Pairwise difference estimation of dynamic panel data models

Michele Aquaro[*] Pavel Čížek[†‡]

November 14, 2012

Abstract

A new estimation procedure of dynamic panel data models with fixed effects is proposed. To improve upon existing estimators, we propose to apply the pairwise-difference data transformation to the generalized method of moments based estimators. A particular focus is given to the long difference (LD) estimation procedure of Hahn et al. (2007), which was proved to retain strong moment conditions even when data are persistent without imposing further assumptions. The bias and asymptotic distribution of the original LD estimator and its proposed extensions are derived. A simulation study is conducted to assess the finite-sample properties of the estimators.

JEL code: C13, C23, C26

Keywords: asymptotic distribution, dynamic panel data, generalized method of moments, long difference, pairwise differences

1 Introduction

The estimation of the dynamic panel data model with fixed effects has been extensively studied in last decades. It is well known that the least square dummy variable (LSDV) estimator is inconsistent when applied to dynamic panels with a small fixed number of

[*]Tilburg University, the Netherlands.

[†]Corresponding author: Tel.: +31 13 466 8723. Fax: +31 13 466 3280. *E-mail:* P.Cizek@uvt.nl.

[‡]CentER, Dept. of Econometrics and OR, Tilburg School of Economics and Management, Tilburg University, the Netherlands.

time periods (Nickell, 1981). As a consequence, the majority of research has focused on the generalized method of moments (GMM) procedures and estimation methods based on instrumental variable (IV) methods (e.g., Anderson and Hsiao, 1982; Holtz-Eakin et al., 1988; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). However, some of these estimators have been found to suffer heavily from various sources of bias. Models in the first differences with instruments in levels can be substantially biased, in particular when the autoregressive parameter is close to unity. Models in levels with instrument in the first differences, specifically designed for persistent series, rely crucially on the stationarity assumption (see Hahn, 1999). As an alternative approach, bias-reduction methods for the LSDV and maximum likelihood estimators have been proposed, see Kiviet (1995), Hahn and Kuersteiner (2002), Bun and Carree (2005), and Gouriéroux et al. (2010).

To improve the estimation when the autoregressive parameter is close to the unit circle, Hahn et al. (2007) suggest to employ the longest difference (LD) of the model, that is, the differences between the last and the first observation for each individual. Using the local-to-unity asymptotics for the autoregressive parameter, the moment conditions defining the LD estimator of Hahn et al. (2007) are chosen from the asymptotically relevant moment condition in order to minimize the estimator's bias. Additionally, to circumvent the non-linearity of the proposed moment conditions, the instruments – being regression residuals – are estimated using an initial consistent estimate.

Although the LD estimator provides a method with a small finite-sample bias without imposing the stationarity assumption, there are two important deficiencies of the method from the practical point of view. First, by using the longest difference of the panel data, the differenced data always contain only one observation for each individual unit irrespective of the number of time periods. Next, the practically applicable asymptotic distribution and variance of the LD estimator is not provided. This is especially important due to the reliance of the LD moment conditions on initially estimated instruments and thus on the properties of the initial estimator.

To rectify these problems and make LD a practically relevant alternative to the GMM estimators such as the one by Blundell and Bond (1998), two steps are necessary. First, we propose to extend the LD estimator by taking more longer differences than just the longest one and show that new estimators have smaller variances while keeping the bias properties almost unchanged. The proposed estimators improve upon the original LD especially for small values autoregressive parameter or larger numbers of time periods. Second, we derive the practically applicable asymptotic-distribution expression for a general class of long-difference estimators — including the original LD estimator — under strong instrument asymptotics. Practical choices and recommendations for the GMM weight matrix are extensively discussed as well. Finally, the theoretical findings are confirmed in finite samples by means of simulation studies.

The rest of the paper is organized as follows. In Section 2, we introduce the dynamic panel data model and the LD estimator. The new estimators are presented in Section 3, where we also study their bias properties. The asymptotic distribution for a finite number of time periods is derived in Section 4. Further, Section 5 contains the results of the Monte Carlo experiments. Finally, Section 6 concludes. The proofs are provided in the Appendix.

2 Long difference estimation of dynamic panels

For a fixed $T \geq 3$ and $n \in \mathbb{N}$, consider the simple dynamic panel data model

$$y_{it} = \alpha y_{it-1} + \eta_i + \varepsilon_{it} \qquad (t = 1, \dots, T; \quad i = 1, \dots, n), \tag{1}$$

where the response variable y_{it} depends on its lagged value y_{it-1} through the unknown autoregressive parameter α, $|\alpha| < 1$, on the unobserved individual fixed effect η_i, and on an idiosyncratic error ε_{it}. Model (1) will be used to describe the estimation concepts, but it can and will be further generalized by including additional explanatory variables; see Section 4.

As the individual effects η_i are not observed, several filtering data transformations have been used in the literature. Many of those rely on the sth difference transformation generically defined as $\Delta^s v_t = v_t - v_{t-s}$ (see Aquaro and Čížek, 2010). More specifically, subtracting (1) at time $t-s$ from its level at time t yields

$$\Delta^s y_{it} = \alpha \Delta^s y_{it-1} + \Delta^s \varepsilon_{it} \qquad (t = s+1, \ldots, T; \quad i = 1, \ldots, n), \tag{2}$$

where the order of the difference s generally ranges from 1 to $T-1$: the Arellano and Bond (1991) use $s = 1$, whereas Hahn et al. (2007) employ $s = T-1$. Aggregating across all time periods and using a vector notation, a more compact notation is $\boldsymbol{D}_s \boldsymbol{y}_i = \alpha \boldsymbol{D}_s \boldsymbol{y}_{i(-1)} + \boldsymbol{D}_s \boldsymbol{\varepsilon}_i$, where $\boldsymbol{D}_s$ is the $(T-s) \times T$ sth difference-operator matrix ($\boldsymbol{D}_s = (\boldsymbol{I}_{T-s}, \boldsymbol{0}) - (\boldsymbol{0}, \boldsymbol{I}_{T-s})$), $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{y}_{i(-1)} = (y_{i0}, \ldots, y_{i(T-1)})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$.

Hahn et al. (2007) propose to estimate α in (1) by using the long $(T-1)$th difference technique of Griliches and Hausman (1986). The model (2) then becomes

$$\Delta^{T-1} y_{iT} = y_{iT} - y_{i1} = \alpha (y_{i(T-1)} - y_{i0}) + \varepsilon_{iT} - \varepsilon_{i1} = \alpha \Delta^{T-1} y_{i(T-1)} + \Delta^{T-1} \varepsilon_{iT} \tag{3}$$

for $i = 1, \ldots, n$. The long difference (LD) estimator itself is based on the following $T-1$ moment conditions, $T \geq 3$:

$$\mathrm{E}[y_{i0} \Delta^{T-1} \varepsilon_{iT}] = \mathrm{E}[y_{i0}(\varepsilon_{iT} - \varepsilon_{i1})] = 0, \tag{4a}$$

$$\mathrm{E}[u_{ir} \Delta^{T-1} \varepsilon_{iT}] = \mathrm{E}\left[u_{ir}(\varepsilon_{iT} - \varepsilon_{i1})\right] = 0 \qquad (r = 2, \ldots, T-1), \tag{4b}$$

where $u_{ir} = y_{ir} - \alpha y_{ir-1} = \eta_i + \varepsilon_{ir}$ (if $T = 2$, only moment condition (4a) makes sense and LD corresponds to the Arellano and Bond (1991) estimator). The operational moment conditions are then obtained by substituting for $\Delta^{T-1} \varepsilon_{iT}$ from (3). The moment conditions however contain also unobservable residuals u_{ir}. To produce moment conditions linear in α, a preliminary consistent estimator $\hat{\alpha}_n^0$ of α has to be used to compute and substitute estimates $\hat{u}_{ir} = y_{ir} - \hat{\alpha}_n^0 y_{ir-1}$ into (4b). Hahn et al. (2007) studied the GMM estimator based on the moment conditions (4) under the local-to-unity asymptotics, that is, assuming $\alpha_n \to 1$ as $n \to +\infty$, to confirm that these moment conditions do not rely on weak instruments in this limit case.

3 A class of long difference estimators

Apart for a more complicated asymptotic distribution of the LD estimator caused by estimating some instruments (see Section 4), an important disadvantage of the LD estimator is that, independently of the number T of time periods available, only a single observation per individual can be used after that data have been transformed by the long difference (3). This drawback is particularly problematic for data with a larger number T of time periods and a small or moderately large number n of individuals. As a remedy, we propose to extend the LD estimator by using multiple pairwise differences.

3.1 Pairwise-difference long-difference estimator

Let S denote the shortest difference considered in estimation, $2 \leq S \leq T-1$. To estimate α in (1), we propose to use the moment conditions of the LD estimator obtained by taking the longest difference Δ^{T-1}, the second longest differences Δ^{T-2}, and so on down to the differences Δ^{S}. This leads to $T^* = (T-S)(T-S+1)/2$ differenced equations consisting of all pairwise differences of lengths greater or equal to S:

$$s = T-1: \qquad \Delta^{T-1} y_{iT} = \alpha \Delta^{T-1} y_{i(T-1)} + \Delta^{T-1} \varepsilon_{iT}, \tag{5}$$

$$s = T-2: \qquad \Delta^{T-2} y_{iT} = \alpha \Delta^{T-2} y_{i(T-1)} + \Delta^{T-2} \varepsilon_{iT}, \tag{6}$$

$$\Delta^{T-2} y_{i(T-1)} = \alpha \Delta^{T-2} y_{i(T-2)} + \Delta^{T-2} \varepsilon_{i(T-1)}, \tag{7}$$

$$\vdots \qquad\qquad \vdots$$

$$s = S \quad : \qquad \Delta^{S} y_{iT} = \alpha \Delta^{S} y_{i(T-1)} + \Delta^{S} \varepsilon_{iT}, \tag{8}$$

$$\vdots$$

$$\Delta^{S} y_{i(S+1)} = \alpha \Delta^{S} y_{iS} + \Delta^{S} \varepsilon_{i(S+1)}. \tag{9}$$

Obviously, if $S = T - 1$, this system of differenced equations reduces to the original LD equation (3). In general, the shortest difference $S < T - 1$ should be chosen so that the number of equations $T^* \leq (T-1)$, which implies that $S > T - \sqrt{2T}$. If $T^* > (T-1)$, some of the moment equations implied by the model could be written as a linear combination of the other ones and would not contribute new information to the system (e.g., in the extreme case of $S = 1$, any sth difference equation could be written as a sum of the consecutive first-differenced equations). This observations is a special case of the equivalence statement in Arellano and Bover (1995).

Using the instruments (4) for each of the above stated equations, $s = S, \ldots, T-1$, leads to the set of the following moment conditions defining the infeasible pairwise-difference long-difference (PD-LD) estimator:

$$\mathrm{E}[y_{i(t-s-1)}\Delta^s \varepsilon_{it}] = \mathrm{E}[y_{i(t-s-1)}(\varepsilon_{it} - \varepsilon_{i(t-s)})] = 0, \tag{10}$$

$$\mathrm{E}[u_{i(t-1)}\Delta^s \varepsilon_{it}] = \mathrm{E}[u_{i(t-1)}(\varepsilon_{it} - \varepsilon_{i(t-s)})] = 0, \tag{11}$$

$$\vdots$$

$$\mathrm{E}[u_{i(t-s+1)}\Delta^s \varepsilon_{it}] = \mathrm{E}[u_{i(t-s+1)}(\varepsilon_{it} - \varepsilon_{i(t-s)})] = 0, \tag{12}$$

where $t = s+1, \ldots, T$ and $s = S, \ldots, T-1$.

