Asymptotically Efficient Method of Moments Estimators for Dynamic Panel Data Models

Jörg Breitung Kazuhiko Hayakawa Sebastian Kripfganz

November 21, 2019

Abstract

In this paper, simple variants of bias correction for dynamic panel data models are proposed and their asymptotic properties are studied when the number of time periods is fixed or tends to infinity with the number of panel units. Our approach can easily generalised to higher-order autoregressive models and cross-sectional dependence. Panel-corrected standard errors are proposed that allow for robust inference in dynamic models with cross-sectionally correlated errors. Monte Carlo experiments suggest that bias-corrected method of moment estimator outperform popular GMM estimators in terms of efficiency and correctly sized tests.

1 Introduction

Dynamic panel data models are now widely used in a wide area of empirical applications. Since the work of Anderson and Hsiao (1981), the instrumental variables and the generalized method of moments (GMM) estimators have been extensively used in the estimation of dynamic panel data models. However, it is known that the GMM estimator by Holtz-Eakin et al. (1988) and Arellano and Bond (1991) suffers from the weak-instruments problem when the persistency of the data is strong, as demonstrated by Blundell and Bond (1998). They also showed that the GMM estimator for models in levels with first-differenced instruments mitigates that problem and proposed the so-called system GMM estimator by combining models in first differences and in levels. Nowadays, the system GMM estimator is most frequently used in practice albeit Bun and Windmeijer (2010) showed that it still suffers from the weak-instruments problem when the variance of the individual specific effects is larger than that of the idiosyncratic errors. As alternatives to the GMM approach, maximum likelihood (ML) estimators and bias-corrected within-groups (WG) estimators were proposed. Hsiao et al. (2002) suggested a transformed ML estimator that adapts the ML approach to the differenced variables. Hayakawa and Pesaran (2015) extended this transformed ML estimator to allow for cross-sectional heteroskedasticity and proposed robust standard errors.

With regard to bias-corrected WG estimators, Kiviet (1995) and Judson and Owen (1999) demonstrate that they are attractive alternatives to GMM estimators. Although the bias-corrected WG estimator of Kiviet (1995) is based on a higher-order expansion of the bias term, the analytical results are based on the unknown parameters that have to be estimated by some consistent initial estimator. Accordingly, the asymptotic distribution of this estimator is unknown. Bun and Carree (2005) proposed an alternative bias-corrected WG estimator which iteratively solves a nonlinear equation with regard to unknown parameters.

In this paper, we consider a simplified variant of the bias-adjusted likelihood approach by Dhaene and Jochmans (2016). We demonstrate that the likelihood scores can be transformed into nonlinear moment conditions that can easily be solved with standard methods, and show that the resulting estimator is asymptotically equivalent to the ML estimators suggested by Hsiao et al. (2002) and

Bai (2013). This approach simplifies the derivation of the asymptotic properties and allows us to develop a bias-corrected method of moment estimator for higher-order dynamic models. Furthermore, we propose "cluster-robust" (resp. "panel-corrected") standard errors that account for cross-sectional dependence. Monte Carlo experiments suggest that these estimators perform well (in terms of efficiency and correctly sized tests) relative to the uncorrected least-squares (WG) or GMM approaches.

The rest of this paper is organized as follows. In Section 2, we show that the adjusted profile scores can be written as nonlinear moment conditions, and derive the asymptotic properties where the time series dimension is fixed or tends to infinity with the number of cross sections diverging. In Section 3, we explore the relationship between the bias-corrected estimator and ML estimators. In Section 4, the model is extended to include higher-order dynamics. In Section 5, we consider the case where errors are cross-sectionally correlated. The results of various Monte Carlo experiments are summarized in Section 6 and Section 7 concludes.

2 Bias-corrected estimators for the baseline model

To motivate our bias-corrected estimator, we first consider the pure autoregressive model

$$y_{it} = \mu_i + \alpha y_{i,t-1} + u_{it}, \qquad t = 1, \dots, T, \quad i = 1, \dots, N,$$

where $u_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. For estimating the parameter α, Dhaene and Jochmans (2016) consider the profile likelihood that results from profiling out the parameters μ_i and σ^2 from the log-likelihood function, yielding the "profile scores"

$$s(\alpha) = \frac{\partial \ell(\alpha)}{\partial \alpha} \;=\; \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{-1,i})(y_{it} - \alpha y_{i,t-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \overline{y}_{-1,i})^2},$$

where $\overline{y}_{-1,i} = T^{-1}\sum_{t=1}^{T} y_{i,t-1}$. The bias-corrected profile likelihood estimator $\widetilde{\alpha}$ results from solving the equation $s(\widetilde{\alpha}) - \mathbb{E}[s(\widetilde{\alpha})] = 0$. Unfortunately, the bias term $\mathbb{E}[s(\alpha)]$ is a complicated function of α as it involves an expectation of a ratio of two random variables both depending on α. To simplify the derivation of the bias function, we first assume that the variance σ^2 is known, resulting in the profile score

$$\widetilde{s}(\alpha) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t-1} - \overline{y}_{-1,i})(y_{it} - \alpha y_{i,t-1}).$$

Our bias-adjusted method of moments estimator is based on the moment condition

$$\mathbb{E}\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[(y_{i,t-1} - \overline{y}_{-1,i})(y_{it} - \alpha y_{i,t-1}) - \frac{\sigma^2}{T} b_T(\alpha) \right] \right) = 0, \qquad (1)$$

where

$$b_T(\alpha) = -\mathbb{E}\left[\frac{1}{\sigma^2 T} \sum_{t=1}^{T} (y_{i,t-1} - \overline{y}_{-1,i})(y_{it} - \alpha y_{i,t-1}) \right]$$

is a polynomial in α that is presented further below. Since σ^2 is unknown, we replace it with an unbiased estimator. Accordingly, our modified profile likelihood estimator is obtained by solving the moment equation

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[(y_{i,t-1} - \overline{y}_{-1,i})(y_{it} - \alpha y_{i,t-1}) - \widehat{\sigma}^2(\alpha) b_T(\alpha) \right] = 0,$$

where $\mathbb{E}[\widehat{\sigma}^2(\alpha)] = \sigma^2$. Solving this moment equation is much simpler than solving the bias-adjusted profile score equation suggested by Dhaene and Jochmans (2016).

Let us now consider some more details of our alternative bias correction for the first-order dynamic model with strictly exogenous regressors given by

$$y_{it} = \mu_i + \alpha_0 y_{i,t-1} + \boldsymbol{\beta}_0' \mathbf{x}_{it} + u_{it} \quad t = 1, \ldots, T, \quad i = 1, \ldots, N, \qquad (2)$$

where α_0 and the $k \times 1$ vector $\boldsymbol{\beta}_0$ denote the true values of the parameters of interest. For the $k \times 1$ vector of regressors $\mathbf{x}_{it}$ and the error term u_{it} the following assumptions are imposed:

Assumption 1. *(i) The errors u_{it} are independent across i and t with $\mathbb{E}(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_i^2 < C$ for some constant $C < \infty$. (ii) The regressors are strictly exogenous with $\mathbb{E}(\mathbf{x}_{it} u_{is}) = 0$ and $\mathbb{E}||u_{it} u_{is} \mathbf{x}_{it} \mathbf{x}_{is}'|| < \infty$ for all $t, s \in \{1, \ldots, T\}$ and $i \in \{1, \ldots, N\}$. (iii) $\mathbb{E}|u_{it}|^{4+\delta} < \infty$ for all i and t and some $\delta > 0$. (iv) For the initial values we assume $\mathbb{E}(y_{i0})^2 < \infty$ for all i and $\mathbb{E}(y_{i0} u_{it}) = 0$ for all i and $t \in \{1, \ldots, T\}$.*

This set of assumptions is standard in the literature on dynamic panel data models (e.g. Arellano and Bond, 1991, and Ahn and Schmidt, 1995). Note that we do not impose any stationarity assumption on the initial values. Therefore, the process may start at any fixed or random level in the finite past. If T is fixed, assumption (1) does not rule out nonstationary regressors or instable dynamic processes with $\alpha \geq 1$. However, if T tends to infinity, additional assumptions are required for the limiting distribution of the estimator. It should also be noted that we allow for individual-specific heteroskedasticity. In Remark 3 we show that our estimator is robust to time series heteroskedasticity as $T \to \infty$.

Assuming normally distributed errors and treating y_{i0} as a fixed constant, the first-order condition of the ML (least-squares) estimator $\widehat{\boldsymbol{\theta}} = [\widehat{\alpha}, \widehat{\boldsymbol{\beta}}']'$ results as

$$\mathbf{g}_{NT}(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} g_{\alpha,NT}(\widehat{\boldsymbol{\theta}}) \\ \mathbf{g}_{\beta,NT}(\widehat{\boldsymbol{\theta}}) \end{bmatrix} = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} y_{i,t-1} - \overline{y}_{-1,i} \\ \mathbf{x}_{it} - \overline{\mathbf{x}}_i \end{bmatrix} \left[e_{it}(\widehat{\boldsymbol{\theta}}) - \overline{e}_i(\widehat{\boldsymbol{\theta}}) \right] = \mathbf{0},$$

where $\overline{y}_{-\ell,i} = T^{-1} \sum_{t=1}^{T} y_{i,t-\ell}$, with the index $-\ell$ suppressed for $\ell = 0$, $e_{it}(\boldsymbol{\theta}) = y_{it} - \alpha y_{i,t-1} - \boldsymbol{\beta}' \mathbf{x}_{it}$, and $\overline{e}_i(\boldsymbol{\theta}) = T^{-1} \sum_{s=1}^{T} e_{is}(\boldsymbol{\theta})$.

The probability limit for $N \to \infty$ and fixed T evaluated at the true parameters is obtained as (e.g. Nickell, 1981, and Moon et al., 2015)

$$\begin{aligned} \operatorname*{plim}_{N\to\infty} g_{\alpha,NT}(\boldsymbol{\theta}_0) &= \lim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathbb{E} \left[\sum_{t=1}^{T} y_{i,t-1} \left(e_{it}(\boldsymbol{\theta}_0) - \frac{1}{T} \sum_{s=1}^{T} e_{is}(\boldsymbol{\theta}_0) \right) \right] \\ &= -\overline{\sigma}^2 \left(\frac{T-2}{T^2} + \frac{T-3}{T^2}\alpha_0 + \frac{T-4}{T^2}\alpha_0^2 + \cdots + \frac{2}{T^2}\alpha_0^{T-3} + \frac{1}{T^2}\alpha_0^{T-2} \right) \\ &= -\frac{1}{T} b_T(\alpha_0) \overline{\sigma}^2, \end{aligned} \tag{3}$$

where $\overline{\sigma}^2 = \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \sigma_i^2$ and

$$\begin{aligned} b_T(\alpha) &= \frac{1}{T} \sum_{t=0}^{T-2} \sum_{s=0}^{t} \alpha^s \\ &= \frac{1}{1-\alpha} + O(T^{-1}). \end{aligned} \tag{4}$$

Note that $b_T(\alpha)$ is a positive monotonously increasing function.

