Dynamic panel data with covariance parameterization

Gabriel Montes-Rojas* Walter Sosa-Escudero† Federico Zincenko‡

October 26, 2017

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 and Bond (1991) and Ahn and Schmidt (1995, 1997). 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.

JEL Classification: C12, C23.

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

*E-mail: gabriel.montes@fce.uba.ar. CONICET-IIEP-BAIRES-Universidad de Buenos Aires

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

‡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. 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. 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. 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 (see, e.g., Crepon and Mairesse, 2004) and following Robertson and Sarafidis (2015), we parameterize the covariances between regressors and these effects. The moment conditions implied by these parameterizations 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. The asymptotic framework corresponds to the large N, finite and small T case. Among other results, 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. 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 is computationally easy to implement and performs fairly comparable to Arellano and Bond (1991) and Ahn and Schmidt (1995), the benchmark cases in the literature.

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. Our proposed model is a particular case of the framework proposed by Robertson and Sarafidis (2015), which works for a general multi-factor structure, where we explicitly develop the canonical case of an autoregressive fixed-effects panel data model with weakly exogenous regressors. 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 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

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

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. 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, $\text{var}(\varepsilon_{it}) > 0$ 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 one 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 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 $t = 2, \ldots, T$ 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.

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 $h = (T-1) + h_y + h_x$, $\quad h_y = T(T-1)/2$, $h_x = k(T+2)(T-1)/2$, where

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

and

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

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 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

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

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, $E(y_{i1}\mu_i), E(y_{i2}\mu_i), \ldots, E(y_{iT}\mu_i)$, $E(x_{i1}\mu_i), \ldots, E(x_{iT}\mu_i)$. 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 parameterization can be changed to accomodate the same covariance parameter across t or for some covariates to be uncorrelated with μ_i, thus generating a mixed (i.e., random- and fixed-effects structure). 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., $E(y_{i1}\mu_i)$.

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:* $E(I_2\varepsilon_i) = E(u_i) - E(\mu_i)\iota_2 = 0_{2\times 1}$,

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

and

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

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

Let $\tau_{1\circ}^{y} = E(y_{i1}\mu_i)$ and $\tau_{t\circ}^{x} = E(x_{it}\mu_i)$ for $t = 1, \ldots, T$. We consider these covariances as parameters of our model. Let $\tau_\circ = (\tau_{1\circ}^{y}, \tau_{1\circ}^{x\prime}, \ldots, \tau_{T\circ}^{x\prime})'$ 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 $E(y_{i2}\mu_i), \ldots, E(y_{iT}\mu_i)$ 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 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

$$\underset{h_y\times 1}{\Psi_Y(\theta)} = (\tau_1^y, \tau_1^y, \psi_2(\theta), \tau_1^y, \psi_2(\theta), \psi_3(\theta), \tau_1^y, \ldots, \tau_1^y, \ldots, \psi_{T-1}(\theta))',$$
$$\underset{h_x\times 1}{\Psi_X(\theta)} = (\tau_1^{x\prime}, \tau_2^{x\prime}, \tau_1^{x\prime}, \tau_2^{x\prime}, \tau_3^{x\prime}, \tau_1^{x\prime}, \ldots, \tau_1^{x\prime}, \ldots, \tau_T^{x\prime})',$$

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 $\{[t(t-1)/2]+1 : t = 1, \ldots, T-1\}$ of $\Psi_Y(\theta)$, $\psi_2(\theta)$ occupies positions $\{[t(t-1)/2]+2 : t = 2, \ldots, T-1\}$ of $\Psi_Y(\theta)$, and in general for $2 \le j \le T-1$, $\psi_j(\theta)$ occupies positions $\{[t(t-1)/2]+j : t = j, \ldots, T-1\}$ 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 $t = 1, \ldots, T$ and $l = \max\{t-1, 1\}, \ldots, T-1$, 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 $\psi_t(\theta_\circ) = E(y_{it}\mu_i)$ 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

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

and

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

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 $\theta = (\kappa', \sigma^2_\varepsilon, \sigma^2_\mu, \tau')$, $\kappa = (\alpha, \gamma, \beta')'$, 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 $\theta \in \Theta$.

Example ($k = 1$ & $T = 3$, cont.). *The function $g_i(\cdot)$ is given by*

$$\underset{10\times 1}{g_i(\theta)} = \begin{pmatrix} y_{i2} - \alpha - \gamma y_{i1} - \beta x_{i2} \\ y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3} \\ y_{i1}(y_{i2} - \alpha - \gamma y_{i1} - \beta x_{i2}) \\ y_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ y_{i2}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i1}(y_{i2} - \alpha - \gamma y_{i1} - \beta x_{i2}) \\ x_{i2}(y_{i2} - \alpha - \gamma y_{i1} - \beta x_{i2}) \\ x_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i2}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i3}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \tau^y_1 \\ \tau^y_1 \\ \gamma\tau^y_1 + \tau^x_2\beta + \sigma^2_\mu \\ \tau^x_1 \\ \tau^x_2 \\ \tau^x_1 \\ \tau^x_2 \\ \tau^x_3 \end{pmatrix}. \tag{9}$$

Write $\tilde{\mathbf{Z}}_i = (\tilde{\mathbf{Z}}_{Yi}, \tilde{\mathbf{Z}}_{Xi})_{(T-2)\times\tilde{h}}$, where $\tilde{\mathbf{Z}}_{Yi}$ and $\tilde{\mathbf{Z}}_{Xi}$ are constructed by removing the last of row of $\mathbf{Z}_{Yi}$ and $\mathbf{Z}_{Xi}$, respectively, and $\tilde{h} = [k(T+1)+T-1](T-2)/2$. Denote further $\Delta y_{i,-1} = (y_{i2} - y_{i1}, \ldots, y_{i,T-1} - y_{i,T-2})'$, $\Delta\tilde{x}_i = (x_{i3} - x_{i2}, \ldots, x_{iT} - x_{i,T-1})'$, 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 and it implies a verifiable condition for any given sample. 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, $\theta_\circ$, as the unique solution of a system of moment conditions based on $g_i(\cdot)$.