To express the PD-LD estimator explicitly as a GMM estimator, let us first write the moment conditions (10)–(12) for a single equation in a more compact form as

$$\mathrm{E}(\boldsymbol{z}_{its}\Delta^s \varepsilon_{it}) = \mathbf{0} \qquad (t = s+1, \ldots, T; \quad s = S, \ldots, T-1), \tag{13}$$

where $\boldsymbol{z}_{its}$ is a $s \times 1$ vector $\boldsymbol{z}_{its} = (y_{i(t-s-1)}, u_{i(t-1)}, \ldots, u_{i(t-s+1)})'$. Furthermore, writing the equations (5)–(9) in the matrix form, the PD-LD estimator is based on the following differenced equations

$$\boldsymbol{D}\boldsymbol{y}_i = \alpha \boldsymbol{D}\boldsymbol{y}_{i(-1)} + \boldsymbol{D}\boldsymbol{\varepsilon}_i \qquad (i = 1, \ldots, n), \tag{14}$$

where $\boldsymbol{D}$ is a $T^* \times T$ partitioned matrix, $\boldsymbol{D} = (\boldsymbol{D}'_S, \dots, \boldsymbol{D}'_{T-1})'$. Hence, the complete set of the PD-LD moment conditions can be expressed in the matrix form as $\mathrm{E}[\boldsymbol{Z}'_i \boldsymbol{D} \boldsymbol{\varepsilon}_i] = 0$, where $\boldsymbol{Z}_i = \mathrm{diag}(\{\boldsymbol{z}'_{its}\}_{(t,s)\in\mathcal{T}})$, $\mathcal{T} = \{(t,s) : t = s+1, \dots, T; s = S, \dots, T-1\}$, denotes a block-diagonal matrix with T^* blocks $\boldsymbol{z}'_{its}$ indexed by $t = s+1, \dots, T$ and $s = S, \dots, T-1$.

As $\boldsymbol{z}_{its}$ in (13) is only partially observable, this PD-LD estimator is infeasible and a preliminary consistent estimator is needed to construct instruments. Let $\hat{\alpha}_n^0$ denote a preliminary consistent estimator of α (e.g., the Arellano-Bond estimator). The feasible instruments to be used in PD-LD are then $\hat{\boldsymbol{z}}_{its} = (y_{i(t-1-s)}, \hat{u}_{i(t-1)}, \dots, \hat{u}_{i(t+1-s)})'$, where $\hat{u}_{ir} = y_{ir} - \hat{\alpha}_n^0 y_{i(r-1)}$; the corresponding feasible matrix representation will be denoted $\hat{\boldsymbol{Z}}_i$. Denoting the inverse weight matrix $\hat{\boldsymbol{V}}_n$, the feasible PD-LD estimator – being a standard GMM estimator with linear moment conditions – can be then defined as follows:

$$\hat{\alpha}_n^{\text{PD-LD}} = \left(\boldsymbol{y}^{*\prime}_{-1} \hat{\boldsymbol{Z}} \hat{\boldsymbol{V}}_n^{-1} \hat{\boldsymbol{Z}}' \boldsymbol{y}^*_{-1}\right)^{-1} \boldsymbol{y}^{*\prime}_{-1} \hat{\boldsymbol{Z}} \hat{\boldsymbol{V}}_n^{-1} \hat{\boldsymbol{Z}}' \boldsymbol{y}^*, \tag{15}$$

where $\boldsymbol{y}^* = ([\boldsymbol{D}\boldsymbol{y}_1]', \dots, [\boldsymbol{D}\boldsymbol{y}_n]')'$ and $\boldsymbol{y}^*_{-1} = ([\boldsymbol{D}\boldsymbol{y}_{1(-1)}]', \dots, [\boldsymbol{D}\boldsymbol{y}_{n(-1)}]')'$ are the differenced variables and $\hat{\boldsymbol{Z}} = (\hat{\boldsymbol{Z}}'_1, \dots, \hat{\boldsymbol{Z}}'_n)'$ is an estimate of $\boldsymbol{Z} = (\boldsymbol{Z}'_1, \dots, \boldsymbol{Z}'_n)'$. In other words, $\hat{\boldsymbol{Z}}\boldsymbol{y}^* = \sum_{i=1}^n \hat{\boldsymbol{Z}}_i \boldsymbol{y}^*_i$ and $\hat{\boldsymbol{Z}}\boldsymbol{y}^*_{-1} = \sum_{i=1}^n \hat{\boldsymbol{Z}}_i \boldsymbol{y}^*_{i(-1)}$, where $\boldsymbol{y}^*_i = \boldsymbol{D}\boldsymbol{y}_i$ and $\boldsymbol{y}^*_{i(-1)} = \boldsymbol{D}\boldsymbol{y}_{i(-1)}$.

By using $\hat{\alpha}_n^{\text{PD-LD}}$ to re-estimate $\hat{\boldsymbol{z}}_{its}$, one can iterate to another LD estimator, which will be referred to as PD-LD1. Eventually, the procedure can be further iterated, yielding PD-LD2, PD-LD3, and so on.

3.2 Mixed-distance long-difference estimator

Loosely speaking, the idea of taking the longest differences is based on the fact that moment conditions based on such a data transformation do not become weak when α approaches one (Hahn et al., 2007). Considering its pairwise-difference extensions, there are other alternative choices of $T-1$ differenced equations than just (5)–(9). For instance, one could make use of all possible pairwise differences from the shortest one $S = 2$ to the longest one $T-1$ and take only one equation for each $s, S \leq s \leq T-1$, in order to fulfill the condition

Table 1: Asymptotic bias and variance

Estimator	Limited # of instruments	Unlimited # of instruments
AB	$O(n^{-1}T^{-1})$	$O(n^{-1})$
LD	$O(n^{-1}\|\alpha\|^{T})$	$O(n^{-1}T\|\alpha\|^{T})$
MD-LD	$O(n^{-1}T^{-1})$	$O(n^{-1}T^{-1})$
PD-LD	$O(n^{-1}\|\alpha\|^{T-\sqrt{2T}})$	$O(n^{-1}T\|\alpha\|^{T-\sqrt{2T}})$

Note: AB refers to the Arellano-Bond estimator with the model in forward orthogonal deviations derived by Bun and Kiviet (2006).

that the number of employed equations $T^* \leq T-1$. As a reference example, let $\hat{\alpha}_n^{\text{MD-LD}}$ be the GMM estimator based on the following $(T-2)(T-1)/2$ moment conditions:

$$\mathrm{E}(\boldsymbol{z}_{its}\Delta^s\varepsilon_{it}) = \boldsymbol{0} \qquad (t = s+1; \quad s = 2, \ldots, T-1). \tag{16}$$

This estimator will be referred to as the mixed-difference long-difference estimator (MD-LD) as it relies both on short and long differences. It will be shown that including shorter differences affects unfavourably the bias of the estimator, at least for a large T.

3.3 Finite sample bias

To compare the LD, PD-LD, and MD-LD estimators, we first derive the leading terms of their finite sample biases. For the sake of simplicity, we compare the methods in the infeasible setting. In this section, we also use the two-stage least squares weight matrix for all estimators (e.g., the inverse weight matrix for the PD-LD in (15) will be $\hat{\boldsymbol{V}}_n = \boldsymbol{V}_n = \sum_i \boldsymbol{Z}_i'\boldsymbol{Z}_i$ instead of a general one), which happens to be the optimal weight matrix for the infeasible LD estimator. Other (asymptotic) properties are studied under more general assumptions in Section 4.

To derive the biases of the long-difference estimators, we need to impose the following conditions (using one high-level assumption for simplicity):

B.1 For all $i \in \mathbb{N}$ and $t \in \mathbb{N}$, idiosyncratic shocks ε_{it} are mutually independent, have finite second moments, and $\mathrm{E}(\varepsilon_{it}|y_{i(t-1)}, \ldots, y_{i0}, \eta_i) = 0$ and $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$.

B.2 Let individual effects η_i be independently distributed across individuals with finite second moments.

B.3 Denoting $q_{nT^*} = \boldsymbol{y}^{*\prime}_{-1}\boldsymbol{Z}\boldsymbol{V}_n^{-1}\boldsymbol{Z}'\boldsymbol{y}^*_{-1}$, let p-lim $q_{nT^*}/(nT^*) = \bar{q} > 0$ for $nT^* \to \infty$, where $T^* \leq T-1$; in particular, $T^* = 1$ in the case of the LD estimator.

Theorem 1. *Let y_{it} be generated by* (1) *with $0 < |\alpha| < 1$ and let $\hat{\alpha}_{nT}^{LD}$, $\hat{\alpha}_{nT}^{PD\text{-}LD}$, and $\hat{\alpha}_{nT}^{MD\text{-}LD}$ be the infeasible two-stage least squares estimators based on moment conditions* (4), (13), *and* (16), *respectively. Additionally, suppose that Assumptions B.1–B.3 hold. When all possible instruments are included, the finite-sample biases of each estimator in the LD class are given by*

$$B_{LD} = O((nT^*_{LD})^{-1}) \cdot \left(-\frac{\sigma_\varepsilon^2}{\alpha^2}(T-1)\alpha^T\right) = O(n^{-1}T|\alpha|^T), \tag{17}$$

$$B_{MD\text{-}LD} = O((nT^*_{MD})^{-1}) \cdot \left[-\frac{\sigma_\varepsilon^2}{\alpha}\left(\frac{\alpha^2-\alpha^T}{(1-\alpha)^2} - \frac{(T-1)\alpha^T-\alpha^2}{1-\alpha}\right)\right] = O(n^{-1}T^{-1}), \tag{18}$$

$$\begin{aligned} B_{PD\text{-}LD} = O((nT^*_{PD})^{-1}) \cdot \Bigg\{ -\frac{\sigma_\varepsilon^2}{\alpha}\Bigg[T\left(\frac{\alpha^S-\alpha^T}{(1-\alpha)^2} + \frac{(S-1)\alpha^S-(T-1)\alpha^T}{1-\alpha}\right) \\ - 2\frac{\alpha^S-\alpha^T}{(1-\alpha)^3} + \frac{[2T-3]\alpha^T-[2S-3]\alpha^S}{(1-\alpha)^2} + \frac{(T-1)^2\alpha^T-(S-1)^2\alpha^S}{1-\alpha}\Bigg]\Bigg\} \\ = O(n^{-1}T|\alpha|^{T-\sqrt{2T}}), \end{aligned} \tag{19}$$

where the leading terms in bounds $O(\cdot)$ are determined for $n \to \infty$ or $T \to \infty$ and $T^*_{LD} = 1$ in the case of LD, $T^*_{MD} = T-2$ in the case of ML-LD, and $S = \lceil T - \sqrt{2T} \rceil$ and $T^*_{PD} = (T-S)(T-S+1)/2$ in the case of PD-LD.