Since $\operatorname{plim}_{N\to\infty} g_{\alpha,NT}(\boldsymbol{\theta}_0) \neq 0$, the FE estimator is inconsistent if T is small. Using

$$\mathbb{E} \left[\frac{1}{T-1} \sum_{t=1}^{T} e_{it}(\boldsymbol{\theta}_0)(e_{it}(\boldsymbol{\theta}_0) - \overline{e}_i(\boldsymbol{\theta}_0)) \right] = \sigma_i^2,$$

we obtain the set of moment conditions

$$\mathbb{E}[\mathbf{m}_{Ti}(\boldsymbol{\theta}_0)] = \mathbb{E}\begin{bmatrix} m_{Ti,\alpha}(\boldsymbol{\theta}_0) \\ \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}_0) \end{bmatrix} = \mathbb{E}\left[\mathbf{w}_{it}(\boldsymbol{\theta}_0)(e_{it} - \overline{e}_i(\boldsymbol{\theta}_0))\right] = \mathbf{0}, \tag{5}$$

where

$$m_{Ti,\alpha}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\left(y_{i,t-1} + \frac{b_T(\alpha)}{T-1}e_{it}(\boldsymbol{\theta})\right)(e_{it}(\boldsymbol{\theta}) - \overline{e}_i(\boldsymbol{\theta})) \tag{6}$$

$$\mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) = \sum_{t=1}^{T}\mathbf{x}_{it}(e_{it}(\boldsymbol{\theta}) - \overline{e}_i(\boldsymbol{\theta})) \tag{7}$$

$$\mathbf{w}_{it}(\boldsymbol{\theta}) = \left[y_{i,t-1} + \frac{b_T(\alpha)}{T-1}e_{it}(\boldsymbol{\theta}),\ \mathbf{x}'_{it}\right]'. \tag{8}$$

The bias-corrected method of moments estimator $\widehat{\boldsymbol{\theta}}_{bc}$ is obtained by solving the moment conditions $\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$ based on the recursion

$$\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0 - \left[\sum_{i=1}^{N}\nabla_\theta\mathbf{m}_{Ti}(\boldsymbol{\theta}^0)\right]^{-1}\left(\sum_{i=1}^{N}\mathbf{m}_{Ti}(\boldsymbol{\theta}^0)\right),$$

where $\boldsymbol{\theta}^0$ is the initial value, $\boldsymbol{\theta}^1$ denotes the updated value, and

$$\nabla_\theta\mathbf{m}_{Ti}(\boldsymbol{\theta}^0) = \left.\frac{\partial\mathbf{m}_{Ti}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\theta^0}.$$

The entries of the matrix $\nabla'_\theta\mathbf{m}_{Ti}(\boldsymbol{\theta})$ are provided in Appendix A. The limiting distribution for $N \to \infty$ and fixed T is presented in the following theorem.

Theorem 1. *(i) Under Assumption 1, the limiting distribution of* $\widehat{\boldsymbol{\theta}}_{bc} = [\widehat{\alpha}_{bc}, \widehat{\boldsymbol{\beta}}'_{bc}]'$ *for fixed* T *and* $N \to \infty$ *is given by*

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{D}_T(\boldsymbol{\theta}_0)^{-1}\mathbf{S}_T(\boldsymbol{\theta}_0)\mathbf{D}_T(\boldsymbol{\theta}_0)^{-1}\right),$$

where

$$\mathbf{S}_T(\boldsymbol{\theta}) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_{Ti}(\boldsymbol{\theta})\mathbf{m}_{Ti}(\boldsymbol{\theta})', \quad \mathbf{D}_T(\boldsymbol{\theta}) = \begin{bmatrix} D_{T,\alpha\alpha}(\boldsymbol{\theta}) & \mathbf{D}_{T,\alpha\beta}(\boldsymbol{\theta})' \\ \mathbf{D}_{T,\alpha\beta}(\boldsymbol{\theta}) & \mathbf{D}_{T,\beta\beta}(\boldsymbol{\theta}) \end{bmatrix},$$

$$D_{T,\alpha\alpha}(\boldsymbol{\theta}) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}(y_{i,t-1} - \overline{y}_{-1,i}) - \sigma^2\nabla_\alpha b_T(\alpha) - \frac{2}{T-1}\sigma^2 b_T(\alpha)^2,$$

$$\mathbf{D}_{T,\alpha\beta}(\boldsymbol{\theta}) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)y_{i,t-1}, \quad \mathbf{D}_{T,\beta\beta}(\boldsymbol{\theta}) = \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\mathbf{x}'_{it},$$

and $\nabla_\alpha b_T(\alpha) = T^{-1}\sum_{t=1}^{T-2}\sum_{s=1}^{t} s\alpha^{s-1}$.

(ii) Let $\mathbf{z}_{it} = [y_{i,t-1}, \mathbf{x}_{it}']'$ *and assume*

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{z}_{it}-\overline{\mathbf{z}}_i)(\mathbf{z}_{it}-\overline{\mathbf{z}}_i)'\right] \;\to\; \mathbf{V}_z \ ,$$

as $N, T \to \infty$ *and* $N/T \to \kappa, 0 \leq \kappa < \infty$*, where* $\mathbf{V}_z$ *is a positive-definite and finite matrix. The limiting distribution for* $N, T \to \infty$ *is given by*

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2\mathbf{V}_z^{-1}).$$

Remark 1. The covariance matrix of the bias-corrected method of moments estimator can be consistently estimated by the small-sample analog

$$\widehat{\mathbf{V}}_{NT}(\widehat{\boldsymbol{\theta}}_{bc}) = \left[\sum_{i=1}^{N}\nabla_\theta'\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})\right]^{-1}\left(\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})'\right)\left[\sum_{i=1}^{N}\nabla_\theta'\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})'\right]^{-1}, \tag{9}$$

and the usual test statistics can be employed for inference. For example, we may examine the linear hypothesis $H_0 : \mathbf{R}\boldsymbol{\theta} - \mathbf{r} = \mathbf{0}$ by the Wald statistic $(\mathbf{R}\widehat{\boldsymbol{\theta}}_{bc} - \mathbf{r})'\widehat{\mathbf{V}}_{NT}(\widehat{\boldsymbol{\theta}}_{bc})^{-1}(\mathbf{R}\widehat{\boldsymbol{\theta}}_{bc} - \mathbf{r})$.

Remark 2. As shown by Hahn and Kuersteiner (2002) and Bai (2013), the asymptotic variance for the case $N/T \to \kappa$ is identical to the lower variance bound, which is equivalent to the asymptotic variance in the case of no individual specific constants. Therefore, whenever $T \to \infty$, the estimator is asymptotically efficient. It is important to note that according to (ii) the estimator does not involve an asymptotic bias. In contrast to the LSDV estimator and the GMM estimators, inference is valid also for all values of κ. This finding suggests that the estimator is particularly attractive in macro panels, where N and T are of similar magnitude (cf. Breitung, 2015).

Remark 3. According to Assumption 1, the bias-corrected method of moments estimator is robust against heteroskedasticity across panel units. On the other hand, the bias correction assumes that the error variances within each group are constant over time. However, it is not difficult to see that the inconsistency becomes less severe as T becomes large. Let $\sigma_{it}^2 = \mathbb{E}(u_{it}^2)$ such that

$$\begin{aligned}
\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1}\overline{u}_i\right) &= \left(\frac{1}{T}\sum_{t=1}^{T-1}\sigma_{it}^2\right) + \alpha_0\left(\frac{1}{T}\sum_{t=1}^{T-2}\sigma_{it}^2\right) + \alpha_0^2\left(\frac{1}{T}\sum_{t=1}^{T-3}\sigma_{it}^2\right) + \cdots \\
&= b_T(\alpha_0)\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{it}^2\right) + \frac{1}{T}\left[\sigma_{iT}^2 + \alpha_0(\sigma_{i,T-1}^2 + \sigma_{iT}^2)\right. \\
&\quad \left. + \alpha_0^2(\sigma_{i,T-2}^2 + \sigma_{i,T-1}^2 + \sigma_{iT}^2) + \cdots\right].
\end{aligned}$$

Let $C_i < \infty$ be some upper bound of the variances $\{\sigma_{i1}^2, \ldots, \sigma_{i,T-1}^2\}$ such that

$$\frac{1}{T}\left[\sigma_{iT}^2 + \alpha_0(\sigma_{i,T-1}^2 + \sigma_{iT}^2) + \alpha_0^2(\sigma_{i,T-2}^2 + \sigma_{i,T-1}^2 + \sigma_{iT}^2) + \cdots\right] < C_i \sum_{t=1}^{T-1} t\alpha_0^{t-1} .$$

Since $\sum_{t=1}^{T-1} t\alpha_0^{t-1}$ is bounded as $T \to \infty$, it follows that

$$\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1}\overline{u}_i\right) = b_T(\alpha_0)\left(\frac{1}{T}\sum_{t=1}^{T}\sigma_{it}^2\right) + O(T^{-1}).$$

For $T \to \infty$ we have $(T-1)^{-1}\sum_{t=1}^{T}(e_{it}(\boldsymbol{\theta}_0) - \overline{e}_i(\boldsymbol{\theta}_0))^2 \xrightarrow{p} \overline{\sigma}_i^2$, where $\overline{\sigma}_i^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sigma_{it}^2$. It follows that the bias correction is valid under temporal heteroskedasticity whenever T is large.

Remark 4. The model setup corresponds to a fixed-effects panel data model. The estimator can easily be adapted to a random-effects framework, where it is assumed that μ_i is a random variable with $\mathbb{E}(\mu_i) = 0$, $\mathbb{E}(\mu_i^2) = \sigma_\mu^2$, and $\mathbb{E}(\mu_i \mathbf{x}_{it}) = 0$ for all i and t. In this case, $\mathbf{m}_{Ti,\beta}(\boldsymbol{\theta})$ in moment equation (7) is replaced by

$$\mathbf{m}_{Ti,\beta}^*(\boldsymbol{\theta}) = \sum_{t=1}^{T}\mathbf{x}_{it}e_{it}(\boldsymbol{\theta}). \tag{10}$$

The respective estimator is more efficient than the original method of moments estimator from the fixed-effects framework as long as the regressors are uncorrelated with the individual effects. The random-effects framework also allows to include time-invariant regressors.

