Ahn and Schmidt (1995, 1997). Still, Monte Carlo results suggest that the new procedure performs better than these alternatives when covariates are non-stationary. The framework also leads to a very simple test for unobserved effects.

*JEL Classification:* C12, C23.

*Keywords:* Asymptotic efficiency, dynamic panel, GMM estimation, individual effects, short panel.

---

\*E-mail: gmontesrojas@udesa.edu.ar. CONICET-Universidad de San Andrés

<sup>†</sup>E-mail: wsosa@udesa.edu.ar. Universidad de San Andrés-CONICET

<sup>‡</sup>Corresponding author. E-mail: zincenko@pitt.edu. Department of Economics, University of Pittsburgh.

We thank Chengying Luo for excellent research assistance. Financial support from the Central Research Development Fund (University of Pittsburgh) and Programa RAICES (CONICET, Argentina) is gratefully acknowledge.

# 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. Empirical researchers value its simple yet flexible structure that can help explore persistent effects in the form of time-invariant unobserved heterogeneity and/or lagged dependent variables. The theoretical literature has focused on the difficulties in dealing with the endogeneity problems that arise by the simultaneous inclusion of both sources of persistence, that render standard methods (OLS and within estimators) biased and inconsistent, as highlighted by Nickel (1981).

The applied literature has largely favored instrumental variables (IV) and moment conditions estimation strategies, 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. Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Ahn and Schmidt (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. 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 (2009) in their survey.

In spite of their popularity, IV/GMM methods are not free from limitations. Some of them relate to their very nature, in the sense that the initial evolution of the literature has favored exploiting the dynamic structure of the model to progressively derive further moment conditions, that eventually lead to increased asymptotic efficiency, always within the IV/GMM framework. Consequently, many concerns that affect IV strategies translate to the case of dynamic panel models, including the problem of weak instruments (Bun and Windmeijer, 2010, and Stock, Wright, and Togo, 2002) and the related issue of moment multiplicity (Roodman, 2009). The Monte Carlo literature on the subject has clearly reflected the practical relevance of these theoretical concerns, see Judson and Owen (1999) and Harris and Matyas (2004).

In this paper we explore an alternative route to deal with the incidental parameters introduced by the presence of unobserved time invariant effects. In the spirit of Chamberlain's (1982, 1984) classic work, we parameterize the possible relationships between regressors and these effects. The moment conditions implied by these parametrizations are 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. Even though the use of parametrizations to deal with lagged dependent variables and fixed effects has already been implemented in the literature (see, e.g., Crepon and Mairesse, 2004), our framework provides an explicit link between the new strategy and existing methods. In particular, our paper shows that classic 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. Nevertheless, a Monte Carlo experiment suggests that though fairly comparable to Arellano and Bond (1991) and Ahn and Schmidt (1995) in the case of stationary covariates, the new estimator performs better in the non-stationary case.

Parametric 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. 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). Finally, 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 heterogeneities.

The paper is structured as follows. Section 2 presents the model. Section 3 introduces the parameterization together with the corresponding moment conditions in levels and establishes the key identification condition that leads to the proposed estimator, 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

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

where  $i = 1, \dots, N$  and  $t = 2, \dots, T$ . In this model,  $y_{it}$  is the dependent variable,  $x_{it}$  is a  $k \times 1$  vector of regressors,  $\mu_i$  is the individual effect, and  $\varepsilon_{it}$  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  $\mu_i$  is denoted by  $\sigma_{\mu_{\circ}}^2$ . We will assume that  $T \geq 3$ .

The coefficient  $\gamma_{\circ}$  measures the degree of the state dependence or pure dynamic persistence. The factor  $\mu_i$  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  $N \rightarrow \infty$  and  $T$  is fixed.

Assume that the researcher observes a random sample  $\{[(y_{i1}, y'_i)', (x_{i1}, x_i)'] : i = 1, \dots, N\}$ , where  $y_i = (y_{i2}, \dots, y_{iT})'$  is a  $(T - 1) \times 1$  random vector and  $x_i = (x_{i2}, \dots, x_{iT})$  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 tests will be derived assuming that  $N$  grows to infinity and  $T$  is fixed, which is standard in microeconometrics.

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

**Assumption 1.**  $\{(y_{i1}, x'_{i1}, \dots, x'_{iT}, \mu_i, \varepsilon'_i) : i = 1, \dots, N\}$  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. The covariance between  $x_{it}$  and  $\mu_i$  is allowed to

vary across  $t$ , but not across  $i$ . Assumption 1 implies that  $\{(y_{i1}, y'_i, x'_{i1}, \dots, x'_{iT}) : i = 1, \dots, N\}$  are i.i.d. random vectors with finite fourth moments.

**Assumption 2.** *The following conditions hold.*

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

The first one imposes the usual normalizing restriction  $E(\mu_i) = 0$ , since the model contains a common intercept. The second requires that the errors  $\varepsilon_{it}$  be uncorrelated across  $t = 1, \dots, T$ . 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  $\varepsilon_{it}$  can also be relaxed as long as  $E(\varepsilon_{it})$  is constant across periods; in such case,  $E(\varepsilon_{it})$  will be ‘captured’ by  $\alpha_o$ . The variance of  $\varepsilon_{it}$  is allowed to vary across  $t$ . The third part of Assumption 2 together with  $E(\varepsilon_{it}\mu_i) = 0$  imply 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  $t = 2, \dots, T$  and  $1 \leq s < t$ . Observe that the individual effects  $\mu_i$  are allowed to be correlated with  $(x_{i1}, x_i)'$ . No restrictions are imposed on the relationship between  $y_{i1}$  and  $\mu_i$ .