Similarly, when the number of instruments used for each moment equation is limited to *be at most $\bar{m} \in \mathbb{N}$, the finite-sample biases are bounded by*

$$|B_{LD}| \leq O((nT^*_{LD})^{-1}) \cdot \sigma^2_\varepsilon \bar{m} \alpha^{T-2} = O(n^{-1}|\alpha|^T), \tag{20}$$

$$|B_{MD-LD}| \leq O((nT^*_{MD})^{-1}) \cdot \frac{\sigma^2_\varepsilon \bar{m}}{|\alpha|} \cdot \frac{|\alpha|^2 - |\alpha|^T}{1-|\alpha|} = O(n^{-1}T^{-1}), \tag{21}$$

$$|B_{PD-LD}| \leq O((nT^*_{PD})^{-1}) \times \tag{22}$$

$$\times \frac{\sigma^2_\varepsilon \bar{m}}{|\alpha|} \left(T \frac{|\alpha|^S - |\alpha|^T}{1-|\alpha|} + \frac{|\alpha|^S - |\alpha|^T}{(1-|\alpha|)^2} + \frac{|(S-1)|\alpha|^S - (T-1)|\alpha|^T|}{1-|\alpha|} \right)$$

$$= O(n^{-1}|\alpha|^{T-\sqrt{2T}}). \tag{23}$$

Results concerning the leading terms are summarized in Table 1 (the order of bias of the Arellano-Bond estimator as derived in Bun and Kiviet (2006) is also reported). In general, the orders of biases are smaller when the number of employed instruments is limited, but the ranking of methods is not affected by the number of instruments. Taking into account (18), the infeasible LD and PD-LD methods exhibit the smallest biases, especially if T is large or α is small. The relatively small increase in bias of PD-LD relative to LD is substantially compensated by the fact that PD-LD uses $nT^* \approx n(T-1)$ observations compared to n observations used by LD (see also Section 4), which will complement the bias properties of PD-LD by a smaller variance of estimates compared to LD.

4 Asymptotic distribution

4.1 Asymptotic normality

In this section, the asymptotic distribution for the class of the long-difference estimators is derived. Although the asymptotic distribution of the LD estimator is derived in Hahn et al. (2007), its given there only for the limit case of $\alpha \to 1$, without any exogenous variables, and in a form difficult for practical use.

For deriving the asymptotic distribution of different LD estimators, it is useful to generalize and derive this result for a model with exogenous variables. Let $\boldsymbol{x}_{it} = (x_{it1}, \ldots, x_{itK})'$

be a set of K exogenous or predetermined variables. Assume $T \geq 3$ is fixed and y_{it} follows

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \eta_i + \varepsilon_{it}, \\ &= \boldsymbol{w}_{it}'\boldsymbol{\theta} + \eta_i + \varepsilon_{it}, \end{aligned} \qquad (t = 1, \ldots, T; \quad i = 1, \ldots, n), \tag{24}$$

where $\boldsymbol{w}_{it} = (w_{itk})_{k=1}^{K+1} = (y_{i(t-1)}, \boldsymbol{x}_{it}')'$ and the parameter of interest is $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}')'$ with the true value $\boldsymbol{\theta}^0$. Let $\boldsymbol{W}_i$ denote the $T \times (K+1)$ matrix $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iT})'$. The assumptions concerning the data-generating process (24), which is allowed to be heterogeneous across individuals i, follow.

A.1 Let $\{\boldsymbol{y}_i, \boldsymbol{W}_i, \eta_i\}_{i=1}^n$ be a sequence of independently distributed random vectors with uniformly bounded finite $(2+\delta)$th moments for some $\delta > 0$.

A.2 For all i and t, $\mathrm{E}(\varepsilon_{it} | \boldsymbol{w}_{it}, \ldots, \boldsymbol{w}_{i1}, \eta_i) = 0$.

Next, the initial estimator $\hat{\boldsymbol{\theta}}_n^0$ will be assumed to be a consistent GMM estimator of $\boldsymbol{\theta}^0$ based on moment conditions $\mathrm{E}[\boldsymbol{\psi}(\boldsymbol{y}_i, \boldsymbol{W}_i, \boldsymbol{\theta})] = \mathrm{E}[\boldsymbol{\psi}_i(\boldsymbol{\theta})] = \mathbf{0}$, where $\boldsymbol{\psi}$ is a $F \times 1$ vector of functions, $F \geq K+1$. The sample counterpart of these moment conditions will be denoted $\boldsymbol{f}_n(\boldsymbol{\theta}) = 1/n \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta})$. We thus assume that

A.3 Estimator $\hat{\boldsymbol{\theta}}_n^0$ is $\sqrt{n}$-consistent and asymptotically normal for a fixed T and $n \to \infty$; in particular, $\hat{\boldsymbol{\theta}}_n^0 \xrightarrow{p} \boldsymbol{\theta}^0$ in probability and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0) = O_p(1)$.

The instruments used in this class of the LD estimators can be generically denoted as $\hat{\boldsymbol{z}}_{its} = \boldsymbol{z}_{its} - \boldsymbol{\mathcal{W}}_{its}(\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta})$, where explanatory variables $\boldsymbol{\mathcal{W}}_{its} = (\mathbf{0}, \boldsymbol{w}_{it}, \cdots, \boldsymbol{w}_{i(t+1-s)})'$ with $\mathbf{0}$ being an appropriately sized matrix of zeros, the initial estimator is assumed to be asymptotically linear in its moment conditions,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0) = \boldsymbol{\Lambda} \cdot \sqrt{n} \boldsymbol{f}_n(\boldsymbol{\theta}^0) + o_p(1) = \boldsymbol{\Lambda} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}^0) + o_p(1), \tag{25}$$

and $\boldsymbol{\Lambda}$ is the result of the stochastic expansion of the initial estimator (see for example Arellano, 2003, p. 187).[1]

[1] Suppose the Arellano-Bond estimator is chosen as preliminary estimator. Let $\boldsymbol{Z}_i^{\mathrm{AB}}$ and $\boldsymbol{A}_n$ denote the

Next, let $\mathrm{E}[\boldsymbol{\tau}_i(\boldsymbol{\theta})] = \mathbf{0}$ be a general expression for the R moment conditions implied by the method, $R \geq K + 1$: after substituting for $\Delta^s \varepsilon_{it}$ from model (24), it consists of (4) for LD, (13) for PD-LD, or (16) for MD-LD, respectively, and additionally, of moment conditions implied by $\boldsymbol{x}$'s variables (in general, these depend on whether each x_{itk} is weakly or strictly exogenous or predetermined). Further, the combined vector of moment conditions for the initial and chosen LD-type estimators will be denoted $\boldsymbol{\rho}_i(\boldsymbol{\theta}) = (\boldsymbol{\tau}_i(\boldsymbol{\theta})', \boldsymbol{\psi}_i(\boldsymbol{\theta})')'$. By Assumption A.1, $\boldsymbol{\rho}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\rho}_n(\boldsymbol{\theta})$ are n independent random vectors. We however have to impose additional assumptions, again taking into account the individual heterogeneity.

A.4 The moment conditions $\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)$ at $\boldsymbol{\theta}^0$ have uniformly bounded finite $(2+\delta)$th moments for some $\delta > 0$. Moreover, $E[\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)] = 0$ and the variance matrix $\boldsymbol{\Sigma} = \lim_{n\to\infty} \sum_{i=1}^n \mathrm{var}[\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)]/n$ exists and is a finite positive definite matrix.

A.5 (a) Let $\boldsymbol{\omega}_{tsk} = \lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}(\boldsymbol{z}_{its} \Delta^s w_{itk})/n$ exist and be finite for all t, s, and k, and additionally, let $\boldsymbol{\omega}_k = (\boldsymbol{\omega}'_{(S+1)Sk}, \ldots, \boldsymbol{\omega}'_{T(T-1)k})'$ be the kth column of the full-rank matrix $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{K+1})$.

(b) Similarly, let $\boldsymbol{P}_{ts} = \lim_{n\to\infty} \sum_{i=1}^n \mathrm{E}(\boldsymbol{\mathcal{W}}_{its} \Delta^s \varepsilon_{it})/n$ exist and be finite for all t and s and let $\boldsymbol{P} = (\boldsymbol{P}'_{(S+1)S}, \ldots, \boldsymbol{P}'_{T(T-1)})'$ have a full rank.

(c) Matrix $\boldsymbol{\Lambda}$ has a full rank.

(d) Finally, $\sum_{i=1}^n \mathrm{E}(\boldsymbol{\mathcal{W}}_{its} \Delta^s w_{itk})/n$ is assumed to exist and to be uniformly bounded in $n \in \mathbb{N}$ for all s, t, and k.

A.6 Let $\hat{\boldsymbol{V}}_n$ be a $\dim(\boldsymbol{\tau}) \times \dim(\boldsymbol{\tau})$ inverse weight matrix such that $\hat{\boldsymbol{V}}_n \xrightarrow{p} \boldsymbol{V}$ as $n \to \infty$, where $\boldsymbol{V}$ is a positive definite matrix.

If the standard, but stronger assumption that random variables in Assumption A.1 are independent and identically distributed is used, the above mentioned assumptions would

corresponding matrix of instruments and the weight matrix, respectively. Then, the matrix $\boldsymbol{\Lambda}$ will be the probability limit of

$$\boldsymbol{\Lambda}_n = -n \left[\sum_{i=1}^n (\boldsymbol{Z}_i^{\mathrm{AB}\prime} \boldsymbol{D}_1 \boldsymbol{W}_i)' \boldsymbol{A}_n \sum_{i=1}^n (\boldsymbol{Z}_i^{\mathrm{AB}\prime} \boldsymbol{D}_1 \boldsymbol{W}_i) \right]^{-1} \sum_{i=1}^n (\boldsymbol{Z}_i^{\mathrm{AB}\prime} \boldsymbol{D}_1 \boldsymbol{W}_i)' \boldsymbol{A}_n,$$

where $\boldsymbol{D}_1$ is the $(T-1) \times T$ first difference-operator matrix.

simplify: for example, the moment conditions $\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)$ would have to possess only finite second moments and their variance matrix would be defined simply as $\boldsymbol{\Sigma} = \text{var}[\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)]$.

Under the above stated assumptions, the asymptotic distribution of the feasible LD, MD-LD, and PD-LD estimators can be derived.

Theorem 2. *Suppose that Assumptions A.1–A.6 hold. Then for a fixed T and $n \to \infty$, $\hat{\boldsymbol{\theta}}_n$ is consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{d} \text{N}(\mathbf{0}, \boldsymbol{\Xi}), \tag{26}$$

where $\boldsymbol{\Xi} = (\boldsymbol{\Omega}'\boldsymbol{V}^{-1}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}'\boldsymbol{V}^{-1}\boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{\Omega}(\boldsymbol{\Omega}'\boldsymbol{V}^{-1}\boldsymbol{\Omega})^{-1}$ and $\boldsymbol{M} = (\boldsymbol{I}_R, -\boldsymbol{P}\boldsymbol{\Lambda})$.