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

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

$$\mathrm{D}^{AB} g_i(\theta) = \tilde{\mathbf{Z}}_i\left(\Delta y_i - \gamma\Delta y_{i,-1} - \Delta\mathbf{x}_i\beta\right) = \tilde{\mathbf{Z}}_i\Delta y_i - \tilde{\mathbf{Z}}_i\Delta\mathbf{x}_i\begin{pmatrix}\gamma\\ \beta\end{pmatrix} \tag{10}$$

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

$$\underset{3\times 10}{\mathrm{D}^{AB}} = \begin{pmatrix} 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}. \tag{11}$$

Observe that $E[\mathrm{D}^{AB} g_i(\tilde{\theta})] = \mathrm{D}^{AB} E[g_i(\tilde{\theta})] = 0_{\tilde{h}\times 1}$, so equation (10) yields Arellano and Bond (1991)'s 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 $E[g_i(\tilde{\theta})] = 0_{h\times 1}$, 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 $E[g_i(\tilde{\theta})] = 0$, 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 $t = 1, \ldots, T$. 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

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

□

4 Optimal GMM estimator

This section proposes a GMM based estimator for $\theta_\circ$ and derives its asymptotic properties. We begin by constructing an estimator for $\Omega \equiv E[g_i(\theta_\circ)g_i(\theta_\circ)']$, whose inverse will be the optimal weighting matrix in the GMM context. Since $g_i(\theta_\circ) = \mathbf{Z}_i'u_i - E(\mathbf{Z}_i'u_i)$ 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 Ω, it suffices to build a consistent (first-step) estimator of $\theta_\circ$. We suggest using $\dot{\theta} = (\dot{\alpha}, \dot{\gamma}, \dot{\beta}', \dot{\sigma}_\mu^2, \dot{\tau}')$ where $(\dot{\gamma}, \dot{\beta}')$ is the first-step Arellano-Bond estimator of $(\gamma_\circ, \beta_\circ')$,

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

$\dot{\tau} = (\dot{\tau}_1^y, \dot{\tau}_1^{x\prime}, \ldots, \dot{\tau}_T^{x\prime})'$, and $\dot{u}_{it} = y_{it} - \dot{\alpha} - \dot{\gamma} y_{i,t-1} - x_{it}'\dot{\beta}$. It follows that $\dot{\theta} \xrightarrow{P} \theta_\circ$, where $\xrightarrow{P}$ denotes convergence in probability. We highlight that the asymptotic results of the paper will not be affected if $\dot{\theta}$ is replaced by another consistent estimator of θ_0. The natural estimator of Ω is then given by

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

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

The next assumption states that Ω 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 $\theta_\circ$ 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 $\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\beta}', \hat{\sigma}_{\mu}^2, \hat{\tau}')$. From Theorem 2.6 in Newey and McFadden (1994), we have $\hat{\theta} \xrightarrow{P} \theta_\circ$. Since $\bar{g}(\cdot)$ is continuously differentiable, $\hat{\theta}$ can be characterized as a solution of the system of (nonlinear) equations

$$\nabla_\theta \bar{g}(\theta)' \dot{\Omega}^{-1} \bar{g}(\theta) = 0_{[k(T+1)+4]\times 1}, \tag{12}$$

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

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

Two remarks are noteworthy. First, computational differentiation is not required to compute $\nabla_\theta \bar{g}(\theta)$. A closed-form expression of the Jacobian matrix of $\Psi(\theta)$ is provided in the Appendix. Second, the system of equations (12) can be solved by standard numerical methods –such as Newton-Raphson algorithm–. We suggest using the first-step estimator $\dot{\theta}$ as initial value. Observe that $\dot{\theta}$ and $\hat{\theta}$ are close to each other because both estimators consistently approach $\theta_\circ$. Hence, the initial value for solving (12) is close to the global minimum and therefore multiple local minima (or equivalently, multiple solutions to eq. (12)) are not a concern. Further computational aspects are discussed in the next section.

The next assumption is imposed to derive the asymptotic normality of $\hat{\theta}$. Write $G = E[\nabla_\theta g_i(\theta_0)]$.

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

In the rest of this paper, we suppose that Assumptions 1-5 hold. In the Monte Carlo simulations (Section 6 below), we provide concrete examples of data generating processes that satisfy these assumptions. We highlight that Assumptions 3–5 can verified using available data. Specifically, Assumption 3 is based on observable covariates and instruments. Assumption 4 can be verified by checking the covariance matrix estimator $\dot{\Omega}$. Assumption 5 can be corroborated by replacing Ω and G with $\dot{\Omega}$ and $\nabla_\theta \bar{g}(\dot{\theta})$, respectively.

Let $\xrightarrow{D}$ denote convergence in distribution for $N \to \infty$ and T fixed, 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

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

This result follows by Theorem 3.4 in Newey and McFadden (1994). 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 $\hat{\Omega} = (1/N)\sum_{i=1}^{N} g_i(\hat{\theta})g_i(\hat{\theta})'$; 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 $H_0 : (\sigma^2_{\mu\circ}, \tau'_\circ) = 0_{1\times(kT+2)}$, 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, Mátyás, 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. Moreover, we show that our estimator can be obtained at no additional computational cost despite having more parameters in our model.

Two cases are considered throughout this section: $T = 3$ and $T = 4$. The arguments and obtained results can be easily extended to the general case $T \geq 4$.