Assumptions 1-2 are equivalent to Ahn and Schmidt (1995)’s (SA.1 - SA.3) after adding the sequential moment conditions  $E(x_{is}\varepsilon_{it}) = 0$ , for  $s \leq t$ .

### 3 Moment conditions in levels

This section presents the moments conditions derived from Assumptions 1-2 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

$$\mathbf{Z}_{Yi} = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & \dots & \dots & \dots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-1} \end{pmatrix},$$

and

$$\mathbf{Z}_{Xi} = \begin{pmatrix} x'_{i1} & x'_{i2} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & \dots & \dots & \dots & 0_{1 \times k} \\ 0_{1 \times k} & 0_{1 \times k} & x'_{i1} & x'_{i2} & x'_{i3} & 0_{1 \times k} & \dots & \dots & \dots & 0_{1 \times k} \\ 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & x'_{i1} & \dots & \dots & \dots & 0_{1 \times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & \dots & x'_{i1} & \dots & x'_{iT} \end{pmatrix}.$$

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  $\varepsilon_i$  satisfy the moment conditions

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

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

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

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

The matrix  $E(\mathbf{Z}'_i \mu_i)$  contains the covariances between  $\mu_i$  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. The standard dynamic panel data literature applies first-differences to eliminate  $\mu_i$ , which is correlated with  $(y_{i,t-1}, x'_{it})$  invalidating the previously proposed instruments. In our case, we consider these correlations as a free parameters. This approach 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.

As an illustration and to clarify the notation, consider the following example in which  $k = 1$  and  $T = 3$ .

**Example** ( $k = 1$  &  $T = 3$ ). *There are  $h = 10$  moment conditions:  $E(I_2 \varepsilon_i) = E(u_i) - E(\mu_i) \iota_2 = 0_{2 \times 1}$ ,*

$$E(\mathbf{Z}'_{Y_i} \varepsilon_i) = E \begin{pmatrix} y_{i1} \varepsilon_{i2} \\ y_{i1} \varepsilon_{i3} \\ y_{i2} \varepsilon_{i3} \end{pmatrix} = E \begin{pmatrix} y_{i1} u_{i2} \\ y_{i1} u_{i3} \\ y_{i2} u_{i3} \end{pmatrix} - E \begin{pmatrix} y_{i1} \mu_i \\ y_{i1} \mu_i \\ y_{i2} \mu_i \end{pmatrix} = 0_{3 \times 1},$$

and

$$E(\mathbf{Z}'_{X_i} \varepsilon_i) = E \begin{pmatrix} x_{i1} \varepsilon_{i2} \\ x_{i2} \varepsilon_{i2} \\ x_{i1} \varepsilon_{i3} \\ x_{i2} \varepsilon_{i3} \\ x_{i3} \varepsilon_{i3} \end{pmatrix} = E \begin{pmatrix} x_{i1} u_{i2} \\ x_{i2} u_{i2} \\ x_{i1} u_{i3} \\ x_{i2} u_{i3} \\ x_{i3} u_{i3} \end{pmatrix} - E \begin{pmatrix} x_{i1} \mu_i \\ x_{i2} \mu_i \\ x_{i1} \mu_i \\ x_{i2} \mu_i \\ x_{i3} \mu_i \end{pmatrix} = 0_{5 \times 1}.$$

Note that  $h_y = 3$  and  $h_x = 5$ .

Let  $\tau_{1o}^y = E(y_{i1}\mu_i)$  and  $\tau_{to}^x = E(x_{it}\mu_i)$  for  $t = 1, \dots, T$ . We consider these covariances as parameters of our model. Let  $\tau_o = (\tau_{1o}^y, \tau_{1o}^{x'}, \dots, \tau_{To}^{x'})'$  be a  $(kT + 1) \times 1$  vector containing these covariances and  $\theta_o = (\alpha_o, \gamma_o, \beta'_o, \sigma_{\mu_o}^2, \tau'_o)$  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:

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

It can be shown by induction that

$$E(y_{it}\mu_i) = \gamma_\circ^{t-1} \tau_{1\circ}^y + \sum_{j=2}^t \gamma_\circ^{t-j} \tau_{j\circ}^{x'} \beta_\circ + \frac{\gamma_\circ^{t-1} - 1}{\gamma_\circ - 1} \sigma_{\mu_\circ}^2, \quad (4)$$

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

In light of previous discussion, we now parameterize the covariances  $E(\mathbf{Z}'_i \mu_i) \nu_{T-1}$  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  $\Theta$  be a compact subset such that  $\theta_\circ \in \text{interior}(\Theta)$ .<sup>1</sup> In view of expressions (2)-(4), we construct the function  $\Psi : \Theta \rightarrow \mathbb{R}^{h \times 1}$  as

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

where

$$\begin{aligned} \Psi_Y(\theta) &= (\tau_1^y, \tau_1^y, \psi_2(\theta), \tau_1^y, \psi_2(\theta), \psi_3(\theta), \tau_1^y, \dots, \tau_1^y, \dots, \psi_{T-1}(\theta))', \\ & \quad h_y \times 1 \\ \Psi_X(\theta) &= (\tau_1^{x'}, \tau_2^{x'}, \tau_1^{x'}, \tau_2^{x'}, \tau_3^{x'}, \tau_1^{x'}, \dots, \tau_1^{x'}, \dots, \tau_T^{x'})', \\ & \quad h_x \times 1 \end{aligned}$$

and

$$\psi_t(\theta) = \gamma^{t-1} \tau_1^y + \sum_{l=2}^t \gamma^{t-l} \tau_l^{x'} \beta + \frac{\gamma^{t-1} - 1}{\gamma - 1} \sigma_\mu^2, \quad (6)$$

for  $2 \leq t \leq T-1$  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  $\tau_1^y$  occupies the positions  $\{[t(t-1)/2] + 1 : t = 1, \dots, T-1\}$  of  $\Psi_Y(\theta)$ ,  $\psi_2(\theta)$  occupies

---

<sup>1</sup>Although  $\sigma_{\mu_\circ}^2 \geq 0$ ,  $\sigma_\mu^2$  may be allowed to take negative values as the identification result (Lemma 1 below) does not rely on the restriction  $\sigma_\mu^2 \geq 0$ .

positions  $\{[t(t-1)/2]+2 : t = 2, \dots, T-1\}$  of  $\Psi_Y(\theta)$ , and in general for  $2 \leq j \leq T-1$ ,  $\psi_j(\theta)$  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  $\Psi_X^{(j_1:j_2)}(\theta) = \tau_t^x$ .