4.2 Estimating the asymptotic variance

According to the standard GMM theory, an optimal choice of the inverse weight matrix $\boldsymbol{V}_n$ is a consistent estimate of the covariance matrix of the moment conditions $\boldsymbol{\Sigma}$. Assuming for simplicity that data are independent and identically distributed across individuals, this covariance matrix can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_\tau & \boldsymbol{\Sigma}_{\tau\psi} \\ \boldsymbol{\Sigma}_{\psi\tau} & \boldsymbol{\Sigma}_\psi \end{pmatrix}, \tag{27}$$

where $\boldsymbol{\Sigma}_\tau = \text{var}[\boldsymbol{\tau}_i(\boldsymbol{\theta}^0)]$, $\boldsymbol{\Sigma}_\psi = \text{var}[\boldsymbol{\psi}_i(\boldsymbol{\theta}^0)]$, and $\boldsymbol{\Sigma}_{\tau\psi} = \text{cov}[\boldsymbol{\tau}_i(\boldsymbol{\theta}^0), \boldsymbol{\psi}_i(\boldsymbol{\theta}^0)]$ (recall that $\boldsymbol{\tau}$ and $\boldsymbol{\psi}$ refer to the moment conditions of the (PD-)LD and initial estimators, respectively). Since the instruments are estimated rather than given, Theorem 2 implies that $\boldsymbol{V}_{\text{opt}}$ will be equal to

$$\boldsymbol{V}_{\text{opt}} = \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}' = \boldsymbol{\Sigma}_\tau - \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\psi\tau} - (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\psi\tau})' + \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_\psi\boldsymbol{\Lambda}'\boldsymbol{P}'. \tag{28}$$

Considering the part $\boldsymbol{\Sigma}_\tau$, which corresponds to the variance of the moment conditions of the infeasible estimator, note that, because of the complex structure of PD-LD, the covariance matrix $\boldsymbol{\Sigma}_\tau = \mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{D}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{D}'\boldsymbol{Z}_i)$ may be singular. In other words, for a sufficiently large T and number of included instruments in PD-LD, some moment conditions are redundant and $\boldsymbol{\Sigma}_\tau$ is not invertible.

To overcome this problem in computing $\boldsymbol{V}_{\mathrm{opt}}$, several solutions are available. First, one could try to keep all the moment conditions corresponding to $\boldsymbol{\Sigma}_\tau$. This however requires dealing with many linearly dependent moment conditions, which would have to be done as in Carrasco and Florens (2000), for instance. A simple alternative solution – also used in this paper – is to limit the number of instruments in $\boldsymbol{\tau}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i'\boldsymbol{D}(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\theta})$, which equals $\boldsymbol{\tau}_i(\boldsymbol{\theta}^0) = \boldsymbol{Z}_i'\boldsymbol{D}\boldsymbol{\varepsilon}_i$ at $\boldsymbol{\theta}^0$.[2] Denoting $\boldsymbol{\tau}_i^\dagger(\boldsymbol{\theta})$ the vector of moment conditions corresponding to selected instruments and $\boldsymbol{Z}_i^\dagger$ the corresponding matrix of instruments, the optimal inverse weight matrix will be a consistent estimate $\hat{\boldsymbol{V}}_{n\Upsilon}$ of (28), where

$$\hat{\boldsymbol{\Sigma}}_{n\boldsymbol{\tau}} = \sum_{i=1}^{n} \hat{\boldsymbol{Z}}_i^{\dagger\prime}\hat{\boldsymbol{\Upsilon}}_i\hat{\boldsymbol{Z}}_i^\dagger \tag{31}$$

with $\hat{\boldsymbol{\Upsilon}}_i = \boldsymbol{D}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\boldsymbol{D}'$ and $\hat{\boldsymbol{Z}}_i^\dagger$ being computed by using a preliminary consistent estimator ($\boldsymbol{\Lambda}$, $\boldsymbol{P}$, and other terms in (28) can be estimated by the respective sample averages).

For several reasons, we do not pay more attention to the estimation of the optimal weights $\boldsymbol{V}_{\mathrm{opt}}$. It is well known that a part of the bias of GMM estimators stems from a poorly estimated weight matrix (Newey and Smith, 2004). For either small values of n or large number of instruments (which depends on T when all instruments are included), weights in $\hat{\boldsymbol{V}}_{n\Upsilon}$ may be highly imprecise. A simple alternative to (31) is to employ the

[2] Clearly, there are more ways to do so. To prevent the linear dependence of the PD-LD moment conditions and instruments thereof are selected here in the following way:

$$\mathrm{E}(y_{i(t-s-1)}\Delta^s\varepsilon_{it}) = 0 \qquad (t = s+1,\dots,T; \quad s = S,\dots,T-1) \tag{29}$$

and for all $t = s+1,\dots,T$, $s = S,\dots,T-1$:

$$\begin{cases} \mathrm{E}(u_{i(t-1)}\Delta^s\varepsilon_{it}) = 0 & \text{if } s > S; \\ \mathrm{E}\left[\begin{pmatrix} u_{i(t-1)} \\ \vdots \\ u_{i(t-s+1)} \end{pmatrix} \Delta^s\varepsilon_{it}\right] = \mathbf{0} & \text{if } s = S. \end{cases} \tag{30}$$

.

weights of the standard two-stage least squares and use instead

$$\hat{\boldsymbol{V}}_{nI} = \sum_{i=1}^{n} \hat{\boldsymbol{Z}}_i' \boldsymbol{I} \hat{\boldsymbol{Z}}_i. \tag{32}$$

There are a couple of advantages of this weighting matrix $\hat{\boldsymbol{V}}_{nI}$: (i) it can be computed directly based on the initial estimate, (ii) it does not impose constrains on the number of instruments (the full proposed matrix $\boldsymbol{Z}_i$ can be used), and finally, (iii) finite sample results for weighting matrix (32) are rather close to or even better than the ones for weighting matrix (31), especially as the sample size increases. See Section 5 for more details.

5 Monte Carlo simulation

5.1 Design

In this section, the finite sample performance of the proposed estimators is evaluated by Monte Carlo simulations. The data-generating proces for y_{it} follows model (1) with $\alpha = 0.1, 0.5, 0.9$, $n = 25, 50, 100, 400, 1600, 3200$, $T = 6, 12, 24$, $\eta_i \sim \mathrm{N}(0, \sigma_\eta^2)$, and $\varepsilon_{it} \sim \mathrm{N}(0, 1)$. In order to measure the sensitivity of the estimators to the stationarity assumption, the initial observations at time $t = 0$ are generated by

$$y_{i0} \sim \mathrm{N}\left(\frac{\eta_i}{1-\alpha_J}, \frac{\sigma_\varepsilon^2}{1-\alpha^2}\right), \tag{33}$$

which leads to mean-stationary series y_{it} if $\alpha_J = \alpha$ and to non-stationary sequences if $\alpha_J \neq \alpha$. Each model is evaluated using 1000 replications.

Results are reported for the LD estimator and for the proposed estimators MD-LD, PD-LD, and PD-LD1, where the last one denotes the iterated PD-LD estimator based on PD-LD used as the preliminary estimator. The Arellano and Bond (AB, 1991) twostep GMM estimator[3] and the system Blundell and Bond (BB, 1998) estimator[4] are also reported, serving as reference estimators as well as preliminary estimators for LD, MD-LD, and PD-LD. All methods are compared by means of the root mean squared errors (RMSE) unless stated otherwise.

5.2 Weight matrix

Before presenting a full comparison of estimators, we will briefly revisit the choice of the GMM weight matrix. As mentioned in Section 4.2, the finite sample performance of a GMM estimator can be heavily affected by the choice of the weight matrix. The difference between using weights (31) and (32) is documented in Table 2 for various models with $\sigma_\eta^2 = 1$. Let PD-LD-I and PD-LD1-I denote the PD-LD estimators when the inverse weight matrix (32) is used and let PD-LD-$\hat{\Upsilon}$ and PD-LD1-$\hat{\Upsilon}$ denote PD-LD when weights (31) are in use.

As shown in Table 2, PD-LD-$\hat{\Upsilon}$ seems to perform only slightly better than PD-LD-I and only for small values of n (the main exception is the case of $n = 25$, $T = 6$, and $\alpha = 0.9$). More specifically, PD-LD-$\hat{\Upsilon}$ can perform slightly better than PD-LD-I if the initial estimator is reliable, but PD-LD-$\hat{\Upsilon}$ can perform much worse than PD-LD-I if the initial estimator is imprecise. Consequently, it seems that using weights (32) is a more robust strategy, which – in the cases when it is worse than PD-LD-$\hat{\Upsilon}$ – matches the optimally weighted alternative once the sample size is sufficiently large. We therefore recommend and use in further simulations the PD-LD estimator based on the weighting matrix $\hat{\boldsymbol{V}}_{nI}$ defined in (32).

[3]The (optimal) inverse weight matrix is $\sum_i \boldsymbol{Z}_i^{\mathrm{AB}\prime} \boldsymbol{H} \boldsymbol{Z}_i^{\mathrm{AB}}$, where $\boldsymbol{Z}_i^{\mathrm{AB}}$ is the matrix of instruments and $\boldsymbol{H}$ is a $(T-1) \times (T-1)$ tridiagonal matrix with 2 on the main diagonal, -1 on the first two sub-diagonals, and zeros elsewhere (see Arellano and Bond, 1991, p. 279).

[4]The inverse weight matrix is $\sum_i \boldsymbol{Z}_i^{\mathrm{BB}\prime} \boldsymbol{G} \boldsymbol{Z}_i^{\mathrm{BB}}$, where $\boldsymbol{Z}_i^{\mathrm{BB}}$ is the matrix of instruments and $\boldsymbol{G}$ is a partitioned matrix, $\boldsymbol{G} = \mathrm{diag}(\boldsymbol{H}, \boldsymbol{I})$, where $\boldsymbol{H}$ is as in Arellano-Bond and $\boldsymbol{I}$ is the identity matrix (see Kiviet, 2007, Eq. (38)).

Table 2: The root mean squared errors of the PD-LD estimators using the two-stage least-squares weighting matrix and the asymptotically optimal weighting matrix. The three sections of the table represent results for $\alpha = 0.1, 0.5$, and 0.9.