3 Relationship to maximum likelihood estimation

In this section, we show that our bias-corrected method of moments estimator is similar to the ML estimation procedures proposed by Hsiao et al. (2002) and Bai (2013). In particular, we show that the first-order conditions of the ML estimation procedures can be decomposed into two terms. The first term results from the first-order condition of the least-squares estimator, $g_{\alpha,NT}(\boldsymbol{\theta})$, whereas the second term acts as a bias correction. Accordingly, the main differences between these approaches are the assumptions on the initial condition that result in different bias correction terms. An important advantage of our approach is that we do not need to impose any assumption on the initial condition y_{i0} other than it has finite variance and is uncorrelated with all subsequent errors $u_{i1}, \ldots, u_{iT}$.

To simplify the discussion, we focus on the simple AR(1) model without exogenous variables and with homoskedastic errors. Let

$$\mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{pmatrix} = \mathbf{y}_i - \alpha \mathbf{y}_{-1,i} \quad \text{and } \mathbf{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix} = \mathbf{e}_i - \mu_i \boldsymbol{\iota}_T,$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_{-1,i} = (y_{i0}, y_{i1}, \ldots, y_{i,T-1})'$, and $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. The (conditional) log-likelihood function is given by

$$\begin{aligned} \ell(\alpha, \sigma^2, \boldsymbol{\mu}) &= \frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N} \mathbf{u}_i'\mathbf{u}_i \\ &= \frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(u_{it} - \overline{u}_i)^2 - \frac{T}{2\sigma^2}\sum_{i=1}^{N}\overline{u}_i^2 \\ &= \frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it} - \overline{e}_i)^2 - \frac{T}{2\sigma^2}\sum_{i=1}^{N}\overline{u}_i^2. \end{aligned} \tag{11}$$

Note that the last term in this likelihood function depends on α and μ_i, therefore, dropping this term (by concentrating out μ_i) results in the within-groups estimator that is known to be biased. The first derivative of the likelihood function is given by

$$\frac{\partial \ell(\alpha, \sigma^2, \mu)}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-1}(e_{it} - \bar{e}_i) + \frac{T}{\sigma^2} \sum_{i=1}^{N} \overline{u}_i \overline{y}_{-1,i}.$$

As $\overline{u}_i$ is unknown, we replace the last term with its expectation,

$$\frac{T}{\sigma^2} \mathbb{E}(\overline{u}_i \overline{y}_{-1,i}) = b_T(\alpha),$$

which yields the nonlinear moment conditions of the previous section.

In the literature, two other likelihood-based estimators were suggested. Hsiao et al. (2002) assume that $u_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for $t = 1, \ldots, T$ and $\Delta y_{i1} \overset{iid}{\sim} \mathcal{N}(b, \omega)$. The initial condition implies

$$\begin{aligned} y_{i1} - y_{i0} &= (\alpha - 1) y_{i0} + \mu_i + u_{i1} \\ &= u_{i1} - u_{i0} + b. \end{aligned}$$

and from $\mathbb{E}(u_{i1} u_{i0}) = 0$ it follows that $\mathbb{E}(u_{i0}^2) = (\omega - 1)\sigma^2$. The (Gaussian) log-likelihood function is given by

$$\ell(\alpha, \sigma^2, \omega) = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln(\sigma^2) - \frac{N}{2} \ln(|\mathbf{\Omega}|) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \mathbf{e}_i^{*\prime} \mathbf{D}' \mathbf{\Omega}^{-1} \mathbf{D} \mathbf{e}_i^*,$$

where $\mathbf{e}_i^* = (u_{i0}, \mathbf{e}_i')'$, $\mathbf{D} = (\mathbf{0}, \mathbf{I}_T) - (\mathbf{I}_T, \mathbf{0})$ is a $T \times (T+1)$ matrix, $\mathbf{\Omega} = \mathbf{D}\mathbf{D}' - (2 - \omega) \boldsymbol{\varphi}_T \boldsymbol{\varphi}_T'$ is a $T \times T$ matrix, and $\boldsymbol{\varphi}_T$ is the first column of the $T \times T$ identity matrix $\mathbf{I}_T$ such that

$$\mathbf{\Omega}^{-1} = (\mathbf{D}\mathbf{D}')^{-1} + \frac{(2 - \omega)(T+1)}{|\mathbf{\Omega}|} (\mathbf{D}\mathbf{D}')^{-1} \boldsymbol{\varphi}_T \boldsymbol{\varphi}_T' (\mathbf{D}\mathbf{D}')^{-1},$$

where $|\mathbf{\Omega}| = T(\omega - 1) + 1$. Using $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} = \mathbf{I}_{T+1} - (T+1)^{-1} \boldsymbol{\iota}_{T+1} \boldsymbol{\iota}_{T+1}'$, the log-likelihood function results as

$$\begin{aligned} \ell(\alpha, \sigma^2, \omega) = &- \frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln(\sigma^2) - \frac{N}{2} \ln(|\mathbf{\Omega}|) \\ &- \frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{t=0}^{T} (u_{it} - \bar{u}_i^*)^2 - \frac{(2 - \omega)(T+1)}{2\sigma^2 |\mathbf{\Omega}|} \sum_{i=1}^{N} (u_{i0} - \bar{u}_i^*)^2 \\ = &- \frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln(\sigma^2) - \frac{N}{2} \ln(|\mathbf{\Omega}|) \\ &- \frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} (e_{it} - \bar{e}_i)^2 - \frac{T}{2\sigma^2 |\mathbf{\Omega}|} \sum_{i=1}^{N} (u_{i0} - \bar{u}_i)^2, \end{aligned}$$

using the relationship

$$u_{i0} - \bar{u}_i^* = \frac{T}{T+1}(u_{i0} - \bar{u}_i)$$

between the within-group average $\bar{u}_i^* = (T+1)^{-1}\sum_{t=0}^{T} u_{it}$ that includes the initial value u_{i0} and the within-group average $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$ that does not include u_{i0}.

Obviously, the last term in the log-likelihood function yields a bias correction. Instead of $T\overline{u}_i^2$ in the log-likelihood function (11), this approach employs an adjusted term, where $\overline{u}_i$ is replaced with the mean of the adjusted residuals, $T^{-1}\sum_{t=1}^{T}(u_{it} - u_{i0})$. This adjustment cancels the individual effect but implies a different variance. Indeed it is not difficult to show that $T\mathbb{E}(\overline{u}_i - u_{i0})^2 = T(\omega - 1)\sigma^2 + \sigma^2 = \sigma^2|\mathbf{\Omega}|$, which explains the different denominator of the last term. Upon differentiation, the last term yields the bias adjustment given by

$$\widetilde{b}(\alpha, \omega) = \frac{T}{\sigma^2|\mathbf{\Omega}|}\sum_{i=1}^{N}(y_{i0} - \bar{y}_{i,-1})(u_{i0} - \bar{u}_i). \tag{12}$$

In Appendix A it is shown that $\mathbb{E}[\widetilde{b}_T(\alpha, \omega)] = b_T(\alpha)$. An important drawback of this approach is that it involves an additional parameter ω that has to be estimated, although the bias does not involve this parameter.

A similar comment applies to the ML framework of Bai (2013). In the AR(1) model without exogenous variables, the individual effects can be treated as a random effect. Again an initial condition is required. We follow Bai (2013) and assume $y_{i0} = 0$ for all i. It is not difficult to show that the derivative of the Gaussian log-likelihood function results as

$$\frac{\partial \ell(\alpha, \sigma^2, \pi)}{\partial \alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}(e_{it} - \overline{e}_i) + \frac{T}{T\pi_N + \sigma^2}\sum_{i=1}^{N}\overline{y}_{i,-1}(\overline{y}_i - \alpha\overline{y}_{i,-1}),$$

where $\pi_N = N^{-1}\sum_{i=1}^{N}\mu_i^2$. Profiling out $\mathbb{E}(\overline{y}_i - \alpha\overline{y}_{-1,i})^2 = T\pi + \sigma^2$ and σ^2 yields the first-order condition

$$g_{NT}(\alpha) + \frac{\widehat{\sigma}^2 T \sum\limits_{i=1}^{N}\overline{y}_{i,-1}(\overline{y}_i - \alpha\overline{y}_{i,-1})}{\sum\limits_{i=1}^{N}(\overline{y}_i - \alpha\overline{y}_{i,-1})^2} = 0,$$

where

$$g_{NT}(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}(e_{it} - \bar{e}_i),$$
$$\widehat{\sigma}^2 = \frac{1}{(T-1)N}\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it} - \bar{e}_i)^2.$$

It is not difficult to show that, if $y_{i0} = 0$ for all i, then

$$\frac{1}{N}\sum_{i=1}^{N}(\overline{y}_i - \alpha\overline{y}_{-1,i})^2 \xrightarrow{p} \sigma^2 + T\pi,$$
$$\frac{1}{N}\sum_{i=1}^{N}\overline{y}_{i,-1}(\overline{y}_i - \alpha\overline{y}_{-1,i}) \xrightarrow{p} b_T(\alpha)(\sigma^2 + T\pi),$$

and, thus, the bias correction is asymptotically equivalent to the bias correction applied in Theorem 1. It is important to notice, however, that the validity of this bias correction crucially depends on the initial condition $y_{i0} = 0$.