**Example** ( $k = 1$  &  $T = 3$ , cont.). *In this case,  $\Psi(\theta)$  becomes*

$$\Psi(\theta) = \begin{pmatrix} 0_{2 \times 1} \\ \Psi_Y(\theta) \\ \Psi_X(\theta) \end{pmatrix}$$

with

$$\Psi_Y(\theta) = \begin{pmatrix} \tau_1^y \\ \tau_1^y \\ \psi_2(\theta) \end{pmatrix}, \quad \Psi_X(\theta) = \begin{pmatrix} \tau_1^x \\ \tau_2^x \\ \tau_1^x \\ \tau_2^x \\ \tau_3^x \end{pmatrix},$$

and  $\psi_2(\theta) = \gamma\tau_1^y + \tau_2^x\beta + \sigma_\mu^2$ .

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

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

and

$$E(\mathbf{Z}'_{Xi}\mu_i) = \begin{pmatrix} \tau_{1o}^{x'} & \tau_{2o}^{x'} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & \dots & \dots & \dots & 0_{1 \times k} \\ 0_{1 \times k} & 0_{1 \times k} & \tau_{1o}^{x'} & \tau_{2o}^{x'} & \tau_{3o}^{x'} & 0_{1 \times k} & \dots & \dots & \dots & 0_{1 \times k} \\ 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & \tau_{1o}^{x'} & \dots & \dots & \dots & 0_{1 \times k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & 0_{1 \times k} & \dots & \tau_{1o}^{x'} & \dots & \tau_{T_o}^{x'} \end{pmatrix},$$

respectively, which implies

$$E(\mathbf{Z}'_i u_i) - \Psi(\theta_o) = 0_{h \times 1}. \quad (7)$$

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

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

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  $E[g_i(\theta_o)] = 0_{h \times 1}$  since  $u_i = y_i - \alpha_o \iota_{T-1} - \gamma_o y_{i,-1} - x'_i \beta_o$  and eq. (7). After imposing a standard rank condition, we will show that  $\theta_o$  is the unique solution to the (nonlinear) system of equations  $E[g_i(\theta)] = 0_{h \times 1}$  with  $\theta \in \Theta$ .

Write  $\tilde{\mathbf{Z}}_i = (\tilde{\mathbf{Z}}_{Y_i}, \tilde{\mathbf{Z}}_{X_i})_{(T-2) \times \tilde{h}}$ , where  $\tilde{\mathbf{Z}}_{Y_i}$  and  $\tilde{\mathbf{Z}}_{X_i}$  are constructed by removing the last of row of  $\mathbf{Z}_{Y_i}$  and  $\mathbf{Z}_{X_i}$ , respectively, and  $\tilde{h} = [k(T+1) + T - 1](T-2)/2$ . Denote further  $\Delta y_{i,-1} = (y_{i2} - y_{i1}, \dots, y_{iT-1} - y_{iT-2})'$ ,  $\Delta \tilde{x}_i = (x_{i3} - x_{i2}, \dots, x_{iT} - x_{iT-1})'$ , 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.**  $E(\tilde{\mathbf{Z}}'_i \Delta \mathbf{x}_i)$  has rank  $k+1$ .

This assumption is standard in the dynamic panel literature. In particular, it coincides with Assumption 3 in Yamagata (2008) and it rules out perfect multicollinearity among regressors. The next Lemma characterizes the true parameters vector,  $\theta_o$ , as the unique solution of a system of moment conditions based on  $g_i(\cdot)$ .

**Lemma 1** (Identification). *Under Assumptions 1-3,  $\theta_o \in \Theta$  is the unique solution of  $E[g_i(\theta)] = 0_{h \times 1}$ .*

*Proof.* From eqs. (7)-(8)  $\theta_o$  is solution of  $E[g_i(\theta)] = 0_{h \times 1}$ . Next we show that  $\theta_o$  is indeed the unique solution. Suppose that  $\tilde{\theta} = (\tilde{\alpha}, \tilde{\gamma}, \tilde{\beta}', \tilde{\sigma}_\mu^2, \tilde{\tau}')$  satisfies  $E[g_i(\tilde{\theta})] = 0_{h \times 1}$ . We prove first that  $(\tilde{\gamma}, \tilde{\beta}') = (\gamma_o, \beta'_o)$ . It can be easily shown that there exists a non stochastic  $\tilde{h} \times h$  matrix  $D^{AB}$  such that

$$D^{AB} g_i(\theta) = \tilde{\mathbf{Z}}_i (\Delta y_i - \gamma \Delta y_{i,-1} - \Delta \mathbf{x}_i \beta) = \tilde{\mathbf{Z}}_i \Delta y_i - \tilde{\mathbf{Z}}_i \Delta \mathbf{x}_i \begin{pmatrix} \gamma \\ \beta \end{pmatrix} \quad (9)$$

for all  $\theta \in \Theta$ . For instance, when  $k = 1$  and  $T = 3$  ( $\tilde{h} = 3$ ), this matrix becomes

$$\mathbf{D}_{3 \times 10}^{AB} = \begin{pmatrix} 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}. \quad (10)$$