n	25	100	400	1600	3200	25	100	25	100
T			6			12		24	
				$\alpha = 0.1$					
AB*	0.141	0.071	0.035	0.017	0.012	0.092	0.041	0.063	0.028
PD-LD	0.138	0.067	0.034	0.016	0.012	0.095	0.047	0.076	0.039
PD-LD-V	0.135	0.063	0.032	0.016	0.011	0.095	0.042	0.076	0.039
PD-LD1	0.141	0.067	0.034	0.017	0.012	0.095	0.047	0.076	0.039
PD-LD1-V	0.157	0.076	0.038	0.018	0.013	0.095	0.053	0.076	0.039
BB*	0.127	0.067	0.035	0.017	0.013	0.100	0.046	0.157	0.052
PD-LD	0.138	0.065	0.033	0.017	0.012	0.093	0.048	0.075	0.038
PD-LD-V	0.127	0.062	0.031	0.016	0.012	0.093	0.046	0.075	0.038
PD-LD1	0.141	0.066	0.033	0.017	0.012	0.093	0.048	0.075	0.038
PD-LD1-V	0.153	0.074	0.037	0.018	0.013	0.093	0.057	0.075	0.038
				$\alpha = 0.5$					
AB*	0.232	0.108	0.051	0.025	0.018	0.133	0.055	0.080	0.033
PD-LD	0.136	0.072	0.035	0.018	0.013	0.090	0.045	0.069	0.037
PD-LD-V	0.187	0.076	0.034	0.017	0.012	0.090	0.047	0.069	0.037
PD-LD1	0.155	0.082	0.040	0.020	0.014	0.092	0.046	0.069	0.037
PD-LD1-V	0.183	0.084	0.040	0.020	0.014	0.092	0.053	0.069	0.037
BB*	0.139	0.081	0.044	0.021	0.016	0.118	0.058	0.186	0.074
PD-LD	0.135	0.069	0.034	0.017	0.012	0.087	0.045	0.070	0.035
PD-LD-V	0.131	0.067	0.033	0.017	0.012	0.087	0.049	0.070	0.035
PD-LD1	0.157	0.080	0.039	0.019	0.014	0.088	0.046	0.070	0.035
PD-LD1-V	0.152	0.078	0.039	0.020	0.014	0.088	0.054	0.070	0.035
				$\alpha = 0.9$					
AB*	0.570	0.444	0.241	0.102	0.069	0.292	0.202	0.146	0.089
PD-LD	0.202	0.160	0.120	0.072	0.050	0.093	0.064	0.049	0.026
PD-LD-V	0.439	0.268	0.128	0.065	0.046	0.093	0.155	0.049	0.026
PD-LD1	0.186	0.127	0.097	0.072	0.044	0.089	0.050	0.049	0.026
PD-LD1-V	0.389	0.205	0.099	0.063	0.043	0.089	0.129	0.049	0.026
BB*	0.082	0.073	0.051	0.030	0.021	0.054	0.048	0.043	0.029
PD-LD	0.124	0.085	0.052	0.028	0.019	0.067	0.042	0.047	0.023
PD-LD-V	0.109	0.081	0.053	0.029	0.020	0.067	0.043	0.047	0.023
PD-LD1	0.207	0.129	0.067	0.033	0.022	0.091	0.052	0.050	0.027
PD-LD1-V	0.161	0.111	0.063	0.032	0.021	0.091	0.046	0.050	0.027

Note: The symbol '*' denotes the preliminary estimator for PD-LD.

Table 3: The root mean squared errors of all estimator for different sample sizes using $\sigma_\eta^2 = 1$. The three sections of the table represent results for $\alpha = 0.1, 0.5$, and 0.9.

n	25			50			100		
T	6	12	24	6	12	24	6	12	24
				$\alpha = 0.1$					
AB*	0.143	0.089	0.064	0.097	0.058	0.041	0.070	0.041	0.027
LD	0.206	0.209	0.206	0.134	0.148	0.141	0.099	0.103	0.100
MD-LD	0.145	0.120	0.110	0.100	0.081	0.078	0.072	0.059	0.052
PD-LD	0.136	0.094	0.074	0.093	0.067	0.053	0.070	0.047	0.037
PD-LD1	0.139	0.094	0.074	0.094	0.067	0.053	0.071	0.047	0.037
BB*	0.129	0.102	0.157	0.095	0.067	0.091	0.068	0.046	0.050
LD	0.206	0.209	0.206	0.134	0.148	0.141	0.099	0.103	0.100
MD-LD	0.145	0.119	0.110	0.099	0.081	0.078	0.071	0.059	0.052
PD-LD	0.136	0.094	0.074	0.093	0.067	0.053	0.070	0.047	0.037
PD-LD1	0.139	0.094	0.074	0.094	0.067	0.053	0.071	0.047	0.037
				$\alpha = 0.5$					
AB*	0.231	0.129	0.083	0.152	0.083	0.053	0.107	0.053	0.033
LD	0.188	0.184	0.174	0.124	0.127	0.121	0.091	0.087	0.086
MD-LD	0.155	0.112	0.099	0.112	0.076	0.069	0.079	0.054	0.046
PD-LD	0.138	0.088	0.070	0.098	0.064	0.049	0.072	0.043	0.035
PD-LD1	0.160	0.090	0.070	0.109	0.065	0.049	0.083	0.044	0.035
BB*	0.148	0.119	0.184	0.113	0.084	0.127	0.082	0.057	0.074
LD	0.191	0.184	0.174	0.123	0.127	0.121	0.091	0.087	0.086
MD-LD	0.154	0.112	0.099	0.107	0.076	0.069	0.076	0.054	0.046
PD-LD	0.141	0.088	0.070	0.095	0.064	0.049	0.072	0.043	0.035
PD-LD1	0.165	0.090	0.070	0.109	0.065	0.049	0.084	0.044	0.035
				$\alpha = 0.9$					
AB*	0.579	0.296	0.146	0.516	0.256	0.122	0.447	0.201	0.089
LD	0.197	0.119	0.096	0.169	0.088	0.066	0.146	0.066	0.045
MD-LD	0.233	0.125	0.073	0.207	0.099	0.056	0.182	0.082	0.040
PD-LD	0.201	0.093	0.047	0.180	0.076	0.035	0.158	0.063	0.026
PD-LD1	0.189	0.089	0.048	0.152	0.068	0.037	0.124	0.051	0.026
BB*	0.086	0.053	0.041	0.079	0.052	0.036	0.070	0.047	0.029
LD	0.171	0.114	0.096	0.120	0.084	0.065	0.095	0.059	0.045
MD-LD	0.142	0.085	0.065	0.108	0.065	0.047	0.085	0.053	0.034
PD-LD	0.130	0.064	0.045	0.094	0.052	0.032	0.080	0.041	0.022
PD-LD1	0.212	0.088	0.048	0.154	0.073	0.037	0.121	0.052	0.026

Note: The symbol '*' denotes the preliminary estimator for LD, MD-LD and PD-LD.

Table 4: The root mean squared errors of all estimator for low and high values of the variance of the individual effects, $\sigma_\eta^2 = 0.25, 1, 4$. The three sections of the table represent results for $\alpha = 0.1, 0.5$, and 0.9.

(n, T)	$(100, 6)$			$(50, 12)$			$(25, 24)$		
$\sigma_\eta^2/\sigma_\varepsilon^2$	1/4	1	4	1/4	1	4	1/4	1	4
				$\alpha = 0.1$					
AB*	0.062	0.070	0.076	0.053	0.062	0.062	0.063	0.062	0.064
LD	0.097	0.104	0.101	0.145	0.148	0.144	0.213	0.208	0.211
MD-LD	0.070	0.071	0.069	0.082	0.084	0.082	0.106	0.109	0.107
PD-LD	0.066	0.069	0.068	0.066	0.066	0.066	0.070	0.075	0.074
PD-LD1	0.066	0.070	0.069	0.066	0.066	0.066	0.070	0.075	0.074
BB*	0.060	0.068	0.081	0.074	0.069	0.071	0.176	0.159	0.091
LD	0.100	0.102	0.100	0.145	0.141	0.138	0.201	0.200	0.203
MD-LD	0.069	0.069	0.070	0.084	0.080	0.083	0.104	0.107	0.108
PD-LD	0.067	0.067	0.065	0.065	0.065	0.065	0.073	0.075	0.075
PD-LD1	0.068	0.068	0.066	0.065	0.065	0.065	0.073	0.075	0.075
				$\alpha = 0.5$					
AB*	0.086	0.108	0.122	0.071	0.087	0.092	0.080	0.082	0.085
LD	0.089	0.093	0.091	0.127	0.128	0.123	0.183	0.175	0.178
MD-LD	0.079	0.079	0.077	0.079	0.078	0.078	0.094	0.095	0.096
PD-LD	0.069	0.070	0.069	0.063	0.061	0.062	0.067	0.071	0.070
PD-LD1	0.078	0.078	0.080	0.064	0.062	0.064	0.067	0.071	0.070
BB*	0.076	0.083	0.118	0.107	0.087	0.104	0.236	0.187	0.083
LD	0.094	0.092	0.098	0.128	0.123	0.120	0.175	0.176	0.175
MD-LD	0.077	0.077	0.089	0.080	0.075	0.081	0.095	0.098	0.096
PD-LD	0.070	0.069	0.077	0.060	0.061	0.061	0.069	0.071	0.071
PD-LD1	0.081	0.080	0.083	0.062	0.062	0.062	0.069	0.071	0.071
				$\alpha = 0.9$					
AB*	0.345	0.470	0.494	0.221	0.257	0.270	0.139	0.147	0.148
LD	0.139	0.151	0.152	0.088	0.089	0.086	0.092	0.094	0.098
MD-LD	0.165	0.186	0.192	0.098	0.101	0.103	0.072	0.074	0.075
PD-LD	0.146	0.163	0.166	0.076	0.076	0.076	0.045	0.048	0.048
PD-LD1	0.130	0.146	0.129	0.068	0.067	0.067	0.046	0.049	0.049
BB*	0.076	0.071	0.088	0.086	0.051	0.081	0.154	0.044	0.066
LD	0.095	0.097	0.114	0.083	0.081	0.093	0.100	0.093	0.092
MD-LD	0.089	0.089	0.104	0.074	0.064	0.083	0.076	0.066	0.062
PD-LD	0.082	0.082	0.098	0.057	0.048	0.064	0.048	0.047	0.044
PD-LD1	0.118	0.127	0.152	0.065	0.066	0.078	0.049	0.050	0.050

Note: The symbol '*' denotes the preliminary estimator for LD, MD-LD and PD-LD.

Table 5: The biases and root mean squared errors of all estimator under the non-stationarity of the initial condition; $\sigma_\eta^2 = 1$ and $\alpha_J = 0.3$ are used. The three sections of the table represent results for $\alpha = 0.1, 0.5,$ and 0.9.

(n, T)	$(100, 6)$		$(50, 12)$		$(25, 24)$	
	RMSE	Bias	RMSE	Bias	RMSE	Bias
			$\alpha = 0.1$			
AB*	0.064	-0.019	0.059	-0.027	0.062	-0.044
LD	0.096	-0.002	0.142	0.001	0.198	0.010
MD-LD	0.072	-0.007	0.083	-0.004	0.103	0.002
PD-LD	0.065	-0.003	0.066	-0.001	0.075	0.003
PD-LD1	0.066	-0.003	0.066	-0.001	0.075	0.003
BB*	0.065	-0.018	0.070	-0.042	0.158	-0.150
LD	0.096	0.006	0.147	0.001	0.205	-0.013
MD-LD	0.065	-0.002	0.082	-0.004	0.103	-0.005
PD-LD	0.065	0.001	0.068	-0.003	0.072	-0.002
PD-LD1	0.065	0.002	0.068	-0.003	0.072	-0.002
			$\alpha = 0.5$			
AB*	0.139	-0.071	0.094	-0.067	0.088	-0.076
LD	0.087	-0.008	0.117	-0.004	0.170	0.002
MD-LD	0.077	-0.019	0.073	-0.013	0.094	-0.005
PD-LD	0.068	-0.013	0.059	-0.008	0.070	-0.006
PD-LD1	0.076	-0.006	0.061	-0.008	0.070	-0.006
BB*	0.125	0.094	0.070	-0.015	0.174	-0.164
LD	0.090	0.024	0.115	-0.006	0.170	-0.004
MD-LD	0.085	0.040	0.074	-0.008	0.097	-0.007
PD-LD	0.076	0.030	0.058	-0.005	0.071	-0.005
PD-LD1	0.077	0.007	0.059	-0.005	0.071	-0.005
			$\alpha = 0.9$			
AB*	0.072	-0.029	0.052	-0.034	0.056	-0.048
LD	0.044	-0.007	0.032	-0.004	0.035	-0.000
MD-LD	0.042	-0.007	0.030	-0.006	0.031	-0.003
PD-LD	0.041	-0.008	0.024	-0.006	0.020	-0.003
PD-LD1	0.039	-0.004	0.023	-0.004	0.020	-0.003
BB*	0.208	0.207	0.146	0.145	0.072	0.063
LD	0.081	0.069	0.041	0.025	0.035	-0.001
MD-LD	0.086	0.075	0.043	0.030	0.031	0.001
PD-LD	0.082	0.073	0.038	0.030	0.019	-0.002
PD-LD1	0.046	0.018	0.024	0.001	0.020	-0.003

Note: The symbol '*' denotes the preliminary estimator for LD, MD-LD and PD-LD.
The initial observations are generated by $y_{i0} \sim \mathrm{N}(\frac{\eta_i}{1-0.3}, \frac{\sigma_\varepsilon^2}{1-\alpha^2})$.