4 Higher-order dynamics

In this section, the bias-corrected method of moments estimator is generalized to an autoregressive model of order p. To simplify the discussion, the (strictly exogenous) regressors are neglected. The treatment of the additional regressors are considered at the end of this section. Consider the autoregressive model of order p,

$$y_{it} = \mu_i + \alpha_1 y_{i,t-1} + \cdots + \alpha_p y_{i,t-1} + u_{it}, \quad t = 1, \ldots, T,\ i = 1, \ldots, N, \qquad (13)$$

where Assumption 1 applies. It is convenient to rewrite the model in companion form:

$$\begin{bmatrix} 1 & 0 & & \cdot & \cdot & 0 & \cdot & 0 \\ -\alpha_1 & 1 & 0 & \cdot & 0 & 0 & \cdot & 0 \\ -\alpha_2 & -\alpha_1 & 1 & \cdot & 0 & 0 & \cdot & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & & -\alpha_p & -\alpha_{p-1} & \cdot & 1 \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ \vdots \\ y_{iT} \end{bmatrix} = \mu_i \boldsymbol{\imath}_T + \begin{bmatrix} \alpha_1 & \alpha_2 & \cdot & \alpha_p \\ \alpha_2 & \alpha_3 & \cdot & 0 \\ \alpha_3 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 \end{bmatrix} \begin{bmatrix} y_{i0} \\ y_{i,-1} \\ \vdots \\ y_{i,-p+1} \end{bmatrix} + \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix},$$

or

$$\mathbf{A}_T(\boldsymbol{\alpha})\mathbf{y}_i = \mu_i \boldsymbol{\imath}_T + \mathbf{B}_T(\boldsymbol{\alpha})\mathbf{y}_i^0 + \mathbf{u}_i \ , \tag{14}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]'$. Accordingly, we obtain

$$\mathbf{y}_i = \mu_i \mathbf{A}_T(\boldsymbol{\alpha})^{-1}\boldsymbol{\imath}_T + \mathbf{A}_T(\boldsymbol{\alpha})^{-1}\mathbf{B}_T(\boldsymbol{\alpha})\mathbf{y}_i^0 + \mathbf{A}_T(\boldsymbol{\alpha})^{-1}\mathbf{u}_i \ .$$

A similar representation exists for the the vector of lags $\mathbf{y}_{-\ell,i} = [0, \ldots, 0, y_1, \ldots, y_{i,T-\ell}]'$:

$$\mathbf{y}_{-\ell,i} = \mu_i \mathbf{C}_{T,\ell}(\boldsymbol{\alpha})\boldsymbol{\imath}_T + \mathbf{R}_{T,\ell}(\boldsymbol{\alpha})\mathbf{y}_i^0 + \boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha})\mathbf{u}_i \ .$$

The matrix $\boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha})$ is defined similar to $\boldsymbol{\Psi}_{T,0}(\boldsymbol{\alpha}) = \mathbf{A}_T(\boldsymbol{\alpha})^{-1}$, dropping the first column and expanding it with a column of zeros from the right. The matrices $\mathbf{C}_{T,\ell}(\boldsymbol{\alpha})$ and $\mathbf{R}_{T,\ell}(\boldsymbol{\alpha})$ are not important as they refer to terms that are uncorrelated with $u_{i1}, \ldots, u_{iT}$.

Let $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\imath}_T\boldsymbol{\imath}_T'$. The expectation of the first-order conditions results as

$$\mathbb{E}(\mathbf{y}_{-\ell,i}'\mathbf{M}_T\mathbf{u}_i) = \sigma^2 tr(\boldsymbol{\Psi}_{T,\ell}\mathbf{M}_T) \ = \ -\frac{\sigma^2}{T}\boldsymbol{\imath}_T'\boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha})\boldsymbol{\imath}_T \ .$$

That is σ^2/T times the sum of all elements of $\boldsymbol{\Psi}_{T,\ell}$. For example, consider $p = 1$ and $T = 4$. For this case we obtain

$$\boldsymbol{\Psi}_{T,0}(\alpha) \ = \ \mathbf{A}_T(\alpha)^{-1} \ = \ \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ \alpha^2 & \alpha & 1 & 0 \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Psi}_{T,1}(\alpha) \ = \ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ \alpha^2 & \alpha & 1 & 0 \end{bmatrix},$$

and therefore, the sum of the elements of $\boldsymbol{\Psi}_{T,1}(\alpha)$ equal to $\sum_{t=0}^{2}\sum_{s=0}^{t}\alpha^s$, as required. An alternative – for our purpose less convenient – interpretation of the elements of the matrix $\mathbf{A}_T(\boldsymbol{\alpha})^{-1}$ is obtained from the moving-average representation of the AR(p) model. Let

$$y_{it} - \mathbb{E}(y_{it}|\mu_i, \mathbf{x}_{it}, x_{i,t-1}, \ldots) \ = \ \phi_0 u_{it} + \phi_1 u_{i,t-1} + \phi_2 u_{i,t-2} + \ldots,$$

with $\phi_0 = 1$. For $j \geq k$ the (j,k)-elements of the matrix $\mathbf{A}_T(\boldsymbol{\alpha})^{-1}$ are equal to ϕ_{j-k}. All other elements are equal to zero.

The moments for the bias-corrected estimator in the AR(p) model result as

$$m_{Ti,\alpha_\ell}(\boldsymbol{\alpha}) = \sum_{t=1}^{T} \left(y_{i,t-\ell} + \frac{b_{T,\ell}(\boldsymbol{\alpha})}{T-1} e_{it} \right) (e_{it} - \overline{e}_i), \tag{15}$$

where $b_{T,\ell}(\boldsymbol{\alpha}) = (\boldsymbol{\imath}_T' \boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha}) \boldsymbol{\imath}_T)/T$. For the model with strictly exogenous variables $\mathbf{x}_{it}$, the complete set of moments for the parameter vector $\boldsymbol{\theta} = [\alpha', \boldsymbol{\beta}']'$ results as

$$\mathbf{m}_{Ti}(\boldsymbol{\theta}) = \begin{bmatrix} m_{Ti,\alpha_1}(\boldsymbol{\theta}) \\ \vdots \\ m_{Ti,\alpha_p}(\boldsymbol{\theta}) \\ \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) \end{bmatrix} = \sum_{t=1}^{T} \mathbf{w}_{it}(\boldsymbol{\theta})(e_{it} - \overline{e}_i), \tag{16}$$

where $\mathbf{m}_{Ti,\beta}(\boldsymbol{\theta})$ is defined as in equation (7), and

$$\mathbf{w}_{it}(\boldsymbol{\theta}) = \begin{bmatrix} y_{i,t-1} + \frac{b_{T,1}(\boldsymbol{\alpha})}{T-1} e_{it} \\ \vdots \\ y_{i,t-p} + \frac{b_{T,p}(\boldsymbol{\alpha})}{T-1} e_{it} \\ \mathbf{x}_{it} \end{bmatrix}.$$

Note that for $T \to \infty$ and $t/T \to c > 0$ it follows that

$$\begin{aligned} \mathbb{E}(y_{i,t-\ell} \overline{u}_i) &= -\frac{\sigma^2}{T} \left(\sum_{j=0}^{\infty} \phi_j \right) \\ &= -\frac{\sigma^2}{T(1 - \alpha_1 - \cdots - \alpha_p)}, \end{aligned}$$

and therefore it follows for $T \to \infty$ that

$$b_{T,\ell}(\alpha) \to -\frac{1}{1 - \alpha_1 - \cdots - \alpha_p}.$$

The bias-corrected method of moments estimator results from solving $\sum_{i=1}^{N} \mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$. The computational details are similar to the case of an AR(1) model. The required derivatives $\partial b_{T,\ell}(\alpha)/\partial \alpha_\ell$ with $\ell = 1, \ldots, p$ are presented in Appendix A.

5 Cross-sectional dependence

In many macroeconomic applications, it is reasonable to assume the elements of the error vector $\mathbf{u}_t = [u_{1t}, \ldots, u_{Nt}]'$ are correlated such that

$$\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma}_{u,t}, \qquad t = 1, ..., T.$$

Although the bias-corrected estimator remains consistent under cross correlation, the estimator of the covariance matrix considered in Remark 1 is biased as

$$\frac{1}{NT}\mathbb{E}\left[\left(\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})\right)\left(\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})'\right)\right] \neq \frac{1}{NT}\mathbb{E}\left(\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})'\right)$$

in general. For the asymptotic covariance matrix we need to estimate the expression on the left-hand side consistently. The moments of the bias-corrected estimator can be rewritten as

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{m}_{Ti}(\boldsymbol{\theta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{w}_{it}(\boldsymbol{\theta})(e_{it}-\bar{e}_i) = \frac{1}{NT}\sum_{t=1}^{T}\mathbf{W}_t(\boldsymbol{\theta})'(\mathbf{u}_t-\bar{\mathbf{u}}),$$

where $\mathbf{W}_t(\boldsymbol{\theta}) = [\mathbf{w}_{1t}(\boldsymbol{\theta}),\ldots,\mathbf{w}_{Nt}(\boldsymbol{\theta})]'$, $\mathbf{u}_t = [u_{1t},\ldots,u_{Nt}]'$, and $\bar{\mathbf{u}} = [\bar{u}_1,\ldots,\bar{u}_N]'$. Since the vector $\mathbf{u}_t$ is independent across t, we have as $N\to\infty$ and $N/T\to\kappa$

$$\mathbf{S}(\boldsymbol{\theta}_0) = \lim_{N,T\to\infty}\mathbb{E}\left[\frac{1}{NT}\sum_{t=1}^{T}\mathbf{W}_t(\boldsymbol{\theta}_0)'\boldsymbol{\Sigma}_{u,t}\mathbf{W}_t(\boldsymbol{\theta}_0)\right] = \lim_{N,T\to\infty}\mathbb{E}\left[\frac{1}{NT}\sum_{t=1}^{T}\mathbf{Z}_t'\boldsymbol{\Sigma}_{u,t}\mathbf{Z}_t\right],$$

where $\mathbf{Z}_t = [\mathbf{z}_{1t}',\ldots,\mathbf{z}_{Nt}']'$ and $\mathbf{z}_{it}$ is defined in Theorem 1. Note that the bias correction does not affect the asymptotic covariance matrix as $T\to\infty$. To obtain a robust estimator of the covariance matrix, we may use the cluster-robust estimator

$$(NT)\,\widehat{\mathbf{S}}_{NT}^{rob}(\widehat{\boldsymbol{\theta}}_{bc}) = \sum_{t=1}^{T}\mathbf{W}_t(\widehat{\boldsymbol{\theta}}_{bc})'\widetilde{\mathbf{e}}_t\widetilde{\mathbf{e}}_t'\mathbf{W}_t(\widehat{\boldsymbol{\theta}}_{bc}) \tag{17}$$

instead of $\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc})'$ in Remark 1, where $\widetilde{\mathbf{e}}_t$ denotes the $N\times 1$ vector of mean-adjusted residuals

$$\widetilde{e}_{it} = (y_{it}-\bar{y}_i) - \widehat{\alpha}_{bc}(y_{i,t-1}-\bar{y}_{-1,i}) - \widehat{\boldsymbol{\beta}}_{bc}'(\mathbf{x}_{it}-\bar{\mathbf{x}}_i).$$

As the following theorem shows, the resulting estimator of the asymptotic covariance matrix is consistent in the case of weakly cross-sectionally dependent errors.

Theorem 2. *Let y_{it} be generated by the dynamic panel data model* (4) *with $E(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}_{u,t}$. If the largest eigenvalues of $\boldsymbol{\Sigma}_{u,1},\ldots,\boldsymbol{\Sigma}_{u,T}$ are bounded in N, then as $N\to\infty$ and $N/T\to\kappa<\infty$*

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{bc}-\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\mathbf{D}(\boldsymbol{\theta}_0)^{-1}\mathbf{S}(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\theta}_0)^{-1}\right) \tag{18}$$

where $\widehat{\mathbf{S}}_{NT}^{rob}(\widehat{\boldsymbol{\theta}}_{bc}) \xrightarrow{p} \mathbf{S}(\boldsymbol{\theta}_0)$ and $T^{-1}\mathbf{D}_T(\widehat{\boldsymbol{\theta}}_{bc}) \xrightarrow{p} \mathbf{D}(\boldsymbol{\theta}_0)$.