Observe that  $E[\mathbf{D}^{AB}g_i(\tilde{\theta})] = \mathbf{D}^{AB}E[g_i(\tilde{\theta})] = \mathbf{0}_{\tilde{h} \times 1}$ , so equation (9) yields the Arellano and Bond (1991) system of linear equations:

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

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  $(\tilde{\gamma}, \tilde{\beta}') = (\gamma_\circ, \beta'_\circ)$ . Using the first equation of the system  $E[g_i(\tilde{\theta})] = \mathbf{0}_{h \times 1}$ , 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  $T$ -th equation of  $E[g_i(\tilde{\theta})] = \mathbf{0}$ , it follows that

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

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  $E[g_i(\tilde{\theta})] = \mathbf{0}_{h \times 1}$  related to expression (3), we obtain

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

and therefore

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

□

## 4 Optimal GMM estimator

This section proposes a GMM based estimator for  $\theta_o$  and derives its asymptotic properties. We begin by constructing an estimator for  $\Omega \equiv E[g_i(\theta_o)g_i(\theta_o)']$ , whose inverse will be the optimal weighting matrix in the GMM context. Since  $g_i(\theta_o) = \mathbf{Z}'_i u_i - E(\mathbf{Z}'_i u_i)$  by equations (7)-(8), Assumption 1 (finite fourth moments) implies that  $\Omega = E\{[\mathbf{Z}'_i u_i - E(\mathbf{Z}'_i u_i)][\mathbf{Z}'_i u_i - E(\mathbf{Z}'_i u_i)]'\}$  is finite, symmetric, and positive semidefinite.

To construct a consistent estimator of  $\Omega$ , it suffices to build a consistent (first-step) estimator of  $\theta_o$ . We suggest using  $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\beta}', \hat{\sigma}_\mu^2, \hat{\tau}')'$  where  $(\hat{\gamma}, \hat{\beta}')$  is the first-step Arellano-Bond estimator of  $(\gamma_o, \beta'_o)$ ,

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

$\hat{\tau} = (\hat{\tau}_1^y, \hat{\tau}_1^{x'}, \dots, \hat{\tau}_T^{x'})'$ , and  $\dot{u}_{it} = y_{it} - \hat{\alpha} - \hat{\gamma} y_{i,t-1} - x'_{it} \hat{\beta}$ . It follows that  $\hat{\theta} \xrightarrow{P} \theta_o$ , where  $\xrightarrow{P}$  denotes convergence in probability. We highlight that the asymptotic results of the paper will not be affected if  $\hat{\theta}$  is replaced by another consistent estimator of  $\theta_o$ . The natural estimator of  $\Omega$  is then given by

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

and satisfies  $\hat{\Omega} \xrightarrow{P} \Omega$ ; see Lemma 4.3 of Newey and McFadden (1994).

The next assumption states that  $\Omega$  is positive definite and assures the existence of an optimal weighting matrix.

**Assumption 4.**  $\Omega = E\{[\mathbf{Z}'_i u_i - E(\mathbf{Z}'_i u_i)][\mathbf{Z}'_i u_i - E(\mathbf{Z}'_i u_i)]'\}$  is positive definite.

The optimal GMM estimator of  $\theta_o$  is defined as

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

where  $\bar{g}(\theta) = (1/N) \sum_{i=1}^N g_i(\theta)$  and  $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\beta}', \hat{\sigma}_\mu^2, \hat{\tau}')$ . This minimization problem can be solved by standard numerical optimization methods –such as Newton-Raphson’s algorithm– and we suggest using  $\hat{\theta}$  as initial value. From Theorem 2.6 in Newey and McFadden (1994), we have  $\hat{\theta} \xrightarrow{p} \theta_o$ . Denote the first-order partial derivatives of  $g_i(\theta)$  by

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

and write  $G = E[\nabla_{\theta} g_i(\theta_o)]$ . We remark that computational differentiation is not required. A closed-form expression of the Jacobian matrix of  $\Psi(\theta)$  (denoted by  $\nabla_{\theta} \Psi(\theta)$ ) is provided in the Appendix.

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

**Assumption 5.**  $G' \Omega^{-1} G$  is nonsingular.

In the rest of this paper, we suppose that Assumptions 1-5 hold. Let  $\xrightarrow{d}$  denote convergence in distribution and  $N(\cdot, \cdot)$  be a multivariate normal distribution. Define  $B = G' \Omega^{-1} G$  and  $\Sigma = B^{-1}$ . The asymptotic distribution of our estimator is given by

$$\sqrt{N}(\hat{\theta} - \theta_o) \xrightarrow{D} N(0, \Sigma).$$

This result follows by Theorem 3.4 in Newey and McFadden (1994). The asymptotic variance  $\Sigma$  can be consistently estimated by  $\hat{\Sigma} = \hat{B}^{-1}$ , where  $\hat{B} = \hat{G}' \hat{\Omega}^{-1} \hat{G}$ ,

$$\hat{G} = \frac{1}{N} \sum_{i=1}^N \nabla_{\theta} g_i(\hat{\theta}),$$

and  $\hat{\Omega} = (1/N) \sum_{i=1}^N g_i(\hat{\theta}) g_i(\hat{\theta})'$ ; see Lemma 4.3 and Theorem 4.5 in Newey and McFadden (1994).

Among other hypotheses of interests, the result 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  $H_0 : (\sigma_{\mu_0}^2, \tau_0') = 0_{1 \times (kT+2)}$ , we have

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

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

## 5 Comparison with existing GMM estimators

This section relates our estimator to the ones proposed by Arellano and Bond (1991) and by Ahn and Schmidt (1995).