5.1 Case $T = 3$: comparison with Arellano and Bond (1991)

In this subsection, we suppose that $T = 3$ and $k = 1$ as in the previous examples. We consider Arellano-Bond estimator for comparison purposes as Ahn and Schmidt (1995)'s method requires $T \geq 4$ to incorporate quadratic moment conditions (see next subsection).

In their seminal paper, Arellano and Bond (1991) propose estimating $(\gamma_\circ, \beta_\circ)$ by GMM using the function

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

based on the moment conditions

$$E \begin{pmatrix} y_{i1}\Delta u_{i3} \\ x_{i1}\Delta u_{i3} \\ x_{i2}\Delta u_{i3} \end{pmatrix} = 0_{3\times 1}, \tag{13}$$

where $\Delta u_{i3} = u_{i3} - u_{i2}$. The optimal Arellano-Bond GMM estimator is

$$(\hat{\gamma}^{AB}, \hat{\beta}^{AB}) = \underset{(\gamma,\beta)}{\operatorname{argmin}}\ \bar{g}^{AB}(\gamma, \beta)'(\dot{\Omega}^{AB})^{-1}\bar{g}^{AB}(\gamma, \beta), \tag{14}$$

where $\bar{g}^{AB}(\gamma, \beta) = (1/N)\sum_{i=1}^{N} g_i^{AB}(\gamma, \beta)$ and $\dot{\Omega}^{AB}$ is a consistent estimator of

$$\Omega^{AB} \equiv E\left[g_i^{AB}(\gamma_\circ, \beta_\circ) g_i^{AB}(\gamma_\circ, \beta_\circ)'\right].$$

Its asymptotic distribution is given by

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

where $\Sigma_{\gamma\beta}^{AB} = [G^{AB\prime}(\Omega^{AB})^{-1}G^{AB}]^{-1}$, $G^{AB} = E[\nabla_{(\gamma,\beta)} g_i^{AB}(\gamma_\circ, \beta_\circ)]$, and

$$\underset{3\times 2}{\nabla_{(\gamma,\beta)} g_i^{AB}}(\gamma, \beta) = \begin{pmatrix} \dfrac{\partial g_i^{AB}(\gamma, \beta)}{\partial \gamma} & \dfrac{\partial g_i^{AB}(\gamma, \beta)}{\partial \beta} \end{pmatrix}.$$

Considering the matrix D^{AB} defined in eq. (11) for the case $(k, T) = (1, 3)$, we observe that $\mathrm{D}^{AB}\Psi(\theta) = 0$, $g_i^{AB}(\gamma, \beta) = \mathrm{D}^{AB} g_i(\theta)$ for all $\theta = (\alpha, \gamma, \beta, \sigma_\mu^2, \tau')$, and $\Omega^{AB} = \mathrm{D}^{AB}\Omega\mathrm{D}^{AB\prime}$. 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–.

To compare our estimator $(\hat{\gamma}, \hat{\beta})$ with $(\hat{\gamma}^{AB}, \hat{\beta}^{AB})$, we construct the function

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

where $\theta_{\setminus\gamma\beta} = (\alpha, \tau_1^y, \tau_1^x, \tau_2^x, \tau_3^x, \sigma_\mu^2)$,

$$g_{i,1}^D(\gamma, \beta) = \begin{pmatrix} y_{i1}(\Delta y_{i3} - \gamma \Delta y_{i2} - \Delta x_{i3}\beta) \\ x_{i1}(\Delta y_{i3} - \gamma \Delta y_{i2} - \Delta x_{i3}\beta) \\ x_{i2}(\Delta y_{i3} - \gamma \Delta y_{i2} - \Delta x_{i3}\beta) \\ \Delta y_{i3} - \gamma \Delta y_{i2} - \Delta x_{i3}\beta \end{pmatrix} = \begin{pmatrix} g_i^{AB}(\gamma, \beta) \\ \Delta y_{i3} - \gamma \Delta y_{i2} - \Delta x_{i3}\beta \end{pmatrix}$$

and

$$g_{i,2}^D(\gamma, \beta; \theta_{\setminus\gamma\beta}) = \begin{pmatrix} y_{i3} - \gamma y_{i2} - \beta x_{i3} \\ y_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i1}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i2}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ x_{i3}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \\ y_{i2}(y_{i3} - \alpha - \gamma y_{i2} - \beta x_{i3}) \end{pmatrix} - \begin{pmatrix} \alpha \\ \tau_1^y \\ \tau_1^x \\ \tau_2^x \\ \tau_3^x \\ \gamma\tau_1^y + \tau_2^x\beta + \sigma_\mu^2 \end{pmatrix}.$$

Observe that $g_i^D(\cdot)$ emerges after applying a one-to-one and onto linear transformation to $g_i(\cdot)$. In other words, along the lines of eq. (11), there is a 10×10 nonsingular matrix D such that $g_i^D(\theta) = \mathrm{D}g_i(\theta)$. This yields

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

where $\bar{g}^D(\theta) = (1/N)\sum_{i=1}^N g_i^D(\theta)$ and $\dot{\Omega}^D = D\dot{\Omega}D'$ is a consistent estimator of $\Omega^D = E[g_i^D(\theta_\circ)g_i^D(\theta_\circ)'] = D\Omega D'$. After partitioning

$$\underset{10\times 10}{\Omega^D} = \begin{pmatrix} \underset{4\times 4}{\Omega_{11}^D} & \underset{4\times 6}{\Omega_{12}^D} \\ \underset{6\times 4}{\Omega_{12}^{D\prime}} & \underset{6\times 6}{\Omega_{22}^D} \end{pmatrix}$$

with $\Omega_{11}^D = E[g_{i,1}^D(\gamma, \beta)g_{i,1}^D(\gamma, \beta)']$, $\hat{\theta}$ can be computed by solving the following (linear) system of equations:

$$\begin{pmatrix} \nabla_{(\gamma,\beta)}\bar{g}_1^D(\gamma, \beta)' & \nabla_{(\gamma,\beta)}\bar{g}_2^D(\gamma, \beta; \theta_{\setminus\gamma\beta})' \\ \nabla_{\theta_{\setminus\gamma\beta}}\bar{g}_1^D(\gamma, \beta)' & \nabla_{\theta_{\setminus\gamma\beta}}\bar{g}_2^D(\gamma, \beta; \theta_{\setminus\gamma\beta})' \end{pmatrix} \begin{pmatrix} \dot{\Omega}_{11}^D & \dot{\Omega}_{12}^D \\ \dot{\Omega}_{12}^{D\prime} & \dot{\Omega}_{22}^D \end{pmatrix}^{-1} \begin{pmatrix} \bar{g}_1^D(\gamma, \beta) \\ \bar{g}_2^D(\gamma, \beta; \theta_{\setminus\gamma\beta}) \end{pmatrix} = 0_{8\times 1},$$

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

$$\underset{4\times 2}{\nabla_{(\gamma,\beta)} g_{i,1}^D}(\gamma,\beta) = \begin{pmatrix} \dfrac{\partial g_{i,1}^D(\gamma,\beta)}{\partial \gamma} & \dfrac{\partial g_{i,1}^D(\gamma,\beta)}{\partial \beta} \end{pmatrix},$$

$\dot{\Omega}_{11}^D$ is a consistent estimator of Ω_{11}^D, $\bar{g}_1^D(\gamma,\beta) = (1/N)\sum_{i=1}^N g_{i,1}^D(\gamma,\beta)$, 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 $g_{i,1}^D(\cdot,\cdot)$ does not depends on $\theta_{\setminus\gamma\beta}$ ($\nabla_{\theta_{\setminus\gamma\beta}}\bar{g}_1^D(\gamma,\beta) = 0_{4\times 6}$), our estimator of $(\gamma_\circ, \beta_\circ)$ can be obtained by solving the (linear) system of equations

$$\nabla_{(\gamma,\beta)}\bar{g}_1^D(\gamma,\beta)'(\dot{\Omega}_{11}^D)^{-1}\bar{g}_1^D(\hat{\gamma},\hat{\beta}) = 0_{2\times 1}$$

or, equivalently, by minimizing the objective function $\bar{g}_1^D(\gamma,\beta)'(\dot{\Omega}_{11}^D)^{-1}\bar{g}_1^D(\gamma,\beta)$:

$$(\hat{\gamma},\hat{\beta}) = \underset{(\gamma,\beta)}{\text{argmin}}\ \bar{g}_1^D(\gamma,\beta)'(\dot{\Omega}_{11}^D)^{-1}\bar{g}_1^D(\gamma,\beta). \tag{15}$$

From these expressions, it follows that we can ignore the presence of $\theta_{\setminus\gamma\beta}$, as well as Ω_{12}^D and Ω_{22}^D, when computing $(\hat{\gamma},\hat{\beta})$. So our estimator can be obtained at no additional computational burden.

Estimating the model using moments in levels (when $T = 3$) can essentially be interpreted as adding two terms to Arellano-Bond objective function. This interpretation follows by comparing eqs. (14) and (15). Note first that we incorporate a condition related to $E(\Delta u_{i3}) = 0$ into the objective function through the last row of $\bar{g}_1^D(\gamma,\beta)$. Second, our weighting matrix includes elements related to the interaction between $(\Delta u_{i3})^2 = (\varepsilon_{i3} - \varepsilon_{i2})^2$ and the covariates (y_{i1}, x_{i1}, x_{i2}). Specifically, we can partition

$$\underset{4\times 4}{\Omega_{11}^D} = \begin{pmatrix} \underset{3\times 3}{\Omega^{AB}} & \underset{3\times 1}{\Omega_{11,12}^D} \\ \underset{1\times 3}{\Omega_{11,12}^{D\prime}} & \underset{1\times 1}{\Omega_{11,22}^D} \end{pmatrix}$$

with

$$\Omega_{11,12}^D = E\begin{pmatrix} y_{i1}(\Delta u_{i3})^2 \\ x_{i1}(\Delta u_{i3})^2 \\ x_{i2}(\Delta u_{i3})^2 \end{pmatrix}$$

and $\Omega^D_{11,22} = E[(\Delta u_{i3})^2]$, so the inverse of Ω^D_{11} can be written as

$$(\Omega^D_{11})^{-1} = \begin{pmatrix} (\Omega^{AB})^{-1} + (\Omega^{AB})^{-1}\Omega^D_{11,12}\Upsilon\Omega^{D\prime}_{11,12}(\Omega^{AB})^{-1} & -(\Omega^{AB})^{-1}\Omega^D_{11,12}\Upsilon \\ -\Upsilon\Omega^{D\prime}_{11,12}(\Omega^{AB})^{-1} & \Upsilon \end{pmatrix}$$

with $\Upsilon = \left[\Omega^D_{11,22} - \Omega^{D\prime}_{11,12}(\Omega^{AB})^{-1}\Omega^D_{11,12}\right]^{-1} > 0$.[1] Observe that Arellano and Bond (1991) use only an estimator of $(\Omega^{AB})^{-1}$ as weighting matrix, ignoring the terms $\Omega^D_{11,12}$ and Υ. In the rest of this subsection, we discuss under what conditions the inclusion these terms improves the asymptotic variance.