Remark 5. It is easy to see that the robust cluster approach runs into difficulties if the error dependence is due to common factors. Assume that $u_{it} = \lambda_i f_t + \varepsilon_{it}$, where f_t and ε_{it} are i.i.d. sequences with $\mathbb{E}(f_t^2) = \sigma_f^2$ and $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$. Accordingly, the error covariance matrix is

$$\boldsymbol{\Sigma}_{u,t} = \boldsymbol{\Sigma}_u \;=\; \sigma_f^2 \boldsymbol{\lambda}\boldsymbol{\lambda}' + \sigma_\varepsilon^2 \mathbf{I}_N$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]'$. Since

$$\frac{1}{NT}\sum_{t=1}^{T} \mathbf{W}_t(\boldsymbol{\theta})'\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathbf{W}_t(\boldsymbol{\theta}) = O_p(N),$$

in general, the estimator $\widehat{\boldsymbol{\theta}}_{bc}$ is no longer $\sqrt{NT}$-consistent (cf. Breitung, 2015, Sec. 15.4.3).

6 Small-sample properties

6.1 Data-generating processes

To compare the small-sample properties of the bias-corrected method of moments estimator to alternative estimators suggested in the literature, some Monte Carlo experiments are performed. The data are generated from a slightly simplified homoskedastic version of the dynamic panel data model used by Kiviet et al. (2017) in their simulation exercise:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \sigma_\mu \mu_i + \sigma_u u_{it}, \\ x_{it} &= \gamma x_{i,t-1} + \pi_\mu \mu_i + \pi_\lambda \lambda_i + \sigma_\epsilon \epsilon_{it}. \end{aligned}$$

The regressor x_{it} is strictly exogenous with respect to the idiosyncratic error term u_{it}. The errors u_{it} and ϵ_{it} and the individual-specific effects μ_i and λ_i are drawn from independent standard normal distributions. Following Kiviet et al. (2017), we choose the remaining free parameters to obtain a reasonable characterization of the data-generating process. More details are available in the Online Appendix.

We distinguish between a process with moderate persistence, $\alpha = 0.4$, and high persistence, $\alpha = 0.9$. The process is initialized at $t = -50$ with $y_{i,-50} = x_{i,-50} = 0$, and the first 50 observations are discarded. As a robustness check, we also consider the initialization without burn-in period, $y_{i0} = x_{i0} = 0$, which implies that the observed process starts off its stationary path.

To analyze the estimators' performance under higher-order dynamics, we modify the above data-generating process as follows:

$$y_{it} = \sum_{j=1}^{3} \alpha_j y_{i,t-j} + \beta x_{it} + \sigma_\mu \mu_i + \sigma_u u_{it}.$$

We set $(\alpha_1, \alpha_2, \alpha_3) = (0.48, -0.2, 0.12)$ to achieve $\sum_{j=1}^{3} \alpha_j = 0.4$ and $(\alpha_1, \alpha_2, \alpha_3) = (1.08, -0.45, 0.27)$ to obtain $\sum_{j=1}^{3} \alpha_j = 0.9$. All other parameter values are left unchanged.

A data-generating process with heteroskedasticity across both dimensions is obtained by replacing u_{it} with

$$u_{it} = \sqrt{\frac{3}{4}} \delta_i \tau_t v_{it},$$

where τ_t and v_{it} are independent standard normally distributed, and δ_i is uniformly distributed over the interval $(0, 2)$. To analyze the estimators' performance under cross-sectional error dependence, we consider the following modifications of the data-generating process, where v_{it} in all specifications is independent standard normally distributed and the parameterizations ensure that $Var(u_{it}) = 1$ to keep the signal-to-noise ratio unaffected:

1. Uniform cross-sectional dependence:

   $$u_{it} = \sqrt{\frac{3}{4N}} \sum_{j=1}^{N} \omega_{ij} v_{jt},$$

   with independent uniformly distributed spatial weights ω_{ij} over the interval $(0, 2)$.

2. Interactive random effects:

   $$u_{it} = \sqrt{\frac{3}{7}} (\delta_i \tau_t + v_{it}),$$

   with independent standard normally distributed common factors τ_t and uniformly distributed factor loadings δ_i over the interval $(0, 2)$.

6.2 Simulation results

We compare the performance of the within-groups estimator (WG), our bias-corrected estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. In addition to the average bias and root mean square error (RMSE), we report the empirical size of Wald tests given a nominal size of 5%. We consider a fixed-T robust variance-covariance estimator clustered at the individual level and, for the WG and BC estimators in the simulation designs with cross-sectional dependence, a large-T robust variance-covariance estimator clustered at the time periods. For the AS-GMM and the BB-GMM estimators, the finite-sample Windmeijer (2005) correction is applied.

We consider all sample size combinations of $T \in \{5, 10, 25, 50\}$ and $N \in \{50, 200\}$. The results are based on 1,000 replications for each simulation design. For the BC estimator, we apply the following procedure to deal with the problem that $b_T(\alpha)$ is a higher-order polynomial in α such that the moment function $m_{Ti,\alpha}(\boldsymbol{\theta})$ may have multiple roots, possibly even in the stationary parameter region. Although this problem disappears as T becomes large, for samples with T as small as 5 or 10, we observe a small fraction of estimates that are far away from the true value. If the moment functions are interpreted as profile scores from an adjusted profile likelihood function, the maximum must obey the second-order condition that $\nabla_{\boldsymbol{\theta}} \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}_{bc})$ is negative definite.[1] In our simulations, we check whether the largest eigenvalue of $\nabla_{\boldsymbol{\theta}} \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}_{bc})$ is negative. If this condition is violated, we re-initialize the search algorithm with a random draw for α from the uniform distribution over the interval $(0, 1)$. If necessary, we repeat this process until a solution is found that satisfies the condition of a negative-definite gradient of the moment function. In our experience, this refinement effectively prevents finding an inappropriate solution for the moment conditions.

As detailed results can be found in the Online Appendix, we sketch the main results in the following:

[1] See Dhaene and Jochmans (2016) and Juodis (2018) for a related discussion on multiple solutions of the adjusted profile score and the ML estimators in autoregressive models with fixed effects.

1. The results for the AR(1) model with stationary initial conditions and i.i.d. errors are presented in Tables 1–2 in the Online Appendix. With respect to bias and RMSE, our bias-corrected estimator (BC) performs very similar to the transformed QML estimator proposed by Hsiao et al. (2002). This is expected from our results in Section 3. One should bear in mind, however, that the BC estimator does not require a specific assumption on the initial value of the dynamic process. Furthermore, our estimator is computationally much simpler and can be computed within a fraction of computing time that is required for the QML estimator (in particular if T gets large). All GMM based estimators perform substantially worse.

2. Tables 3–4 present results for autoregressive processes with $p = 3$ lags. As no ML based estimation procedures are readily available for higher-order autoregressive processes, we compare our estimator to the existing GMM estimation procedures. The results suggest that the bias correction effectively removes any bias from the estimator and yields estimates with the lowest RMSE. The only exception is the case with high persistence ($\alpha = 0.9$) and small T ($T \leq 10$), where the system GMM estimator (BB-GMM) performs best.

3. Tables 5–6 display the results for first-order dynamic models with cross-sectional and time-dependent heteroskedasticity. In Remark 3 we argued that the BC estimator is robust against cross-sectional heteroscedasticity for all N and T, whereas robustness against heteroskedasticity across time requires large T. The findings of our Monte Carlo experiments suggest that for moderate persistence the BC estimator is robust against both forms of heteroskedasticity even if T is as small as 10. For processes with higher persistence a larger number of time periods ($T \geq 25$) is required to remove the bias from the estimator. Overall, the BC estimator performs similarly to the QML estimator and both estimators clearly outperform the other GMM estimators in most cases. (Only the system GMM estimator performs better when the persistence is large and T is small.)

4. The performance under cross-sectional dependence is studied in Tables 7–10. The results suggest that inference based on the i.i.d. assumption may

be severely biased whenever the errors are cross-sectionally dependent. For the model with uniform cross-sectional dependence, the significance tests with nominal size of 5 percent reject in more than 50 percent of the cases and sometimes the empirical sizes even exceed 0.9. On the other hand, the cluster-robust standard errors considered in Theorem 2 effectively correct for cross-sectional dependence and yield empirical sizes close to the nominal sizes. The robustification works well even in models where cross-sectional dependence is due to common factors, although the asymptotic theory underlying Theorem 2 does not apply to models with strong dependence.

5. We finally study the small-sample properties of the estimators when the initial condition affects the distribution of the dependent variable (nonstationary initialization). It is well known that in this case the BB-GMM estimator can be severely biased. This is confirmed by our Monte Carlo simulations presented in Tables 11–12. All other estimators are virtually unbiased whenever N is sufficiently large. While the BC and QML estimators perform similarly in the case of i.i.d. errors, the BC estimator turns out to be more efficient than the QML estimator in the model with nonstationary initialization.

7 Conclusion

In this paper, we proposed a simple asymptotically efficient estimator for dynamic panel data models. The estimator is related to the bias-corrected profile score estimator of Dhaene and Jochmans (2016) but results in a much simpler set of moment conditions that can easily be solved with standard methods. An important advantage of our bias-corrected estimator is that an estimator of the asymptotic covariance matrix is readily available and the estimator can easily be generalized to higher order dynamic models. Furthermore, robust standard errors are available that effectively adjust for cross-sectional dependence, a relevant feature in the analysis of macroeconomic panel data. We show that our estimator is related to the QML estimator of Hsiao et al. (2002), where the bias-correction is obtained in a different (and computationally more demanding) manner. Monte

Carlo simulations are carried out to investigate the finite-sample properties. The simulation results reveal that the bias-corrected method of moments estimator performs favorable compared to existing GMM and ML based estimation procedures.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*(1), 5–27.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*(375), 598–606.

Arellano, M. and S. R. 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.

Bai, J. (2013). Fixed-effects dynamic panel models, a factor analytical method. *Econometrica 81*(1), 285–314.

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

Breitung, J. (2015). The analysis of macroeconomic panel data. In B. H. Baltagi (Ed.), *The Oxford Handbook of Panel Data*, Chapter 15, pp. 453–493. Oxford: Oxford University Press.

Bun, M. J. G. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*(2), 200–210.

Bun, M. J. G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal 13*(1), 95–126.

Dhaene, G. and K. Jochmans (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32*(5), 1178–1215.