### 5.1 Comparison with Arellano and Bond (1991)

Without loss of generality and to simplify the discussion, in this subsection, we suppose  $k = 1$  and  $T = 3$ . In their seminal paper, Arellano and Bond (1991) suggest using the moment conditions

$$E \begin{pmatrix} y_{i1} \Delta u_{i3} \\ x_{i1} \Delta u_{i3} \\ x_{i2} \Delta u_{i3} \end{pmatrix} = 0_{3 \times 1}, \quad (11)$$

with  $\Delta u_{i3} = u_{i3} - u_{i2}$ . Exploiting these conditions, Arellano and Bond (1991) propose estimating  $(\gamma_0, \beta_0)$  by GMM using the function

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

The optimal Arellano-Bond GMM estimator is

$$(\hat{\gamma}^{AB}, \hat{\beta}^{AB}) = \underset{(\gamma, \beta)}{\operatorname{argmin}} \bar{g}^{AB}(\gamma, \beta)' W^{AB} \bar{g}^{AB}(\gamma, \beta), \quad (12)$$

where  $\bar{g}^{AB}(\gamma, \beta) = (1/N) \sum_{i=1}^N g_i^{AB}(\gamma, \beta)$ , and  $W^{AB}$  is optimal weighting matrix associated with  $g_i^{AB}(\cdot)$ :

$$W^{AB} = \{E [g_i^{AB}(\gamma_o, \beta_o) g_i^{AB}(\gamma_o, \beta_o)']\}^{-1}.$$

The specific form of the weighting matrix  $W^{AB}$  depends on assumptions about  $\varepsilon_i$ —such as homoskedasticity—. Moreover,  $W^{AB}$  can be replaced by any consistent estimator as the asymptotic variance is not affected.

Before proceeding, we observe that  $D^{AB}\Psi(\theta) = 0$  and  $g_i^{AB}(\gamma, \beta) = D^{AB}g_i(\theta)$  for all  $\theta = (\alpha, \gamma, \beta, \sigma_\mu^2, \tau')$ , where  $D^{AB}$  has been defined in eq. (10) for the case  $(k, T) = (1, 3)$ .<sup>2</sup> The linear transformation  $D^{AB}$  removes not only the individual effects, but also the function  $\Psi(\cdot)$  from  $g_i(\cdot)$ . Since  $W^{AB} = (D^{AB}\Omega D^{AB'})^{-1}$ , the objective function of minimization problem (12) can be written as

$$\bar{g}^{AB}(\gamma, \beta)' W^{AB} \bar{g}^{AB}(\gamma, \beta) = \bar{g}(\theta)' \tilde{W}^{AB} \bar{g}(\theta),$$

with  $\tilde{W}^{AB} = D^{AB'}(D^{AB}\Omega D^{AB'})^{-1}D^{AB}$ . This implies that the Arellano-Bond estimator can be obtained by minimizing the function  $\bar{g}(\cdot)' \tilde{W}^{AB} \bar{g}(\cdot)$ . Since the optimal weighting matrix associated with  $\bar{g}(\cdot)$  is  $\Omega^{-1}$ , our estimator  $(\hat{\gamma}, \hat{\beta})$  is at least as efficient as the Arellano-Bond estimator.

We highlight that if we ignore the first  $T - 1$  moments conditions in equation (2) related to  $E(\varepsilon_i) = 0_{(T-1) \times 1}$ , our estimator of  $(\gamma_o, \beta_o)$  coincides with that of Arellano-Bond when  $T = 3$ . Since we are indeed including these additional conditions and using the optimal weighting matrix, our estimator may be even more efficient. In particular, this occurs when the unconditional mean of  $(y_{it}, x_{it})$  varies across time; see Section 6.

## 5.2 Comparison with Ahn and Schmidt (1995)

In this subsection, we assume  $k = 0$  and  $T = 4$  to simplify the exposition.<sup>3</sup> Model (1) then becomes  $y_{it} = \alpha_o + \gamma_o y_{i,t-1} + u_{it}$  with  $u_{it} = \mu_i + \varepsilon_{it}$  and the (true) parame-

<sup>2</sup>We remark that  $\tilde{W}^{AB}$  does not depend on  $(\gamma, \beta)$ .

<sup>3</sup>When  $T = 3$ , Ahn and Schmidt (1995)'s moment conditions coincide with the ones proposed by Arellano and Bond (1991).



ters are  $(\alpha_o, \gamma_o, \sigma_{\mu_o}^2, \tau_{1_o}^y)$ . Ahn and Schmidt (1995, 1997) focus on estimating  $\gamma_o$  and suggest exploiting supplementary orthogonality conditions derived from the standard assumptions. Specifically, they consider

$$E \begin{pmatrix} y_{i1} \Delta u_{i3} \\ y_{i1} \Delta u_{i4} \\ y_{i2} \Delta u_{i4} \\ u_{i4} \Delta u_{i3} \end{pmatrix} = 0_{4 \times 1}; \quad (13)$$

see eqs. (4a)-(4b) in Ahn and Schmidt (1997) under ‘Case B’. The first 3 conditions are due to Arellano and Bond (1991), while the last one was added by Ahn and Schmidt (1995) to exploit the absence of serial correlation in  $\varepsilon_{it}$ . In this context, the optimal Ahn-Schmidt GMM estimator is given by

$$\hat{\gamma}^{AS} = \underset{\gamma}{\operatorname{argmin}} \bar{g}^{AS}(\gamma)' W^{AS} \bar{g}^{AS}(\gamma), \quad (14)$$

where  $\bar{g}^{AS}(\gamma) = (1/N) \sum_{i=1}^N g_i^{AS}(\gamma)$ ,

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

where  $W^{AS}$  is the optimal weighting matrix associated with  $g_i^{AS}(\cdot)$ , i.e.,  $W^{AS} = \{E[g_i^{AS}(\gamma_o)g_i^{AS}(\gamma_o)']\}^{-1}$ .