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

$$\sqrt{N}\left[\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} \gamma_\circ \\ \beta_\circ \end{pmatrix}\right] \xrightarrow{D} N\left(0, \Sigma_{\gamma\beta}\right),$$

where $\Sigma_{\gamma\beta} = \left[G^{D\prime}_1(\Omega^D_{11})^{-1}G^{D\prime}_1\right]^{-1}$, $G^D_1 = E[\nabla_{(\gamma,\beta)}g^D_{i,1}(\gamma_\circ, \beta_\circ)]$, and

$$\underset{4\times 2}{\nabla_{(\gamma,\beta)}g^D_{i,1}(\gamma, \beta)} = \begin{pmatrix} \dfrac{\partial g^D_i(\gamma,\beta)}{\partial\gamma} & \dfrac{\partial g^D_i(\gamma,\beta)}{\partial\beta} \end{pmatrix}.$$

Clearly, $\Sigma_{\gamma\beta}$ coincide with the $(k+1)\times(k+1)$ sub-matrix of Σ associated with $(\gamma_\circ, \beta'_\circ)$ derived in the previous section. To compare $\Sigma_{\gamma\beta}$ and $\Sigma^{AB}_{\gamma\beta}$, we partition

$$\underset{4\times 2}{G^D_1} = \begin{pmatrix} \underset{3\times 2}{G^{AB}} \\ \underset{1\times 2}{G^D_{1,2}} \end{pmatrix}$$

with $G^D_{1,2} = -E(\Delta y_{i2}\ \Delta x_{i3})$. Then, we write

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

[1] This inverse is obtained by applying eq. (3.2) of Theil (1983). We have that $\Upsilon > 0$ because $\Omega^D = D\Omega D'$ is positive definite (Assumption 4) and so is its inverse.

and as a result

$$\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AB})^{-1} = [G^{AB\prime}(\Omega^{AB})^{-1}\Omega_{11,12}^{D} - G_{1,2}^{D\prime}]\Upsilon[G^{AB\prime}(\Omega^{AB})^{-1}\Omega_{11,12}^{D} - G_{1,2}^{D\prime}]'. \quad (16)$$

Recall that the difference $\Sigma_{\gamma\beta}^{AB} - \Sigma_{\gamma\beta}$ is positive (semi)definite if and only if $\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AB})^{-1}$ is positive (semi)definite. From eq. (16) and since $\Upsilon > 0$, it is clear that $\Sigma_{\gamma\beta}^{-1} - (\Sigma_{\gamma\beta}^{AB})^{-1}$ is positive semidefinite, which we already know as our estimator uses the optimal weighting matrix.

The efficiency gain of our estimator comes from two sources: the interaction between $(\varepsilon_{i3} - \varepsilon_{i2})^2$ and the covariates (captured by $\Omega_{11,12}^{D}$) and the variation in the unconditional mean of the covariates across time (captured by $G_{1,2}^{D}$). There is no efficiency gain ($\Sigma_{\gamma\beta}^{AB} = \Sigma_{\gamma\beta}$), e.g., when $\Omega_{11,12}^{D} = 0_{3\times 1}$ and $G_{1,2}^{D} = 0_{1\times 2}$. Among other conditions, we have that $\Omega_{11,12}^{D} = 0_{3\times 1}$ if (y_{i1}, x_{i1}, x_{i2}) has zero mean and is independent of $\Delta u_{i3} = \varepsilon_{i3} - \varepsilon_{i2}$, while $G_{1,2}^{D} = 0_{1\times 2}$ if and only if $E[(\Delta y_{i2}\ \Delta x_{i3})] = 0_{1\times 2}$. The difference $\Sigma_{\gamma\beta}^{AB} - \Sigma_{\gamma\beta}$ is positive definite, e.g., when $\Omega_{11,12}^{D} = 0$ and $E[(\Delta y_{i2}\ \Delta x_{i3})] \neq 0_{1\times 2}$.

5.2 Case $T = 4$: comparison with Ahn and Schmidt (1995)

Here, we set $T = 4$ and assume $k = 0$ to simplify the exposition, so model (1) becomes $y_{it} = \alpha_\circ + \gamma_\circ y_{i,t-1} + u_{it}$ with $u_{it} = \mu_i + \varepsilon_{it}$ and the parameters are $(\alpha_\circ, \gamma_\circ, \sigma_{\mu\circ}^2, \tau_{1\circ}^{y})$. The results of this subsection can be easily extended to the general case $T \geq 4$. We consider Ahn-Schmidt estimator for comparison purposes because it is as efficient as Arellano and Bond (1991)'s estimator.

Ahn and Schmidt (1995, 1997) focus on estimating $\gamma_\circ$ and suggest exploiting supplementary orthogonality conditions derived from the standard assumptions. Specifically, they consider

$$E\begin{pmatrix} y_{i1}\Delta u_{i3} \\ y_{i1}\Delta u_{i4} \\ y_{i2}\Delta u_{i4} \\ u_{i4}\Delta u_{i3} \end{pmatrix} = 0_{4\times 1}; \quad (17)$$

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 is one is the 'quadratic' moment condition that was added by Ahn and Schmidt (1995) to exploit the absence of serial correlation in ε_{it}. The optimal Ahn-Schmidt GMM estimator is

$$\hat{\gamma}^{AS} = \underset{\gamma}{\operatorname{argmin}}\ \bar{g}^{AS}(\gamma)'(\dot{\Omega}^{AS})^{-1}\bar{g}^{AS}(\gamma), \tag{18}$$

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

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

and $\dot{\Omega}^{AS}$ is a consistent estimator of $\Omega^{AS} = E[g_i^{AS}(\gamma_\circ)g_i^{AS}(\gamma_\circ)']$. Observe that $g_i^{AS}(\gamma) = \tilde{\mathrm{D}}^{AS}(\gamma)g_i(\alpha, \gamma, \sigma_\mu^2, \tau_1^y)$ with

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

The asymptotic distribution of the Ahn-Schmidt estimator is given by $\sqrt{N}(\hat{\gamma}^{AS} - \gamma_\circ) \xrightarrow{D} N\left(0, \Sigma_\gamma^{AS}\right)$, where $\Sigma_\gamma^{AS} = [G^{AS\prime}(\Omega^{AS})^{-1}G^{AS}]^{-1}$, $G^{AS} = E[\nabla_\gamma g_i^{AS}(\gamma_\circ)]$, and

$$\nabla_\gamma \underset{4\times 1}{g_i^{AS}}(\gamma) = \frac{\partial g_i^{AS}(\gamma)}{\partial\gamma}.$$