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. and M. H. Pesaran (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models. *Journal of Econometrics 188*(1), 111–134.

Hayakawa, K., M. Qi, and J. Breitung (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. *Econometric Reviews 38*(9), 1055–1088.

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

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

Judson, R. A. and A. L. Owen (1999). Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters 65*(1), 9–15.

Juodis, A. (2018). First difference transformation in panel VAR models: Robustness, estimation, and inference. *Econometric Reviews 37*(6), 650–693.

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., M. Pleus, and R. Poldermans (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics 5*(1), 14.

Moon, H. R., B. Perron, and P. C. B. Phillips (2015). Incidental parameters and dynamic panel modeling. In B. H. Baltagi (Ed.), *The Oxford Handbook of Panel Data*, Chapter 4, pp. 111–148. Oxford: Oxford University Press.

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

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126*(1), 25–51.

Appendix A: Computational details

The derivative of the moments $\mathbf{m}_{Ti}(\boldsymbol{\theta})$

Consider the matrix of derivatives

$$\nabla'_\theta \mathbf{m}_{Ti}(\boldsymbol{\theta}) = \begin{bmatrix} \nabla_\alpha m_{Ti,\alpha}(\boldsymbol{\theta}) & \nabla'_\beta m_{Ti,\alpha}(\boldsymbol{\theta}) \\ \nabla_\alpha \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) & \nabla'_\beta \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) \end{bmatrix}.$$

We obtain

$$\begin{aligned}
\nabla_\alpha m_{Ti,\alpha}(\boldsymbol{\theta}) &= -\left(\sum_{t=1}^{T} y_{i,t-1}(y_{i,t-1} - \overline{y}_{-1,i})\right) \\
&\quad + \frac{\nabla_\alpha b_T(\alpha)}{T-1}\sum_{t=1}^{T}(e_{it} - \overline{e}_i)^2 - \frac{2b_T(\alpha)}{T-1}\left(\sum_{t=1}^{T} y_{i,t-1}(e_{it} - \overline{e}_i)\right), \\
\mathbb{E}\left[\nabla_\alpha m_{Ti,\alpha}(\boldsymbol{\theta})\right] &= \mathbb{E}\left[-\left(\sum_{t=1}^{T} y_{i,t-1}(y_{i,t-1} - \overline{y}_{-1,i})\right)\right] + \sigma^2 \nabla_\alpha b_T(\alpha) + \frac{2\sigma^2}{T-1} b_T(\alpha)^2,
\end{aligned}$$

where $\nabla_\alpha b_T(\alpha) = T^{-1}\sum_{t=1}^{T-2}\sum_{s=1}^{t} s\alpha^{s-1}$. Furthermore,

$$\begin{aligned}
\nabla'_\beta m_{Ti,\alpha}(\boldsymbol{\theta}) &= -\sum_{t=1}^{T} y_{i,t-1}(x_{it} - \overline{x}_i)' - 2\left(\frac{b_T(\alpha)}{T-1}\sum_{t=1}^{T}(e_{it} - \overline{e}_i)x'_{it}\right), \\
\mathbb{E}\left[\nabla'_\beta m_{Ti,\alpha}(\boldsymbol{\theta})\right] &= -\mathbb{E}\left[\sum_{t=1}^{T} y_{i,t-1}(x_{it} - \overline{x}_i)'\right], \\
\nabla_\alpha \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}(x_{it} - \overline{x}_i) y_{i,t-1}, \\
\nabla'_\beta \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}) &= -\sum_{t=1}^{T}(x_{it} - \overline{x}_i) x'_{it}.
\end{aligned}$$

The derivatives of $b_{T,\ell}(\boldsymbol{\alpha})$

First, we consider the derivative

$$\frac{\partial \mathbf{A}_T(\boldsymbol{\alpha})^{-1}}{\partial \alpha_s} = -\mathbf{A}_T(\boldsymbol{\alpha})^{-1}\frac{\partial \mathbf{A}_T(\boldsymbol{\alpha})}{\partial \alpha_s}\mathbf{A}_T(\boldsymbol{\alpha})^{-1},$$

where for $j-k=s$, $j>k$, the (j,k)-th element of the matrix $\partial \mathbf{A}_0(\boldsymbol{\alpha})/\partial \alpha_s$ is equal to -1, whereas all other elements are zero. From this matrix we obtain the matrix $\partial \boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha})/\partial \alpha_s$ by expanding from the right ℓ columns of zeros and dropping the first ℓ columns of $\partial \mathbf{A}_T(\boldsymbol{\alpha})^{-1}/\partial \alpha_s$. This gives

$$\frac{\partial b_{T,\ell}(\boldsymbol{\alpha})}{\partial \alpha_s} = \frac{1}{T}\boldsymbol{\imath}_T' \left(\frac{\partial \boldsymbol{\Psi}_{T,\ell}(\boldsymbol{\alpha})}{\partial \alpha_s} \right) \boldsymbol{\imath}_T \ .$$

The expectation of $\widetilde{b}_T(\alpha, \omega)$

Treating the variance parameters σ^2 and ω as known, the first derivative becomes

$$\frac{\partial \ell(\alpha, \sigma^2, \omega)}{\partial \alpha} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t-1} - \bar{y}_{i,-1})(u_{it} - \bar{u}_i) + \frac{T}{\sigma^2 |\boldsymbol{\Omega}|} \sum_{i=1}^{N} (y_{i0} - \bar{y}_{i,-1})(u_{i0} - \bar{u}_i).$$

We rewrite $b_T(\alpha) = T^{-1} \sum_{t=0}^{T-2} \sum_{s=0}^{t} \alpha^s$ such that

$$\begin{aligned}
b_T(\alpha) &= \frac{1}{1-\alpha}\left(1 - \frac{1-\alpha^T}{T(1-\alpha)}\right) \\
b_T(\alpha) &= \frac{T}{|\boldsymbol{\Omega}|-1}\left(\frac{\omega-1}{1-\alpha} - \frac{(\omega-1)(1-\alpha^T)}{T(1-\alpha)^2}\right) \\
\left(1 - \frac{1}{|\boldsymbol{\Omega}|}\right) b_T(\alpha) &= \frac{T}{|\boldsymbol{\Omega}|}\left(\frac{\omega-1}{1-\alpha} - \frac{(\omega-1)(1-\alpha^T)}{T(1-\alpha)^2}\right) \\
b_T(\alpha) &= \frac{T}{|\boldsymbol{\Omega}|}\left(\frac{\omega-1}{1-\alpha} - \frac{(\omega-1)(1-\alpha^T)}{T(1-\alpha)^2} + \frac{b_T(\alpha)}{T}\right).
\end{aligned}$$

Inserting these results in the expectation of the scores yields

$$\mathbb{E}\left[\frac{1}{N}\frac{\partial \ell(\alpha, \sigma^2, \omega)}{\partial \alpha}\right] = \frac{1}{\sigma^2}\left(-b_T(\alpha)\sigma^2\right) + \frac{T}{\sigma^2|\boldsymbol{\Omega}|}\left(\frac{(\omega-1)\sigma^2}{1-\alpha} - \frac{(\omega-1)\sigma^2(1-\alpha^T)}{T(1-\alpha)^2} + \frac{b_T(\alpha)\sigma^2}{T}\right) = 0.$$

Appendix B: Proofs

Proof of Theorem 1

(i) A first-order Taylor series expansion of the moment conditions yields

$$\mathbf{0} = \sum_{i=1}^{N} \mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}}_{bc}) \ \simeq \ \sum_{i=1}^{N} \mathbf{m}_{Ti}(\boldsymbol{\theta}_0) + \left(\sum_{i=1}^{N} \nabla_\theta' \mathbf{m}_{Ti}(\boldsymbol{\theta}_0)\right)(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0)$$

and

$$(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \simeq -\left(\sum_{i=1}^{N} \nabla'_\theta \mathbf{m}_{Ti}(\boldsymbol{\theta}_0)\right)^{-1} \sum_{i=1}^{N} \mathbf{m}_{Ti}(\boldsymbol{\theta}_0).$$

Since $\mathbf{m}_{T1,\alpha}(\boldsymbol{\theta}_0), \ldots, \mathbf{m}_{TN,\alpha}(\boldsymbol{\theta}_0)$ are independent with expectation zero, the central limit theorem yields

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{m}_{Ti}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}_T(\boldsymbol{\theta}_0)).$$

From the results for the derivative $\nabla'_\theta \mathbf{m}_{Ti}(\boldsymbol{\theta})$ in Appendix A, we obtain

$$\frac{1}{NT} \sum_{i=1}^{N} \nabla'_\theta \mathbf{m}_{Ti}(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{D}_T(\boldsymbol{\theta}_0).$$

Hence, the limiting distribution results as stated in (i).

(ii) Let $y_{it} = \xi_{it} + v_{it}$, where $v_{it} = \sum_{j=0}^{t-1} \alpha_0^j u_{i,t-j}$ and ξ_{it} contains the current and past values of regressors $\mathbf{x}_{it}$, individual effects and initial conditions y_{i0}. Then, after some algebra, we have

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} m_{Ti,\alpha}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} (y_{i,t-1} - \overline{\xi}_{-1,i}) u_{it} \right) - \sqrt{\frac{N}{T}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} v_{i,t-1} u_{is} \right) \right] + \sqrt{\frac{N}{T}} b_T(\alpha_0) \frac{1}{N} \sum_{i=1}^{N} s_i^2,$$

where $(T-1)^{-1} s_i^2 = \sum_{t=1}^{T} u_{it}(u_{it} - \bar{u}_i)$. Since $\mathbb{E}\left(T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} v_{i,t-1} u_{is}\right) = b_T(\alpha_0)\sigma^2$ and $\mathbb{E}(s_i^2) = \sigma^2$, it follows that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} m_{Ti,\alpha}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t-1} - \overline{\xi}_{-1,i}) u_{it} \ + \ o_p(1).$$

Furthermore,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} (\mathbf{x}_{it} - \overline{\mathbf{x}}_i) u_{it} \right).$$

Since $\mathbb{E}(\overline{y}_{-1,i}) = \mathbb{E}(\overline{\xi}_{-1,i})$ and by using the fact that $(y_{i,t-1} - \overline{\xi}_{-1,i})$ and $(\mathbf{x}_{it} - \overline{\mathbf{x}}_i)$ are uncorrelated with u_{it} we obtain

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{m}_{Ti}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{V}_z).$$

From the results of Appendix A it follows that

$$\frac{1}{NT}\sum_{i=1}^{N}\nabla_{\alpha}m_{Ti,\alpha}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}(y_{i,t-1}-\overline{y}_{-1,i}) + O_p(T^{-1}) - O_p(T^{-1}),$$

$$\frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\beta}\mathbf{m}_{Ti,\alpha}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}(\mathbf{x}_{it}-\overline{\mathbf{x}}_i)' - O_p(T^{-1}),$$

$$\frac{1}{NT}\sum_{i=1}^{N}\nabla_{\alpha}\mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\overline{\mathbf{x}}_i)y_{i,t-1},$$

$$\frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\beta}\mathbf{m}_{Ti,\beta}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it}-\overline{\mathbf{x}}_i)\mathbf{x}'_{it},$$

and, therefore, as N and T tend to infinity,

$$\frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\theta}\mathbf{m}_{Ti}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{z}_{it}-\overline{\mathbf{z}}_i)(\mathbf{z}_{it}-\overline{\mathbf{z}}_i)' + O_p(T^{-1}) \xrightarrow{p} -\mathbf{V}_z.$$

With these results the limiting distribution of $\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{bc}-\boldsymbol{\theta}_0)$ follows as $\mathcal{N}(0, \sigma^2\mathbf{V}_z^{-1})$.