To compare  $\hat{\gamma}^{AS}$  with our estimator, define the matrix

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

for  $|\gamma| < 1$ .<sup>4</sup> Following similar arguments to the ones in previous subsection, we observe that  $D^{AS}(\gamma)\Psi(\alpha, \gamma, \sigma_{\mu}^2, \tau_1^y) = 0$  and  $g_i^{AS}(\gamma) = D^{AS}(\gamma)g_i(\alpha, \gamma, \sigma_{\mu}^2, \tau_1^y)$  for every  $(\alpha, \gamma, \sigma_{\mu}^2, \tau_1^y)$ . Since  $W^{AS} = [D^{AS}(\gamma_o)\Omega D^{AS}(\gamma_o)']^{-1}$ , the objective function of minimization problem (14) can be written as

$$\bar{g}^{AS}(\gamma)' W^{AS} \bar{g}^{AS}(\gamma) = \bar{g}(\alpha, \gamma, \sigma_{\mu}^2, \tau_1^y)' \tilde{W}^{AS}(\gamma) \bar{g}(\alpha, \gamma, \sigma_{\mu}^2, \tau_1^y)$$

---

<sup>4</sup>This matrix can be compared with the transformation of eq. (A.7) in Ahn and Schmidt (1995).

with  $\tilde{W}^{AS}(\gamma) = D^{AS}(\gamma)'[D^{AS}(\gamma_0)\Omega D^{AS}(\gamma_0)']^{-1}D^{AS}(\gamma)$ .<sup>5</sup> Then, Ahn-Schmidt estimator can be obtained by minimizing the objective function  $\bar{g}(\gamma, \cdot)'\tilde{W}^{AS}(\gamma)\bar{g}(\gamma, \cdot)$  with respect to  $\gamma$ . As discussed in the previous subsection, the optimal weighting matrix for  $\bar{g}(\cdot)$  is  $\Omega^{-1}$ , so our estimator  $\hat{\gamma}$  is at least as efficient as  $\hat{\gamma}^{AS}$ . Again, since we are exploiting additional moments conditions related to  $E(\varepsilon_i) = 0_{(T-1)\times 1}$ , our estimator may be even more efficient; e.g., when the mean of  $(y_{it}, x'_{it})$  varies across time.

## 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

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

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

We let  $\mu_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\mu^2)$ ,  $\epsilon_{it} \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$ ,  $v_{it} \stackrel{i.i.d.}{\sim} N(0, \sigma_v^2)$ . We fix  $\alpha = 0$ ,  $\beta = 1$ ,  $\sigma_\mu^2 = 53/108$ ,  $\sigma_\epsilon^2 = 55/20$  and  $\sigma_v^2 = 1$  as in Yamagata (2008, p. 141) and different parameter values specifications:  $\gamma \in \{0, 0.25, 0.5, 0.75\}$ ,  $\rho_\tau \in \{0, 0.25\}$ . The sample sizes considered are  $N \in \{200, 300\}$  and  $T \in \{4, 8\}$ .<sup>6</sup> 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  $(\alpha, \gamma, \beta)$  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

---

<sup>5</sup>We remark that  $\tilde{W}^{AS}(\gamma)$  depends on  $\gamma$ ; however, this does not affect the asymptotic distribution as it depends on  $\tilde{W}^{AS}(\gamma_0)$ ; see e.g. Hall (2005, sec. 3.7).

<sup>6</sup>Although not reported, we also considered additional sample sizes and distributions as in Wu and Zhu (2012).

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. Tables 1 and 2 report bias and RMSE for estimating  $\gamma$  and  $\beta$ , respectively. For this case the three procedures under comparison perform similarly. 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 Tables 1 and 2. The top graph corresponds to the estimation of  $\gamma$  (Table 1) and the bottom graph to the estimation of  $\beta$  (Table 2). The graph shows that the three estimators behave similarly for this case.

[ INSERT TABLES 1 AND 2 HERE ]

[ INSERT FIGURE 1 HERE ]

Interestingly when  $X_t$  is non-stationary (Tables 3 and 4 for  $\gamma$  and  $\beta$ , respectively), our proposed estimator systematically over-performs both AS and AB in terms of RMSE and bias for the ‘small T’ case ( $T = 4$ ). For example, when  $\gamma = 0.25$  and  $\rho_\tau = 0.25$ , the RMSE of our procedure when estimating  $\beta$  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 Tables 3 and 4 graphically, organized as in Figure 1.

[ INSERT TABLES 3 AND 4 HERE ]

[ INSERT FIGURE 2 HERE ]

As a by-product, the proposed set of moment conditions leads to a simple procedure to estimate  $\sigma_\mu^2$  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  $\tau$  parameters are themselves 0. The test would proceed by directly estimating  $\sigma_\mu^2$  and then testing down the corresponding null through a simple Wald-type test. Monte Carlo results for bias and RMSE for the estimation of  $\sigma_\mu^2$  are reported in Tables 5 and 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 Wu and Zhu (2012), Table 1, case (i), p.486, results for  $\gamma_2^\mu$ , simulations for  $N = 200, T = 8$  and  $N = 300, T = 8$ . This corresponds to  $\rho_\mu = 0$  and  $\gamma = 0.5$  in our experiments, for  $X_t$  stationary.

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  $\gamma$  increases, pointing out that, as in Zinchenko, Montes-Rojas, and Sosa-Escudero (2014), dynamic persistence and unobserved heterogeneity are confounding factors. In fact, our model can be seen as an extension of Zinchenko et al. (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, 1997), implying no efficiency loss.