5.3 Simulation results

First, an overview of the behaviour of all estimators is given for many different sample sizes and $\sigma_\eta^2/\sigma_\varepsilon^2 = 1$, see Table 3. By taking only the difference between the last and the first observation per individual, the LD estimator yields almost no benefit from a larger number T of time periods, unless $\alpha = 0.9$ (then there are more available informative instruments as T increases). In particular, LD performs poorly when α is close to zero. These weaknesses are amended by the proposed estimators. Among these, PD-LD has an overall good performance for all combinations of n and T: (i) it always performs better than LD and MD-LD; (ii) it exhibits smaller RMSEs than AB for $\alpha \geq 0.5$ and is rather close to AB for $\alpha = 0.1$; and (iii) it outperforms BB for $\alpha \leq 0.5$ and – if BB is the initial estimator – PD-LD has similar or smaller RMSE compared to BB for $\alpha = 0.9$ except for the smallest sample size $n = 25$ and $T = 6$. Finally, it is interesting to note that the precision of the PD-LD estimates does not depend much on the initial estimator except for $\alpha = 0.9$, where AB gets very imprecise and substantially biased.

Further, the estimators in the LD class also do not seem to be affected by different ratios of $\sigma_\eta^2/\sigma_\varepsilon^2$. This is documented in Table 4. In the performed experiments, the AB estimator is not substantially influenced by variations in the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$ either. On the contrary, the BB estimator is the most sensitive, in particular when T is large.

Finally, we examine the sensitivity of the estimators to misspecification of the initial condition assumption; Table 5 summarizes now both the RMSE and biases for all estimates. The initial observations y_{i0} are defined as in (33) and $\alpha_J = 0.3$ for all $\alpha \in \{0.1, 0.5, 0.9\}$. It is well known that the BB estimator loses its predominant source of information when y_{it} is mean-nonstationary (see Hahn, 1999). On the contrary, all estimators in the LD class are not substantially affected by different assumptions about y_{i0}. In particular, the biases of LD estimators are almost zero if AB is used as the initial estimator. (Note that the AB estimator actually benefits from mean-nonstationarity, especially when α is close to one, as documented in Hayakawa (2009).) In the other case of the initial BB estimator, LD and PD-LD substantially reduce the bias of the initial estimator, and surprisingly, PD-LD1 even manages to eliminate the bias almost completely (i.e., despite a sizeable upward bias of BB for $\alpha = 0.9$). Finally, note that PD-LD and PD-LD1 exhibit the smallest RMSE of all estimators if $\alpha \geq 0.5$.

Altogether, the PD-LD estimator performs equally well or better than existing methods in the majority of simulated models. The reported experiments show that these results are not overly sensitive to the values of the autoregressive parameter, to the variance of errors, or to the specification of initial observations.

6 Conclusion

To our knowledge, the idea of applying multiple pairwise differences to dynamic linear panel data models is new. This data transformation is presented and applied here to the long-difference estimator of Hahn et al. (2007) to improve its behavior for data with many time periods and for the values of the autoregressive coefficient far from one. We derive the finite-sample bias of the method and the asymptotic distribution of the proposed estimators. Our results indicate that the PD-LD estimator has a smaller variance than the original LD estimator, while preserving its very small bias. In finite samples, simulation results confirm that the proposed pairwise-difference transformation improves the LD estimator in all simulation settings, and in particular, when the time span increases or when α is small. Compared to the existing IV/GMM type of estimator, PD-LD seems to be very competitive without imposing additional restrictive assumptions.

A Appendix

A.1 Finite sample bias

Let us first state and prove the following lemma, which will be used for evaluating of the bias expressions.

Lemma 1. *Let $J \in \mathbb{N}$ and $|\gamma| < 1$. Then it holds for $1 \leq K < L, K \in \mathbb{N}, L \in \mathbb{N}$, that*

$$\sum_{j=K}^{L} j\gamma^j = \frac{\gamma^K - \gamma^{L+1}}{(1-\gamma)^2} - \frac{L\gamma^{L+1} - (K-1)\gamma^K}{1-\gamma}, \tag{34}$$

$$\sum_{j=K}^{L} j^2\gamma^j = 2\frac{\gamma^K - \gamma^{L+1}}{(1-\gamma)^3} - \frac{(2L-1)\gamma^{L+1} - (2K-3)\gamma^K}{(1-\gamma)^2} - \frac{L^2\gamma^{L+1} - (K-1)^2\gamma^K}{1-\gamma}. \tag{35}$$

Proof. The proof follows directly from $(1-\gamma)\sum_{j=0}^{J}\gamma = 1 - \gamma^{J+1}$:

$$\begin{aligned}\sum_{j=0}^{J} j\gamma^j &= \sum_{j=1}^{J}\sum_{l=j}^{J}\gamma^l = \sum_{j=1}^{J}\gamma^j\frac{1-\gamma^{J-j+1}}{1-\gamma} = \frac{1}{1-\gamma}\left(\sum_{j=0}^{J}\gamma^j - 1 - J\gamma^{J+1}\right)\\ &= \frac{1-\gamma^{J+1}}{(1-\gamma)^2} - \frac{J\gamma^{J+1}+1}{1-\gamma},\end{aligned}$$

and using the above result,

$$\begin{aligned}\sum_{j=0}^{J} j^2\gamma^j &= \sum_{j=1}^{J}[\sum_{l=1}^{j}(2l-1)]\gamma^l = \sum_{j=1}^{J}(2j-1)\sum_{l=j}^{J}\gamma^l = \frac{1}{1-\gamma}\left(\sum_{j=1}^{J}(2j-1)\gamma^j - J^2\gamma^{J+1}\right)\\ &= 2\frac{1-\gamma^{J+1}}{(1-\gamma)^3} - \frac{(2J-1)\gamma^{J+1}+\gamma+1}{(1-\gamma)^2} - \frac{J^2\gamma^{J+1}}{1-\gamma}.\end{aligned}$$

Writing now sums $\sum_{j=K}^{L} a_j$ as $\sum_{j=0}^{L} a_j - \sum_{j=0}^{K-1} a_j$ implies the results of the lemma. □

Proof of Theorem 1. In this proof we follow Bun and Kiviet (2006, Appendix A). Results are fully derived for the infeasible PD-LD estimator only. For LD and MD-LD, the proof develops identically except for the final evaluation of the biases as functions of the autoregressive parameter α. We thus proceed with the proof for PD-LD and only the final evaluation is done for each estimator separately.

The estimation error of the unfeasible PD-LD estimator in (15) (obtained after substituting from the model equations (14)) is given by

$$\hat{\alpha}_{nT} - \alpha = \frac{\boldsymbol{y}^{*\prime}_{-1}\boldsymbol{Z}\boldsymbol{V}_n^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}^*}{\boldsymbol{y}^{*\prime}_{-1}\boldsymbol{Z}\boldsymbol{V}_n^{-1}\boldsymbol{Z}'\boldsymbol{y}^*_{-1}} = \frac{\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*}{q_{nT^*}}, \tag{36}$$

where $\boldsymbol{\varepsilon}^* = ([\boldsymbol{D}\boldsymbol{\varepsilon}_1]', \ldots, [\boldsymbol{D}\boldsymbol{\varepsilon}_n]')'$. Suppose that $\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*) = O(N^*)$, where N^* is some function of n and/or T^* to be derived yet. Assuming that either or both n and T^* can get large, Bun and Kiviet (2006, Eq. (31)–(33)) showed that the first-order bias approximation of $\hat{\alpha}_{nT}$ is given by

$$\mathrm{E}(\hat{\alpha}_{nT} - \alpha) = \frac{\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*)}{\bar{q}} + O(N^*(nT^*)^{-3/2}) = B + O(N^*(nT^*)^{-3/2}), \tag{37}$$

where $B = \mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*)/\bar{q} = O(N^*(nT^*)^{-1})$ is the leading term of the bias. Note that the term $(nT^*)^{-1}$ in the previous expressions follows from Assumption B.3 as $q_{nT^*}/(nT^*) \to \bar{q} > 0$ in probability and $q_{nT^*} = O((nT^*)^{-1})$ for $nT^* \to \infty$.

Next, let us derive N^*. First, we can rewrite (36) in a more convenient form. Let $\boldsymbol{G}$ be an $nT^* \times nT^*$ permutation matrix which changes the order of the rows of $\boldsymbol{Z}$, $\boldsymbol{y}^*_{-1}$, $\boldsymbol{y}^*$ such that observations are organized by individuals first ($i = 1, \ldots, n$), then by pairwise differences ($s = S, \ldots, T-1$), and last by time periods ($t = s+1, \ldots T$). As $\boldsymbol{Z}_i$ is block diagonal, $\boldsymbol{G}'\boldsymbol{Z}$ will be block diagonal as well with blocks $\boldsymbol{Z}_{(S+1)S}, \ldots, \boldsymbol{Z}_{T(T-1)}$, where $\boldsymbol{Z}_{ts} = ((\boldsymbol{z}_{its})_{i=1}^n)'$ is $n \times m_{ts}$ and m_{ts} denotes the number of instruments. The inverse weight matrix used here is $\boldsymbol{V}_n = \boldsymbol{Z}'\boldsymbol{Z}$ and is thus also block diagonal. Given that

$$(\boldsymbol{Z}'\boldsymbol{Z})^{-1} = (\boldsymbol{Z}'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{Z})^{-1} = \mathrm{diag}\left((\boldsymbol{Z}'_{(S+1)S}\boldsymbol{Z}_{(S+1)S})^{-1}, \ldots, (\boldsymbol{Z}'_{T(T-1)}\boldsymbol{Z}_{T(T-1)})^{-1}\right), \tag{38}$$

we can rewrite (36) as

$$\hat{\alpha}_{nT} - \alpha = \frac{\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*}{q_{nT^*}} = \frac{\sum_{s=S}^{T-1}\sum_{t=s+1}^{T} \boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{Z}_{ts}(\boldsymbol{Z}'_{ts}\boldsymbol{Z}_{ts})^{-1}\boldsymbol{Z}'_{ts}\boldsymbol{\varepsilon}^*_{ts}}{\sum_{s=S}^{T-1}\sum_{t=s+1}^{T} \boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{Z}_{ts}(\boldsymbol{Z}'_{ts}\boldsymbol{Z}_{ts})^{-1}\boldsymbol{Z}'_{ts}\boldsymbol{y}^*_{(t-1)s}}, \tag{39}$$