For comparison purposes, we construct the function

$$g_i^{\tilde{D}}(\theta) = \begin{pmatrix} g_{i,1}^{\tilde{D}}(\gamma) \\ g_{i,2}^{\tilde{D}}(\gamma; \alpha, \tau_1^y, \sigma_\mu^2) \end{pmatrix},$$

where

$$g_{i,1}^{\tilde{D}}(\gamma) = \begin{pmatrix} y_{i1}(\Delta y_{i3} - \gamma\Delta y_{i2}) \\ y_{i1}(\Delta y_{i4} - \gamma\Delta y_{i3}) \\ y_{i2}(\Delta y_{i4} - \gamma\Delta y_{i3}) \\ (y_{i4} - \gamma y_{i3})(\Delta y_{i3} - \gamma\Delta y_{i2}) \\ \Delta y_{i3} - \gamma\Delta y_{i2} \\ \Delta y_{i4} - \gamma\Delta y_{i3} \end{pmatrix} = \begin{pmatrix} g_i^{AS}(\gamma) \\ \Delta y_{i3} - \gamma\Delta y_{i2} \\ \Delta y_{i4} - \gamma\Delta y_{i3} \end{pmatrix}$$

and

$$g_{i,2}^{\tilde{D}}(\gamma; \alpha, \tau_1^y, \sigma_\mu^2) = \begin{pmatrix} y_{i4} - \gamma y_{i3} \\ y_{i1}(y_{i3} - \alpha - \gamma y_{i2}) \\ y_{i2}(y_{i3} - \alpha - \gamma y_{i2}) \end{pmatrix} - \begin{pmatrix} \alpha \\ \tau_1^y \\ \gamma\tau_1^y + \sigma_\mu^2 \end{pmatrix}.$$

Observe that $g_i^{\tilde{D}}(\cdot)$ is obtained after applying a one-to-one and onto transformation to $g_i(\cdot)$; more specifically, there is a nonsingular 9×9 matrix $\tilde{\mathrm{D}}(\gamma)$ such that $g_i^{\tilde{D}}(\gamma) = \tilde{\mathrm{D}}(\gamma) g_i(\gamma)$.[2] Following similar arguments to the ones in previous subsection, our estimator of $\gamma_\circ$ can be expressed as

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}}\ \bar{g}_1^{\tilde{D}}(\gamma)'(\dot{\Omega}_{11}^{\tilde{D}})^{-1}\bar{g}_1^{\tilde{D}}(\gamma),$$

where $\bar{g}_1^{\tilde{D}}(\gamma) = (1/N)\sum_{i=1}^N g_{i,1}^{\tilde{D}}(\gamma)$ and $\dot{\Omega}_{11}^{\tilde{D}}$ is a consistent estimator of

$$\Omega_{11}^{\tilde{D}} = E[g_{i,1}^{\tilde{D}}(\gamma_\circ) g_{i,1}^{\tilde{D}}(\gamma_\circ)'].$$

From this expression, essentially, estimating the dynamic panel using moments in levels when $T = 4$ (or more generally, when $T \geq 4$) is equivalent to adding the conditions associated with $(\Delta u_{i3}, \Delta u_{i4})$ to Ahn-Schmidt objective function (18). These conditions are incorporated through the last two rows of $g_{i,1}^{\tilde{D}}(\cdot)$ and the weighting matrix $(\dot{\Omega}_{11}^{\tilde{D}})^{-1}$ without affecting the computational cost. More specifically, $\Omega_{11}^{\tilde{D}}$ can be partitioned as

$$\underset{6\times 6}{\Omega_{11}^{\tilde{D}}} = \begin{pmatrix} \underset{4\times 4}{\Omega^{AS}} & \underset{4\times 2}{\Omega_{11,12}^{\tilde{D}}} \\ \underset{2\times 4}{\Omega_{11,12}^{\tilde{D}\prime}} & \underset{2\times 2}{\Omega_{11,22}^{\tilde{D}}} \end{pmatrix}$$

with

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

and

$$\Omega_{11,22}^{\tilde{D}} = E\begin{pmatrix} (\Delta u_{i3})^2 & \Delta u_{i3}\Delta u_{i4} \\ \Delta u_{i3}\Delta u_{i4} & (\Delta u_{i4})^2 \end{pmatrix},$$

so the terms $\Omega_{11,12}^{\tilde{D}}$ and $\Omega_{11,22}^{\tilde{D}}$ are included in our objective function through $(\dot{\Omega}_{11}^{\tilde{D}})^{-1}$.

[2] $\tilde{\mathrm{D}}(\gamma)$ can be constructed by extending $\tilde{\mathrm{D}}^{AS}(\gamma)$ and can be compared with the transformation of eq. (A.7) in Ahn and Schmidt (1995). We remark that the dependence of $\tilde{\mathrm{D}}(\gamma)$ on γ does not affect the asymptotic distribution of the estimator as it depends; see e.g. Hall (2005, sec. 3.7).