Though mostly of theoretical and modelling interest, Monte Carlo results suggest that the new estimator performs better (in 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 effect. The derived procedure performs better than available alternatives like Wu and Zhu (2012).

## Appendix: Closed-form expression for $\nabla_{\theta}\Psi(\theta)$

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

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

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

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

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

where  $\nabla_{\gamma}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\gamma$  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  $2 \leq j \leq T-1$ .

Second, note that

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

Then,

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

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

The rest of this subsection considers  $\nabla_{\tau}\Psi(\theta)$ . We write

$$\nabla_{\tau}\Psi(\theta) = \begin{pmatrix} 0_{(T-1)\times 1} & 0_{(T-1)\times k} & \dots & 0_{(T-1)\times k} & \dots & 0_{(T-1)\times k} \\ \nabla_{\tau_1^y}\Psi_Y(\theta) & \nabla_{\tau_{x,1}}\Psi_Y(\theta) & \dots & \nabla_{\tau_{x,t}}\Psi_Y(\theta) & \dots & \nabla_{\tau_{x,T}}\Psi_Y(\theta) \\ 0_{h_x\times 1} & \nabla_{\tau_{x,1}}\Psi_X(\theta) & \dots & \nabla_{\tau_{x,t}}\Psi_X(\theta) & \dots & \nabla_{\tau_{x,T}}\Psi_X(\theta) \end{pmatrix},$$

where  $\nabla_{\tau_1^y}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\tau_1^y$ ,  $\nabla_{\tau_{x,t}}\Psi_Y(\theta) \equiv \partial\Psi_Y(\theta)/\partial\tau_{x,t}$  and  $\nabla_{\tau_{x,t}}\Psi_X(\theta) \equiv \partial\Psi_X(\theta)/\partial\tau_{x,t}$ . 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  $\nabla_{\tau_{x,t}}\Psi_X^{(j_1:j_2,:)}(\theta)$  denote the sub-matrix of  $\nabla_{\tau_{x,t}}\Psi_X(\theta)$  from row  $j_1$  to  $j_2$  and containing all columns, then

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

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  $\nabla_{\tau_{x,t}}\Psi_X(\theta)$  are all equal to 0.

## References

- Blundell, M. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115-143.
- Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68:5-27.
- 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*, 87:115-143
- 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.



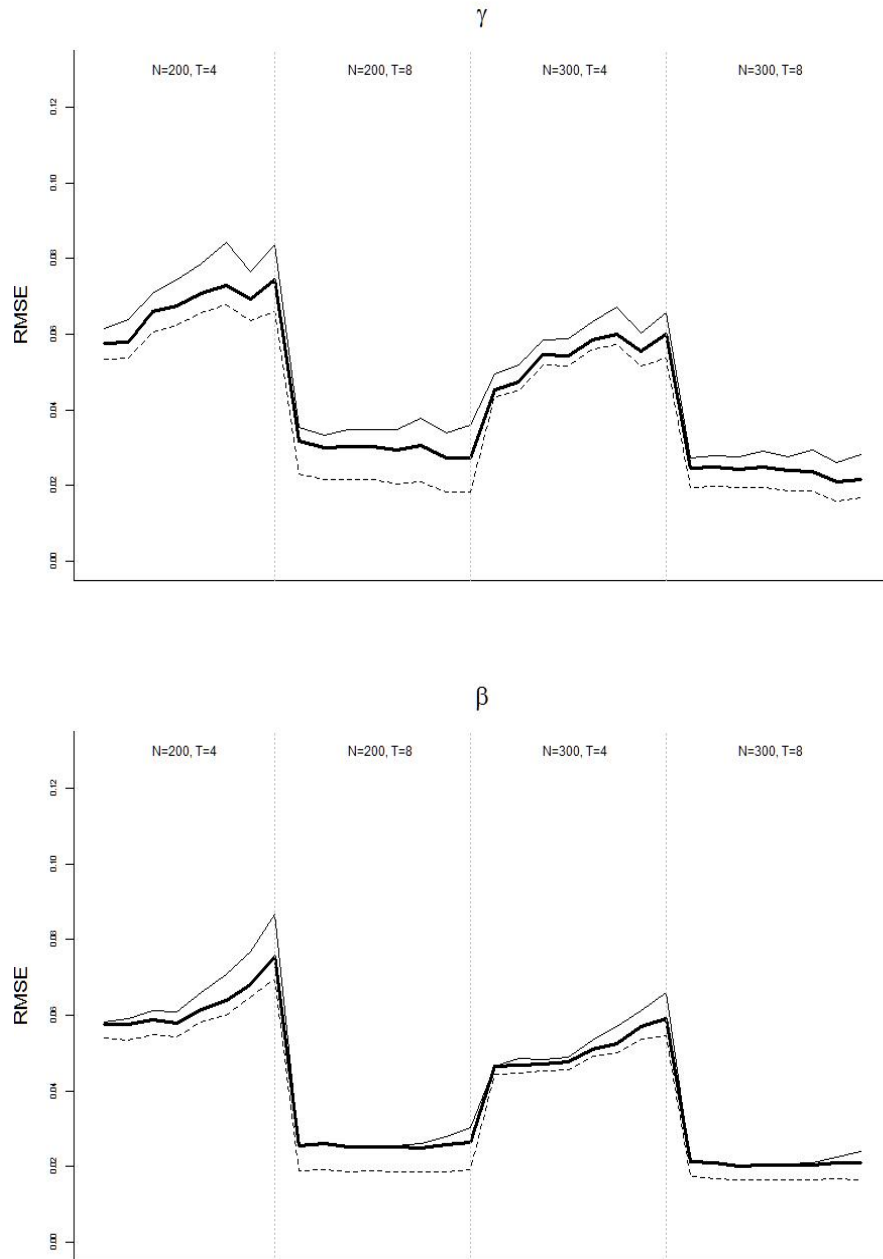
- Chamberlain, G. (1984). Panel Data. In Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, ch. 22, pp. 1247-1318. North-Holland.
- Crepon, B. and J. Mairesse (2008). The Chamberlain approach to panel data: an overview and some simulations. In L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data* (3<sup>rd</sup> ed.), ch. 5, pp. 113-183. Springer.
- Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press.
- Harris, M., L. Matyas, and P. Sevestre (2008). Dynamic models for short panels. In L. Matyas and P. Sevestre (eds.), *The Econometrics of Panel Data* (3<sup>rd</sup> ed.), ch. 8, pp. 249-278. Springer.
- 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.
- 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 R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, ch. 36, pp. 2111-2245. North-Holland.
- Nickel, S. (1981). Bias in dynamic models with fixed effects. *Econometrica*, 49(6):1417-1426.
- Roodman, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135-158.
- 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.