Proof of Theorem 2

First note that

$$\mathbf{w}_{it}(\widehat{\boldsymbol{\theta}}_{bc}) = \mathbf{z}_{it} + \frac{1}{T-1}b_T\left(\widehat{\alpha}_{bc}\right)\boldsymbol{\varphi}_{k+1}\widetilde{e}_{it},$$

where $\boldsymbol{\varphi}_{k+1}$ denotes the first column of the identity matrix $\boldsymbol{I}_{k+1}$. It follows that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}})\mathbf{m}_{Ti}(\widehat{\boldsymbol{\theta}})'\right] &= \frac{1}{NT}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Z}'_t\mathbf{u}_t\mathbf{u}'_t\mathbf{Z}'_t\right] + o_p(1)\\ &\xrightarrow{p}\mathbf{S}(\boldsymbol{\theta}_0).\end{aligned}$$

Define $\mathbf{q}_{N,t} = N^{-1/2}\sum_{i=1}^{N}\mathbf{z}_{it}u_{it}$ such that

$$\mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{m}_{Ti}(\boldsymbol{\theta}_0)\mathbf{m}_{Ti}(\boldsymbol{\theta}_0)'\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{q}_{N,t}\mathbf{q}'_{N,t}\right).$$

Since u_{it} is independent of $\mathbf{z}_{it}$, it follows that $\mathbb{E}(\mathbf{q}_{N,t}) = \mathbf{0}$. Furthermore, if all eigenvalues of $\boldsymbol{\Sigma}_{u,t}$ are bounded for all t, the norm of the covariance matrix $\mathbf{q}_{N,t}$ is bounded as well and

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\mathbf{q}_{N,t}\mathbf{q}'_{N,t}\right) = \mathbf{S}(\boldsymbol{\theta}_0).$$

Online Appendix: Monte Carlo simulation

Details of the data-generating process

By normalizing the long-run variance of x_{it} to unity, we can obtain the parameters π_μ, π_λ, and σ_ϵ^2 by fixing the fraction of the variance of x_{it} that is due to the individual-specific effects (EVF_x) and the variance fraction of $\pi_\mu \mu_i$ in the composite individual-specific effects (IEF_x^μ). With the normalization $\sigma_u^2 = 1$ and by further fixing the direct cumulated effect of μ_i on y_{it} relative to the noise (DEN_y^η) and the signal-to-noise ratio (SNR), we can then pin down σ_μ and β. Following Kiviet et al. (2017), we hold fixed the fraction of the variance of x_{it} that is due to the individual-specific effects (EVF_x) and the variance fraction of $\pi_\mu \mu_i$ in the composite individual-specific effects (IEF_x^μ). This allows us to endogenously determine

$$
\begin{aligned}
\pi_\mu &= (1-\gamma)\sqrt{EVF_x IEF_x^\mu}, \\
\pi_\lambda &= (1-\gamma)\sqrt{EVF_x(1-IEF_x^\mu)}, \\
\sigma_\epsilon^2 &= \sqrt{(1-\gamma^2)(1-EVF_x)}.
\end{aligned}
$$

By further fixing the direct cumulated effect of μ_i on y_{it} relative to the noise (DEN_y^η) and the signal-to-noise ratio (SNR), we obtain[2]

$$
\begin{aligned}
\sigma_\mu &= (1-\alpha)DEN_y^\eta, \\
\beta &= \sqrt{\frac{(1-\alpha\gamma)(SNR-\alpha^2(1+SNR))}{(1+\alpha\gamma)(1-EVF_x)}}.
\end{aligned}
$$

Fixing $\gamma = 0.4$, $EVF_x = IEF_x^\mu = 0.3$, $DEN_y^\eta = 4$, and $SNR = 5$ implies $\beta \approx 2.044$ under moderate persistence and $\beta \approx 0.307$ in the high-persistence case.

Construction of the estimators

Besides the within-groups estimator (WG) and our bias-corrected estimator (BC), we consider the following three generalized method of moments estimators:

[2] See Kiviet et al. (2017, Section 4) for details.

- The one-step Arellano and Bond (1991) GMM estimator (AB-GMM) utilizes the moment conditions $E[y_{i,t-s}\Delta e_{it}] = 0$, $2 \leq s \leq 4$, $t = s, \ldots, T$, and $E[\sum_{t=2}^{T} \Delta x_{it}\Delta e_{it}] = 0$.[3] With a weighting matrix that is optimal under homoskedasticity and absence of serial correlation in the idiosyncratic error component, the one-step GMM estimator equals the two-stage least squares estimator.

- The two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM) utilizes the nonlinear moment conditions $E[e_{iT}\Delta e_{it}]$, $t = 2, \ldots, T-1$, in addition to the moment conditions of the AB-GMM estimator and an intercept. The optimal weighting matrix is computed based on the residuals from an inefficient estimator with block-diagonal weighting matrix. The block corresponding to the AB-GMM moment conditions is identical to the AB-GMM weighting matrix. The block corresponding to the nonlinear moment conditions is an identity matrix.

- The two-step Blundell and Bond (1998) GMM estimator (BB-GMM) utilizes the moment conditions $E[\Delta y_{i,t-1}e_{it}]$, $t = 2, \ldots, T$, in addition to the moment conditions of the AB-GMM estimator and an intercept. The optimal weighting matrix is computed based on the residuals from the inefficient two-stage least squares estimator.

We use the same moment conditions under AR(1) and AR(3) dynamics. In the simulations with AR(1) dynamics, we further consider the Hsiao et al. (2002) quasi-maximum likelihood (QML) estimator with the initial-observations parameterization $E[\Delta y_{i1}|\Delta x_{i1}, \ldots, \Delta x_{iT}] = b + \sum_{s=1}^{3} \pi_s \Delta x_{is}$.

[3] We do not use all available lags of the dependent variable to avoid instrument proliferation. As demonstrated by Hayakawa et al. (2019), 3 lags are sufficient to achieve a reasonable degree of efficiency.

Estimation of the variance-covariance matrix

We consider the following two estimators for the variance-covariance matrix:

$$\hat{\mathbf{V}}_{NT}(\theta) = \frac{1}{NT}\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{S}_{NT}(\hat{\boldsymbol{\theta}})\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}, \tag{19}$$

$$\hat{\mathbf{V}}^{csd}_{NT}(\theta) = \frac{1}{NT}\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{S}^{csd}_{NT}(\hat{\boldsymbol{\theta}})\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}. \tag{20}$$

Variance-covariance estimator (19) is the conventional cluster-robust estimator clustered at the individual level. Estimator (20) is a robust estimator clustered at the time periods. For the WG estimator, we have

$$\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\theta}\mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}}), \quad \mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}})\mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}})',$$

$$\mathbf{S}^{csd}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{t=1}^{T}(\mathbf{y}_{t-1}, \mathbf{X}_t)'\tilde{\mathbf{u}}_t\tilde{\mathbf{u}}'_t(\mathbf{y}_{t-1}, \mathbf{X}_t).$$

Similarly, for the BC estimator we compute

$$\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\theta}\mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}), \quad \mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}})\mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}})',$$

$$\mathbf{S}^{csd}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{NT}\sum_{t=1}^{T}\mathbf{W}_t(\hat{\boldsymbol{\theta}})'\tilde{\mathbf{u}}_t\tilde{\mathbf{u}}'_t\mathbf{W}_t(\hat{\boldsymbol{\theta}}).$$

For the GMM estimators with moment functions $\tilde{\mathbf{m}}_{Ti}(\boldsymbol{\theta})$ and weighting matrix $\mathbf{W}_{NT}$, we get

$$\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\nabla_{\theta}\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})\right)\mathbf{W}_{NT}\left(\frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\theta}\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})\right),$$

$$\mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\nabla_{\theta}\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})\right)\mathbf{W}_{NT}\left(\frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})'\right)\mathbf{W}_{NT}\left(\frac{1}{NT}\sum_{i=1}^{N}\nabla'_{\theta}\tilde{\mathbf{m}}_{Ti}(\hat{\boldsymbol{\theta}})\right).$$

For the two-step AS-GMM and BB-GMM estimators, the variance matrix is estimated as $\hat{\mathbf{V}}^{(1)}_{NT}(\theta) = \frac{1}{NT}\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}$ and the Windmeijer (2005) finite-sample correction is applied. For the QML estimator, $\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})$ is the negative Hessian matrix and $\mathbf{S}_{NT}(\hat{\boldsymbol{\theta}})$ the outer product of the gradient.