The asymptotic distribution of $\hat{\gamma}$ is given by $\sqrt{N}(\hat{\gamma}-\gamma_{\circ}) \xrightarrow{D} N(0,\Sigma_{\gamma})$, where $\Sigma_{\gamma} = [G_1^{\tilde{D}\prime}(\Omega^{\tilde{D}})^{-1}G_1^{\tilde{D}}]^{-1}$, $G_1^{\tilde{D}} = E[\nabla_{\gamma} g_{i,1}^{\tilde{D}}(\gamma_{\circ})]$, and

$$\underset{6\times 1}{\nabla_{\gamma} g_{i,1}^{\tilde{D}}(\gamma)} = \frac{\partial g_{i,1}^{\tilde{D}}(\gamma)}{\partial \gamma}.$$

To compare the asymptotic variances, Σ_{γ}^{AS} and Σ_{γ}, we partition

$$\underset{6\times 1}{G_1^{\tilde{D}}} = \begin{pmatrix} \underset{4\times 1}{G^{AS}} \\ \underset{2\times 1}{G_{1,2}^{\tilde{D}}} \end{pmatrix}$$

with $G_{1,2}^{\tilde{D}} = -E[(\Delta y_{i2}\ \Delta y_{i3})']$, so we can write

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

being $\tilde{\Upsilon} \equiv \left[\Omega_{11,22}^{\tilde{D}} - \Omega_{11,12}^{\tilde{D}\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{\tilde{D}}\right]^{-1}$ positive definite. As in the previous subsection, the efficiency gain from using moment conditions in levels depends crucially on the terms $\Omega_{11,12}^{\tilde{D}}$ and $G_{1,2}^{\tilde{D}}$. Our estimator of $\gamma_{\circ}$ is asymptotically equivalent to the one developed by Ahn and Schmidt (1995) if and only if $G^{AS\prime}(\Omega^{AS})^{-1}\Omega_{11,12}^{\tilde{D}} = G_{1,2}^{\tilde{D}\prime}$.

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, p. 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 $T \in \{4, 8\}$.[3] 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. Tables 1 and 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 Tables 1 and 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.

[INSERT TABLES 1 AND 2 HERE]

[INSERT FIGURE 1 HERE]

When X_t is non-stationary (Tables 3 and 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 previous section's discussion. For example, when $\gamma = 0.25$ and $\rho = 0.25$, the RMSE of our procedure

[3] Although not reported, we also considered additional sample sizes and distributions as in Wu and Zhu (2012).

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 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 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 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 γ_2^μ, simulations for $N = 200, T = 8$ and $N = 300, T = 8$. This corresponds to $\rho = 0$ 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 γ increases, pointing out that, as in Zincenko, 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 Zincenko 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 $\Psi(\theta)$, denoted by $\nabla_\theta \Psi(\theta)$, can partitioned in 5 blocks:

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

It follows immediately that $\nabla_\alpha \Psi(\theta) = 0_{h\times 1}$. We provide below 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 $\nabla_\theta g_i(\theta)$.

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 $\{[t(t-1)/2]+1 : t = 1,\ldots,T-1\}$, $\partial\psi_2(\theta)/\partial\gamma$ occupies positions $\{[t(t-1)/2]+2 : t = 2,\ldots,T-1\}$, and in general $\partial\psi_j(\theta)/\partial\gamma$ occupies positions $\{[t(t-1)/2]+j : t = j,\ldots,T-1\}$ 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,

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

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 $\nabla_{\sigma_\mu^2} \Psi(\theta)$.

Next consider $\nabla_\tau \Psi(\theta)$. 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 $\nabla_{\tau_{x,t}} \Psi_Y(\theta)$ 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 $\nabla_{\tau_{x,t}} \Psi_Y(\theta)$ 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^{rd} 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^{rd} 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.

Robertson, D. and V. Sarafidis (2015). IV estimation of panel with factor residuals. *Journal of Econometrics*, 185:526-541.

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.

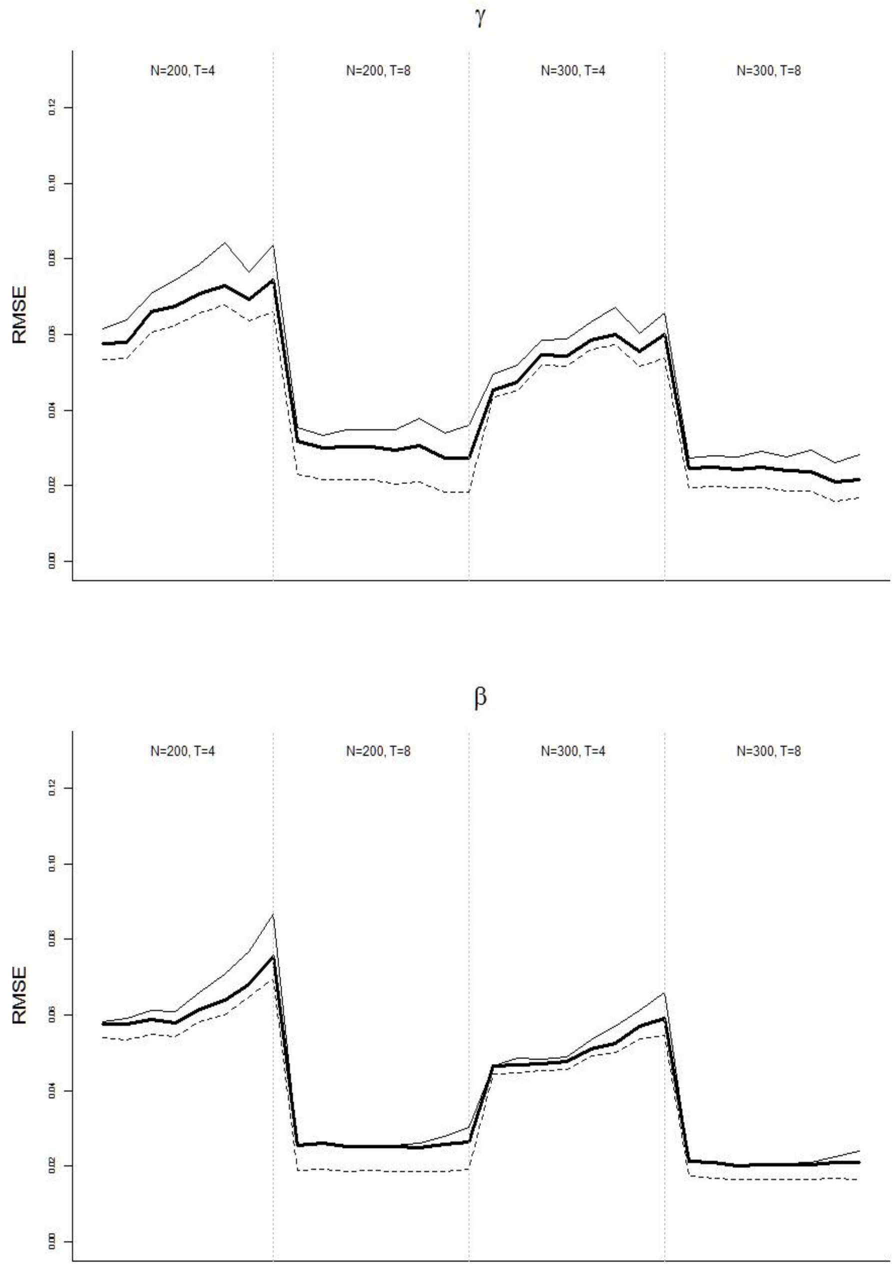
Theil, H. (1983). Linear algebra and matrix methods in econometrics. In Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1, ch. 1, pp. 3-65. North-Holland.