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.

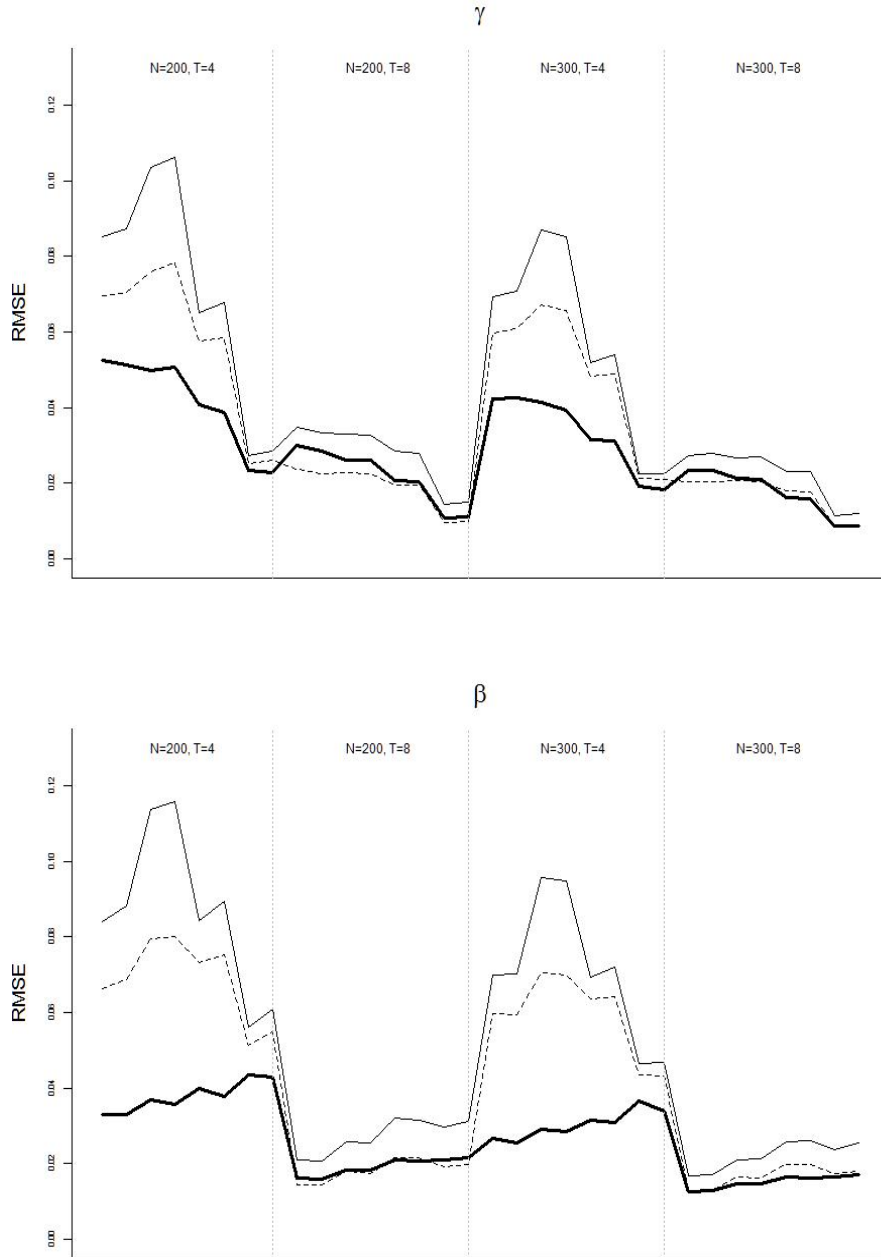
Zincenko, F., W. Sosa-Escudero, and G. Montes-Rojas (2014). Robust tests for time-invariant individual heterogeneity versus dynamic state dependence. *Empirical Economics*, 47(4):1365-1387.

Figure 1: RMSE for the stationary  $X$  case.



Notes: RMSE for the estimation of  $\gamma$  (top panel) and  $\beta$  (bottom panel) in the  $X_t$  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 Tables 1 and 2.

Figure 2: RMSE for the non-stationary  $X$  case.



Notes: RMSE for the estimation of  $\gamma$  (top panel) and  $\beta$  (bottom panel) in the  $X_t$  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 Tables 3 and 4.

Table 1: Bias and RMSE for estimating  $\gamma$ ,  $X$  stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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  $\beta$ ,  $X$  stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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

Table 3: Bias and RMSE for estimating  $\gamma$ ,  $X$  non-stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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  $\beta$ ,  $X$  non-stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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



Table 5: Bias and RMSE for estimating  $\sigma_\mu^2$ ,  $X$  stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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  $\sigma_\mu^2$ ,  $X$  non-stationary

$N$	$T$	$\gamma$	$\rho_\tau$	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