where $\boldsymbol{y}^*_{(t-1)s} = (y^*_{1(t-1)s}, \ldots, y^*_{n(t-1)s})'$ and $\boldsymbol{\varepsilon}^*_{ts} = (\varepsilon^*_{1ts}, \ldots, \varepsilon^*_{nts})'$ using $y^*_{i(t-1)s} = \Delta^s y_{i(t-1)} = y_{i(t-1)} - y_{i(t-1-s)}$ and $\varepsilon^*_{its} = \Delta^s \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i(t-s)}$. We now have to analyze the expectation of the nominator of (39),

$$\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*) = \mathrm{E}\left(\sum_{s=S}^{T-1}\sum_{t=s+1}^{T} \boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}\right) = \sum_{s=S}^{T-1}\sum_{t=s+1}^{T} \mathrm{E}(\boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}), \tag{40}$$

where $\boldsymbol{M}_{ts} = \boldsymbol{Z}_{ts}(\boldsymbol{Z}'_{ts}\boldsymbol{Z}_{ts})^{-1}\boldsymbol{Z}'_{ts}$, with $\mathrm{tr}(\boldsymbol{M}_{ts}) = \mathrm{tr}(\boldsymbol{I}_{m_{ts}}) = m_{ts}$. Next,

$$\begin{aligned} \mathrm{E}(\boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}) &= \mathrm{E}\left[\mathrm{tr}(\boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts})\right] = \mathrm{E}\left[\mathrm{tr}(\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}\boldsymbol{y}^{*\prime}_{(t-1)s})\right] \\ &= \mathrm{E}\left\{\mathrm{E}\left[\mathrm{tr}(\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}\boldsymbol{y}^{*\prime}_{(t-1)s})\middle|\mathcal{I}_{t-1}\right]\right\}, \end{aligned} \tag{41}$$

where $\mathrm{E}(\cdot|\mathcal{I}_{t-1})$ denotes the expectation conditional on the information known up to $t-1$. Note that $\boldsymbol{Z}_{ts}$ and thus $\boldsymbol{M}_{ts}$ contain only relevant stochastic elements that have been observed prior to t. Hence

$$\begin{aligned} \mathrm{E}\left\{\mathrm{E}\left[\mathrm{tr}(\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}\boldsymbol{y}^{*\prime}_{(t-1)s})\middle|\mathcal{I}_{t-1}\right]\right\} &= \mathrm{E}\left\{\mathrm{tr}\left[\boldsymbol{M}_{ts}\,\mathrm{E}\left(\boldsymbol{\varepsilon}^*_{ts}\boldsymbol{y}^{*\prime}_{(t-1)s}\middle|\mathcal{I}_{t-1}\right)\right]\right\} \\ &= \mathrm{E}\left\{\mathrm{tr}\left[\boldsymbol{M}_{ts}\boldsymbol{I}_n\,\mathrm{E}\left(\varepsilon^*_{its}y^*_{i(t-1)s}\middle|\mathcal{I}_{t-1}\right)\right]\right\} \\ &= \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{M}_{ts}\right)\mathrm{E}\left(\varepsilon^*_{its}y^*_{i(t-1)s}\middle|\mathcal{I}_{t-1}\right)\right] \\ &= m_{ts}\,\mathrm{E}\left[\mathrm{E}\left(\varepsilon^*_{its}y^*_{i(t-1)s}\middle|\mathcal{I}_{t-1}\right)\right], \end{aligned} \tag{42}$$

provided that the conditional expectations are independent of index i. Under Assumptions B.1–B.3, this however follows from the definition of the transformed variables $y^*_{i(t-1)s}$ and ε^*_{its}:

$$\begin{aligned} \mathrm{E}\left[\mathrm{E}\left(\varepsilon^*_{its}y^*_{i(t-1)s}\middle|\mathcal{I}_{t-1}\right)\right] &= \mathrm{E}\left[\mathrm{E}\left(\Delta^s\varepsilon_{it}\Delta^s y_{i(t-1)}\middle|\mathcal{I}_{t-1}\right)\right] = -\,\mathrm{E}\left(\varepsilon_{i(t-s)}y_{i(t-1)}\right) \\ &= -\,\mathrm{E}\left\{\varepsilon_{i(t-s)}\left[\left(\sum_{k=0}^{t-2}\alpha^k\right)\eta_i + \alpha^{t-1}y_{i0} + \sum_{k=0}^{t-2}\alpha^k\varepsilon_{i(t-1-k)}\right]\right\} \\ &= -\alpha^{s-1}\sigma^2_\varepsilon, \end{aligned} \tag{43}$$

where $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$ for all i and t. This implies for equation (40) that

$$\sum_{s=S}^{T-1}\sum_{t=s+1}^{T} \mathrm{E}(\boldsymbol{y}^{*\prime}_{(t-1)s}\boldsymbol{M}_{ts}\boldsymbol{\varepsilon}^*_{ts}) = -\sigma_\varepsilon^2 \sum_{s=S}^{T-1}\sum_{t=s+1}^{T} m_{ts}\alpha^{s-1}. \tag{44}$$

Note that derivations in (41)–(43) hold for both LD and MD-LD as well – only the bounds of the sums in equations (40) and (44) will differ (depending on the equations used).

To evaluate the biases of the estimators, we consider first the case with all possible instruments included: $m_{ts} = s$ for all $t > s$. For PD-LD, we obtain by Lemma 1 that

$$\begin{aligned}
\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*) &= -\sigma_\varepsilon^2 \sum_{s=S}^{T-1}\sum_{t=s+1}^{T} s\alpha^{s-1} = -\sigma_\varepsilon^2 \sum_{s=S}^{T-1}(T-s)s\alpha^{s-1} \\
&= -\frac{\sigma_\varepsilon^2}{\alpha}\Bigg[T\left(\frac{\alpha^S-\alpha^T}{(1-\alpha)^2} - \frac{(T-1)\alpha^T-(S-1)\alpha^S}{1-\alpha}\right) \\
&\quad - 2\frac{\alpha^S-\alpha^T}{(1-\alpha)^3} + \frac{[2T-3]\alpha^T-[2S-3]\alpha^S}{(1-\alpha)^2} \\
&\quad + \frac{(T-1)^2\alpha^T-(S-1)^2\alpha^S}{1-\alpha}\Bigg],
\end{aligned} \tag{45}$$

which is of order $O(T^2\alpha^{T-\sqrt{2T}})$ for $T\to\infty$ as $S > T-\sqrt{2T}$. Similarly for MD-LD, it holds

$$\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*) = -\sigma_\varepsilon^2 \sum_{s=2}^{T-1}\sum_{t=s+1}^{s+1} s\alpha^{s-1} = \left[-\frac{\sigma_\varepsilon^2}{\alpha}\left(\frac{\alpha^2-\alpha^T}{(1-\alpha)^2} - \frac{(T-1)\alpha^T-\alpha^2}{1-\alpha}\right)\right], \tag{46}$$

which is of order $O(1)$ when $T\to\infty$. Finally, we have for LD

$$\mathrm{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*) = -\sigma_\varepsilon^2 \sum_{s=T-1}^{T-1}\sum_{t=T}^{T} s\alpha^{s-1} = -\sigma_\varepsilon^2(T-1)\alpha^{T-2} = O(T\alpha^T). \tag{47}$$

Next, the case of a bounded number of instruments is considered: suppose $m_{ts} =$

$\max(s, \bar{m})$ for all $t > s$. For PD-LD we have again by Lemma 1

$$
\begin{aligned}
|\operatorname{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*)| &= \left| -\sigma_\varepsilon^2 \sum_{s=S}^{T-1} \sum_{t=s+1}^{T} m_{ts}\alpha^{s-1} \right| \leq \sigma_\varepsilon^2 \bar{m} \sum_{s=S}^{T-1} (T-s)|\alpha|^{s-1} \\
&= \frac{\sigma_\varepsilon^2 \bar{m}}{|\alpha|} \left(T\frac{|\alpha|^S - |\alpha|^T}{1-|\alpha|} + \frac{|\alpha|^S - |\alpha|^T}{(1-|\alpha|)^2} + \frac{|(S-1)|\alpha|^S - (T-1)|\alpha|^T|}{1-|\alpha|} \right),
\end{aligned} \tag{48}
$$

which is of order $O(T|\alpha|^{T-\sqrt{2T}})$ when $T \to \infty$. Similarly, it holds for MD-LD that

$$
\begin{aligned}
|\operatorname{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*)| &= \left| \sigma_\varepsilon^2 \sum_{s=2}^{T-1} \sum_{t=s+1}^{s+1} m_{ts}\alpha^{s-1} \right| \leq \sigma_\varepsilon^2 \bar{m} \sum_{s=2}^{T-1} \sum_{t=s+1}^{s+1} |\alpha|^{s-1} \\
&= \frac{\sigma_\varepsilon^2 \bar{m}}{|\alpha|} \frac{1-|\alpha|^T}{|\alpha|^2 - |\alpha|},
\end{aligned} \tag{49}
$$

which is of order $O(1)$ when $T \to \infty$. Finally, we can write for LD

$$
\begin{aligned}
|\operatorname{E}(\boldsymbol{g}'_{nT^*}\boldsymbol{\varepsilon}^*)| &= \left| \sigma_\varepsilon^2 \sum_{s=T-1}^{T-1} \sum_{t=T}^{T} m_{ts}\alpha^{s-1} \right| \leq \sigma_\varepsilon^2 \bar{m} \sum_{s=T-1}^{T-1} \sum_{t=T}^{T} |\alpha|^{s-1} = \sigma_\varepsilon^2 \bar{m} |\alpha|^{T-2} \\
&= O(|\alpha|^T).
\end{aligned} \tag{50}
$$

□

A.2 Asymptotic distribution

The common notation will be discussed first. The proof of Theorem 2 is identical for LD, MD-LD, and PD-LD except for the dimensions of the instrument and data matrices used. Similarly to $\boldsymbol{y}^*$ in (15), let $\boldsymbol{W}^* = ([\boldsymbol{D}\boldsymbol{W}_1]', \ldots, [\boldsymbol{D}\boldsymbol{W}_n]')'$ and $\boldsymbol{W}_i^* = \boldsymbol{D}\boldsymbol{W}_i$, $i = 1, \ldots, n$, where $\boldsymbol{D}$ is the difference-operator matrix corresponding to the analyzed estimator. The instrument matrices $\boldsymbol{Z}$ and $\boldsymbol{Z}_i$ are also assumed to be corresponding to the estimator of interest (LD, MD-LD, or PD-LD). We will generically refer to $\hat{\boldsymbol{\theta}}_n$ as one of the estimator in this class, which can be now expressed as

$$\hat{\boldsymbol{\theta}}_n = \left(\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}\hat{\boldsymbol{V}}_n^{-1}\hat{\boldsymbol{Z}}'\boldsymbol{W}^*\right)^{-1}\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}\hat{\boldsymbol{V}}_n^{-1}\hat{\boldsymbol{Z}}'\boldsymbol{y}^*, \tag{51}$$

where $\hat{\boldsymbol{Z}}'\boldsymbol{W}^* = \sum_{i=1}^n \hat{\boldsymbol{Z}}_i'\boldsymbol{W}_i^*$, $\hat{\boldsymbol{Z}}'\boldsymbol{y}^* = \sum_{i=1}^n \hat{\boldsymbol{Z}}_i'\boldsymbol{y}_i^*$, and the instrument matrix $\hat{\boldsymbol{Z}}_i$ refers to the feasible counterpart of $\boldsymbol{Z}_i$. Given that the $T^* \times R$ matrix $\hat{\boldsymbol{Z}}_i$ is block diagonal, $\hat{\boldsymbol{Z}}_i = \text{diag}(\hat{\boldsymbol{z}}_{its}')$, the $R \times (K+1)$ matrix $\hat{\boldsymbol{Z}}_i'\boldsymbol{W}_i^*$ can be conveniently partitioned in vectors in the following way

$$\hat{\boldsymbol{Z}}_i'\boldsymbol{W}_i^* = (\hat{\boldsymbol{z}}_{its}w_{itsk}^*)_{(t,s)\in\mathcal{T},k=1,\dots,K+1}, \tag{52}$$

where $(t,s) \in \mathcal{T}$ is the running row-index with values depending on the type of estimator,

$$\begin{aligned}
\mathcal{T}_{\text{LD}} &= \{(t,s) : t = T;\ s = T-1\}, \\
\mathcal{T}_{\text{MD-LD}} &= \{(t,s) : t = s+1;\ s = 2,\dots,T-1\}, \\
\mathcal{T}_{\text{PD-LD}} &= \{(t,s) : t = s+1,\dots,T;\ s = S,\dots,T-1\},
\end{aligned} \tag{53}$$

and $k = 1,\dots,K+1$ is the column index.