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.

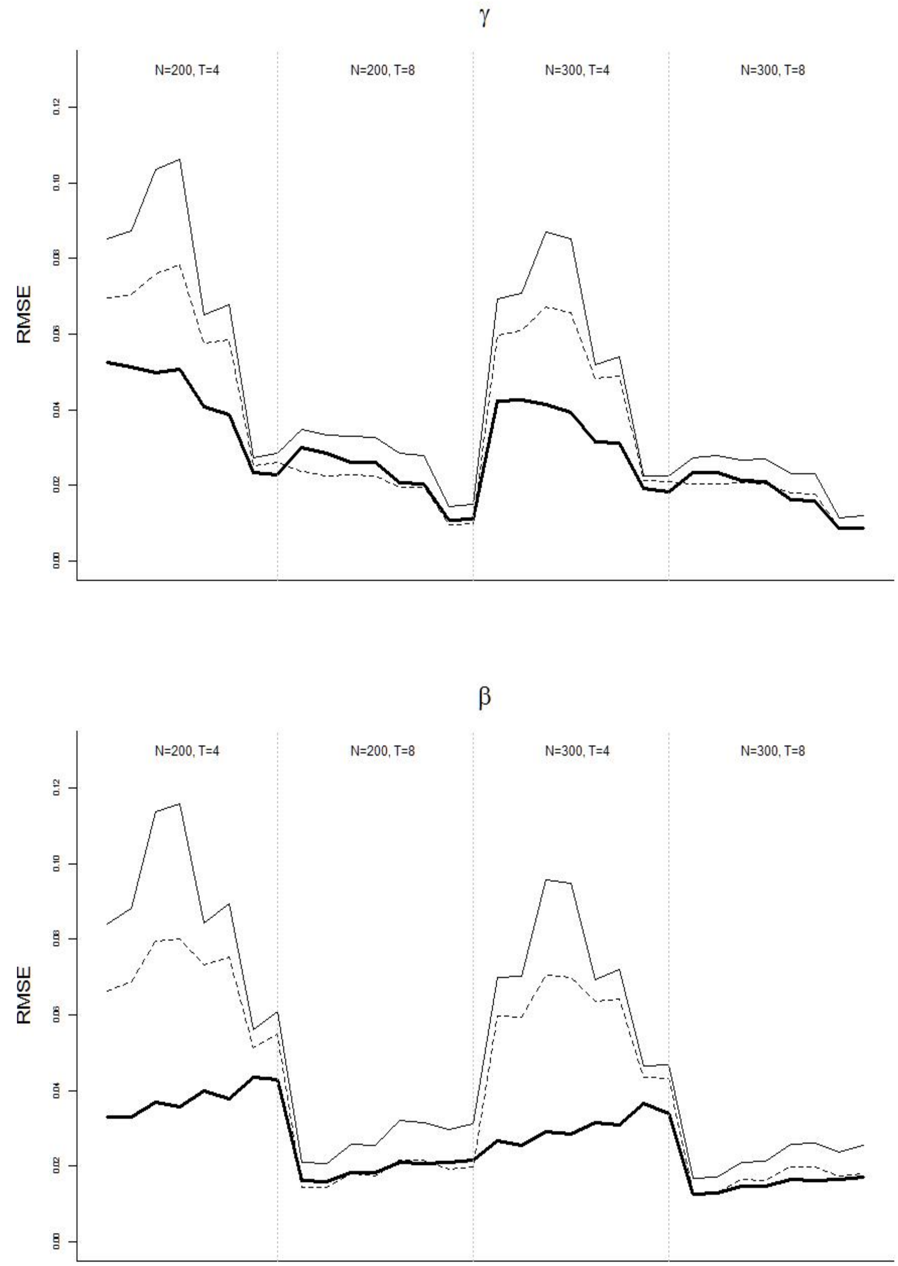
Zincenko, F., W. Sosa-Escudero, and G. Montes-Rojas (2014). Robust tests for time-invariant individual heterogeneity versus dynamic state dependence. *Empirical Economics*, 47(4):1365-1387.

Figure 1: RMSE for the stationary X case.



Notes: RMSE for the estimation of γ (top panel) and β (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 γ (top panel) and β (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 γ, X stationary

				Bias			RMSE		
N	T	γ	ρ	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 stationary

				Bias			RMSE		
N	T	γ	ρ	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 γ, X non-stationary

				Bias			RMSE		
N	T	γ	ρ	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 non-stationary

				Bias			RMSE		
N	T	γ	ρ	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 σ_μ^2, X 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_{μ}^{2}, X 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