The following lemmas will now analyze individual terms of

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) = \left(\frac{\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}}{n}\hat{\boldsymbol{V}}_n^{-1}\frac{\hat{\boldsymbol{Z}}'\boldsymbol{W}^*}{n}\right)^{-1}\frac{\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}}{n}\hat{\boldsymbol{V}}_n^{-1}\frac{\hat{\boldsymbol{Z}}'\boldsymbol{\varepsilon}^*}{\sqrt{n}}, \tag{54}$$

which is obtained by substituting for $\boldsymbol{y}^*$ in (51) from model (24) and where the notation $\boldsymbol{\varepsilon}^* = ([\boldsymbol{D}\boldsymbol{\varepsilon}_1]',\dots,[\boldsymbol{D}\boldsymbol{\varepsilon}_n]')'$ is used.

Lemma 2. *Suppose Assumptions A.1–A.5 hold for a fixed T and $n \to \infty$. Then*

$$\frac{1}{n}\hat{\boldsymbol{Z}}'\boldsymbol{W}^* \xrightarrow{p} \boldsymbol{\Omega}. \tag{55}$$

Proof. We use the decomposition

$$\frac{1}{n}\hat{\boldsymbol{Z}}'\boldsymbol{W}^* = \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_i'\boldsymbol{W}_i^* = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{z}}_{its}w_{itsk}^*\right)_{(t,s)\in\mathcal{T};k=1,\ldots,K+1}, \tag{56}$$

where (t,s) and k are the row and column indices, respectively, of the matrix $\hat{\boldsymbol{Z}}_i\boldsymbol{W}_i^*$. Next, let us analyze the generic vector

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{z}}_{its}w_{itsk}^* &= \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{z}_{its} - \boldsymbol{\mathcal{W}}_{its}(\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0)\right)w_{itsk}^* \\ &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{its}w_{itsk}^* - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\mathcal{W}}_{its}w_{itsk}^*\right)\right](\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0).\end{aligned} \tag{57}$$

First, note that $\sum_{i=1}^{n}\boldsymbol{z}_{its}w_{itsk}^*/n \to \boldsymbol{\omega}_{tsk} = \mathrm{E}(\boldsymbol{z}_{its}w_{itsk}^*)$ in probability as $n \to \infty$ by the law of large numbers (Davidson, 1994, Theorem 20.8) and Assumptions A.1 and A.5. The same argument applies to $\sum_{i=1}^{n}(\boldsymbol{\mathcal{W}}_{its}w_{itsk}^*)/n$. Finally, $\hat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0 = o_p(1)$ follows from the consistency of the preliminary estimator $\hat{\boldsymbol{\theta}}_n^0$ (see Assumption A.3). Consequently, $\sum_{i=1}^{n}\hat{\boldsymbol{z}}_{its}w_{itsk}^*/n \to \boldsymbol{\omega}_{tsk}$ in probability as $n \to \infty$ for any t, s, and k and we can rewrite (56) as

$$\frac{1}{n}\hat{\boldsymbol{Z}}'\boldsymbol{W}^* = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\boldsymbol{z}}_{its}w_{itsk}^*\right)_{ts,k} = \boldsymbol{\Omega} + o_p(1). \tag{58}$$

□

Lemma 3. *Suppose Assumptions A.1–A.5 hold for a fixed T and $n \to \infty$. Then*

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}'\boldsymbol{\varepsilon}^* \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'). \tag{59}$$

Proof. We use again the decomposition

$$\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}'\boldsymbol{\varepsilon}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_i'\boldsymbol{\varepsilon}_i^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\boldsymbol{z}}_{its}\varepsilon_{its}^*\right)_{(t,s)\in\mathcal{T}}. \tag{60}$$

Next, let us analyze the generic vector and substitute for the initial estimator from (25):

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\boldsymbol{z}}_{its}\varepsilon^*_{its} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\boldsymbol{z}_{its} - \frac{1}{\sqrt{n}}\boldsymbol{\mathcal{W}}_{its}\sqrt{n}(\hat{\boldsymbol{\theta}}^0_n - \boldsymbol{\theta}^0)\right)\varepsilon^*_{its} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{z}_{its}\varepsilon^*_{its} - \frac{1}{n}\sum_{i=1}^{n}\left[\boldsymbol{\mathcal{W}}_{its}\left(\boldsymbol{\Lambda}\sqrt{n}\boldsymbol{f}_n(\boldsymbol{\theta}^0) + o_p(1)\right)\varepsilon^*_{its}\right] \qquad (61)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{z}_{its}\varepsilon^*_{its} - \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mathcal{W}}_{its}\varepsilon^*_{its}\right)\boldsymbol{\Lambda}\sqrt{n}\boldsymbol{f}_n(\boldsymbol{\theta}^0) + o_p(1).
\end{aligned}
$$

The law of large numbers (Davidson, 1994, Theorem 20.8) and Assumptions A.1 and A.5 imply that $\sum_{i=1}^{n}\boldsymbol{\mathcal{W}}_{its}\varepsilon^*_{its}/n \to \boldsymbol{P}_{ts} = \mathrm{E}(\boldsymbol{\mathcal{W}}_{its}\varepsilon^*_{its})$ for each t and s.

As $\boldsymbol{P} = (\boldsymbol{P}'_{(S+1)S}, \ldots, \boldsymbol{P}'_{T(T-1)})'$, we can rewrite (60) as

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}'\boldsymbol{\varepsilon}^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\boldsymbol{z}}_{its}\varepsilon^*_{its})_{(t,s)\in\mathcal{T}} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{z}_{its}\varepsilon^*_{its})_{(t,s)\in\mathcal{T}} - (\boldsymbol{P}_{ts})_{(t,s)\in\mathcal{T}}\boldsymbol{\Lambda}\sqrt{n}\boldsymbol{f}_n(\boldsymbol{\theta}^0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{Z}'_i\boldsymbol{\varepsilon}^*_i - \boldsymbol{P}\boldsymbol{\Lambda}\sqrt{n}\boldsymbol{f}_n(\boldsymbol{\theta}^0) + o_p(1) \qquad (62)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\tau}_i(\boldsymbol{\theta}^0) - \boldsymbol{P}\boldsymbol{\Lambda}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}_i(\boldsymbol{\theta}^0) + o_p(1) \\
&= \boldsymbol{M}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\rho}_i(\boldsymbol{\theta}^0) + o_p(1),
\end{aligned}
$$

where $\boldsymbol{M} = (\boldsymbol{I}_R, -\boldsymbol{P}\boldsymbol{\Lambda})$ and $\boldsymbol{\tau}_i(\boldsymbol{\theta}^0) = \boldsymbol{Z}'_i\boldsymbol{\varepsilon}^*_i = \boldsymbol{Z}'_i\boldsymbol{D}\boldsymbol{\varepsilon}_i$ denotes the moment conditions of the LD-type estimator at $\boldsymbol{\theta}^0$.

By Assumption A.1 and A.4, $\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)$ are independent random vectors satisfying $\mathrm{E}[\boldsymbol{\rho}_i(\boldsymbol{\theta}^0)] = \mathbf{0}$. As the second and higher moments exist by Assumptions A.4, the central limit theorem (Davidson, 1994, Theorem 23.12 and 25.6) imply

$$
\frac{1}{\sqrt{n}}\hat{\boldsymbol{Z}}'\boldsymbol{\varepsilon}^* = \boldsymbol{M}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\rho}_i(\boldsymbol{\theta}^0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{M}\boldsymbol{\Sigma}\boldsymbol{M}'). \qquad (63)
$$

□

Proof of Theorem 2. Let $\hat{\boldsymbol{\theta}}_n$ be either LD, PD-LD or MD-LD in (51). By (54), it can be written as

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) = \left[\left(\frac{\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}}{n}\right)\hat{\boldsymbol{V}}_n^{-1}\left(\frac{\hat{\boldsymbol{Z}}^{\prime}\boldsymbol{W}^{*}}{n}\right)\right]^{-1}\left(\frac{\boldsymbol{W}^{*\prime}\hat{\boldsymbol{Z}}}{n}\right)\hat{\boldsymbol{V}}_n^{-1}\left(\frac{\hat{\boldsymbol{Z}}^{\prime}\boldsymbol{\varepsilon}^{*}}{\sqrt{n}}\right).$$

By Assumption A.6 and Lemma 2, it follows for $n \to \infty$

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) = \left(\left[\boldsymbol{\Omega}^{\prime}\boldsymbol{V}\boldsymbol{\Omega}\right]^{-1}\boldsymbol{\Omega}^{\prime}\boldsymbol{V} + o_p(1)\right)\left(\frac{\hat{\boldsymbol{Z}}^{\prime}\boldsymbol{\varepsilon}^{*}}{\sqrt{n}}\right).$$

The claim of the theorem now follows from Lemma 3. □

References

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Aquaro, M. and P. Čížek (2010). One-step robust estimation of fixed-effects panel data models. Technical Report 2010-110, Tilburg University.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The 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.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*, 200–210.

Bun, M. J. and J. F. Kiviet (2006). The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models. *Journal of Econometrics 132*(2), 409–444.

Carrasco, M. and J.-P. Florens (2000). Generalization of gmm to a continuum of moment conditions. *Econometric Theory 16*(06), 797–834.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Gouriéroux, C., P. C. Phillips, and J. Yu (2010). Indirect inference for dynamic panel models. *Journal of Econometrics 157*(1), 68–77.

Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of Econometrics 31*(1), 93 – 118.

Hahn, J. (1999). How informative is the initial condition in the dynamic panel model with fixed effects? *Journal of Econometrics 93*(2), 309–326.

Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics 140*(2), 574–617.

Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica 70*(4), 1639–1657.

Hayakawa, K. (2009). On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics 153*(2), 133–135.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*(6), 1371–1395.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68*(1), 53–78.

Kiviet, J. F. (2007). Judging contending estimators by simulation: tournaments in dynamic panel data models. In G. Phillips and E. Tzavalis (Eds.), *The Refinement of Economet-*

ric Estimation and Test Procedures, Chapter 11, pp. 282–318. Cambridge: Cambridge University Press.

Newey, W. K. and R. J. Smith (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica 72*(1), 219–255.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.