Table 1: Simulation results: baseline model (IID), $N = 50$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 5$												
Bias	-0.077	0.001	-0.038	0.067	0.097	-0.003	-0.005	-0.001	-0.022	0.019	-0.025	-0.025
RMSE	0.086	0.041	0.117	0.150	0.140	0.041	0.093	0.093	0.134	0.148	0.142	0.094
$T = 10$												
Bias	-0.034	0.000	-0.020	0.109	0.129	-0.006	0.015	0.001	-0.008	0.038	-0.042	-0.018
RMSE	0.041	0.023	0.055	0.145	0.143	0.023	0.062	0.060	0.075	0.117	0.103	0.062
$T = 25$												
Bias	-0.014	-0.001	-0.008	0.094	0.142	-0.004	0.013	0.003	0.008	0.025	-0.038	-0.006
RMSE	0.019	0.013	0.023	0.108	0.147	0.013	0.037	0.035	0.044	0.080	0.067	0.035
$T = 50$												
Bias	-0.006	0.000	-0.005	0.078	0.134	-0.002	0.006	0.000	0.006	0.015	-0.043	-0.004
RMSE	0.011	0.009	0.014	0.087	0.138	0.009	0.026	0.025	0.031	0.058	0.059	0.025
	$\alpha = 0.9$											
$T = 5$												
Bias	-0.433	-0.034	-0.442	-0.044	0.018	-0.003	-0.048	-0.001	-0.058	-0.005	-0.008	-0.033
RMSE	0.438	0.124	0.561	0.205	0.091	0.156	0.102	0.095	0.124	0.125	0.122	0.094
$T = 10$												
Bias	-0.223	-0.004	-0.249	0.007	0.031	-0.022	-0.016	0.002	-0.031	0.002	-0.009	-0.039
RMSE	0.226	0.067	0.321	0.084	0.053	0.073	0.064	0.063	0.086	0.090	0.086	0.071
$T = 25$												
Bias	-0.085	0.000	-0.078	0.009	0.027	-0.018	0.004	0.003	0.002	0.013	0.003	-0.024
RMSE	0.087	0.025	0.103	0.046	0.037	0.026	0.034	0.033	0.046	0.055	0.049	0.041
$T = 50$												
Bias	-0.039	-0.001	-0.033	0.006	0.023	-0.009	0.004	0.001	0.005	0.007	-0.002	-0.017
RMSE	0.041	0.012	0.044	0.033	0.035	0.014	0.025	0.024	0.033	0.041	0.035	0.029

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 2: Simulation results: baseline model (IID), $N = 200$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 5$												
Bias	-0.079	-0.001	-0.008	0.026	0.024	-0.004	-0.002	0.000	-0.004	0.011	-0.002	-0.023
RMSE	0.081	0.021	0.058	0.065	0.054	0.021	0.046	0.046	0.063	0.066	0.060	0.051
$T = 10$												
Bias	-0.035	-0.001	-0.005	0.017	0.022	-0.007	0.014	0.000	-0.002	-0.001	-0.012	-0.020
RMSE	0.037	0.011	0.025	0.030	0.034	0.013	0.033	0.030	0.038	0.041	0.042	0.036
$T = 25$												
Bias	-0.013	0.000	-0.002	0.022	0.036	-0.004	0.011	0.001	0.002	0.001	-0.014	-0.008
RMSE	0.015	0.007	0.011	0.027	0.039	0.008	0.021	0.018	0.022	0.028	0.030	0.019
$T = 50$												
Bias	-0.006	0.000	-0.001	0.032	0.060	-0.002	0.005	0.000	0.001	0.005	-0.013	-0.005
RMSE	0.008	0.004	0.007	0.035	0.061	0.005	0.014	0.013	0.016	0.025	0.023	0.014
	$\alpha = 0.9$											
$T = 5$												
Bias	-0.430	-0.006	-0.191	0.000	0.001	-0.002	-0.048	0.000	-0.025	-0.002	-0.004	-0.033
RMSE	0.432	0.082	0.306	0.122	0.058	0.102	0.066	0.047	0.068	0.060	0.059	0.055
$T = 10$												
Bias	-0.221	0.004	-0.083	-0.001	0.008	-0.029	-0.016	0.000	-0.010	-0.004	-0.007	-0.038
RMSE	0.222	0.044	0.124	0.060	0.033	0.040	0.034	0.030	0.041	0.042	0.042	0.047
$T = 25$												
Bias	-0.084	0.000	-0.020	0.006	0.015	-0.016	0.002	0.001	0.001	0.001	-0.003	-0.026
RMSE	0.084	0.012	0.035	0.026	0.022	0.019	0.018	0.017	0.023	0.027	0.026	0.031
$T = 50$												
Bias	-0.039	0.000	-0.009	0.010	0.022	-0.009	0.003	0.000	0.001	0.003	0.000	-0.018
RMSE	0.040	0.006	0.017	0.019	0.024	0.010	0.012	0.012	0.017	0.019	0.018	0.022

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 3: Simulation results: higher-order dynamics, $N = 50$

	α					β				
	WG	BC	AB-GMM	AS-GMM	BB-GMM	WG	BC	AB-GMM	AS-GMM	BB-GMM
	$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.4$									
$T = 5$										
Bias	-0.143	0.000	-0.113	0.103	0.222	-0.038	-0.002	-0.062	0.036	0.026
RMSE	0.156	0.069	0.253	0.215	0.258	0.103	0.098	0.179	0.165	0.143
$T = 10$										
Bias	-0.056	0.001	-0.043	0.199	0.254	0.006	0.003	-0.017	0.064	0.012
RMSE	0.065	0.033	0.102	0.231	0.266	0.062	0.062	0.090	0.126	0.095
$T = 25$										
Bias	-0.020	0.000	-0.018	0.192	0.248	0.007	0.000	-0.002	0.050	0.013
RMSE	0.026	0.016	0.044	0.204	0.253	0.037	0.037	0.048	0.084	0.057
$T = 50$										
Bias	-0.009	0.000	-0.017	0.169	0.219	0.004	0.000	0.001	0.040	0.008
RMSE	0.015	0.011	0.031	0.179	0.225	0.026	0.025	0.033	0.063	0.042
	$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.9$									
$T = 5$										
Bias	-0.558	0.012	-0.588	-0.093	0.035	-0.063	-0.003	-0.074	-0.014	-0.006
RMSE	0.566	0.174	0.719	0.290	0.100	0.108	0.098	0.132	0.130	0.129
$T = 10$										
Bias	-0.282	0.019	-0.341	0.006	0.047	-0.024	0.004	-0.038	-0.004	-0.006
RMSE	0.287	0.093	0.426	0.094	0.067	0.064	0.061	0.089	0.090	0.087
$T = 25$										
Bias	-0.105	0.002	-0.135	0.016	0.031	-0.003	0.001	-0.009	0.007	0.000
RMSE	0.107	0.034	0.171	0.056	0.048	0.036	0.035	0.050	0.055	0.051
$T = 50$										
Bias	-0.048	0.000	-0.080	-0.003	0.014	0.000	-0.001	-0.001	0.007	-0.003
RMSE	0.050	0.014	0.096	0.057	0.053	0.024	0.024	0.034	0.041	0.036

Note: Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 4: Simulation results: higher-order dynamics, $N = 200$

	α					β				
	WG	BC	AB-GMM	AS-GMM	BB-GMM	WG	BC	AB-GMM	AS-GMM	BB-GMM
	$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.4$									
$T = 5$										
Bias	-0.141	0.000	-0.038	0.069	0.085	-0.035	0.000	-0.019	0.033	0.019
RMSE	0.144	0.034	0.128	0.136	0.125	0.059	0.049	0.088	0.091	0.072
$T = 10$										
Bias	-0.057	0.000	-0.012	0.059	0.085	0.003	0.000	-0.005	0.018	0.007
RMSE	0.059	0.016	0.051	0.087	0.102	0.030	0.030	0.042	0.053	0.044
$T = 25$										
Bias	-0.020	0.000	-0.006	0.074	0.114	0.006	-0.001	-0.001	0.017	0.010
RMSE	0.022	0.008	0.023	0.082	0.119	0.019	0.018	0.023	0.033	0.028
$T = 50$										
Bias	-0.010	0.000	-0.004	0.119	0.169	0.005	0.000	0.000	0.034	0.028
RMSE	0.011	0.005	0.014	0.123	0.171	0.014	0.013	0.017	0.041	0.034
	$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.9$									
$T = 5$										
Bias	-0.554	0.032	-0.309	-0.026	0.022	-0.063	0.003	-0.036	-0.004	-0.001
RMSE	0.556	0.125	0.444	0.157	0.081	0.077	0.050	0.076	0.063	0.063
$T = 10$										
Bias	-0.281	0.028	-0.132	0.004	0.026	-0.025	0.002	-0.015	-0.003	-0.005
RMSE	0.282	0.078	0.192	0.077	0.053	0.039	0.030	0.043	0.042	0.042
$T = 25$										
Bias	-0.104	0.001	-0.042	0.012	0.032	-0.004	-0.001	-0.003	0.000	-0.002
RMSE	0.105	0.016	0.063	0.037	0.041	0.018	0.017	0.024	0.027	0.026
$T = 50$										
Bias	-0.048	0.000	-0.023	0.024	0.041	0.001	0.000	0.000	0.004	0.002
RMSE	0.048	0.007	0.034	0.034	0.043	0.013	0.012	0.018	0.020	0.018

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 5: Simulation results: heteroskedasticity, $N = 50$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 5$												
Bias	-0.075	0.001	-0.032	0.041	0.092	-0.004	-0.001	0.002	-0.018	0.014	-0.008	-0.024
RMSE	0.095	0.053	0.116	0.116	0.127	0.047	0.097	0.098	0.133	0.132	0.125	0.100
$T = 10$												
Bias	-0.034	0.000	-0.017	0.079	0.123	-0.008	0.018	0.004	0.001	0.040	-0.023	-0.019
RMSE	0.042	0.024	0.054	0.128	0.138	0.024	0.066	0.062	0.079	0.123	0.088	0.065
$T = 25$												
Bias	-0.013	0.000	-0.008	0.087	0.147	-0.005	0.009	-0.001	0.002	0.027	-0.035	-0.011
RMSE	0.019	0.014	0.024	0.114	0.152	0.015	0.038	0.037	0.046	0.096	0.065	0.039
$T = 50$												
Bias	-0.007	-0.001	-0.006	0.081	0.145	-0.003	0.006	0.000	0.006	0.017	-0.032	-0.005
RMSE	0.012	0.010	0.016	0.096	0.149	0.010	0.027	0.026	0.032	0.068	0.050	0.027
	$\alpha = 0.9$											
$T = 5$												
Bias	-0.405	-0.100	-0.373	-0.096	0.017	-0.058	-0.047	-0.012	-0.054	-0.012	0.006	-0.033
RMSE	0.437	0.249	0.536	0.272	0.066	0.252	0.105	0.096	0.132	0.112	0.102	0.098
$T = 10$												
Bias	-0.217	-0.028	-0.244	-0.017	0.024	-0.041	-0.018	-0.004	-0.034	0.003	0.007	-0.041
RMSE	0.230	0.107	0.307	0.113	0.043	0.103	0.063	0.060	0.085	0.079	0.070	0.072
Size	0.961	0.248	0.364	0.077	0.230	0.326	0.059	0.211	0.113	0.043	0.080	0.150
$T = 25$												
Bias	-0.084	0.000	-0.103	-0.002	0.025	-0.018	-0.001	-0.002	-0.011	0.007	0.009	-0.030
RMSE	0.088	0.033	0.136	0.076	0.032	0.030	0.036	0.035	0.050	0.067	0.044	0.047
$T = 50$												
Bias	-0.040	-0.001	-0.049	0.001	0.026	-0.010	0.004	0.001	0.004	0.011	0.014	-0.018
RMSE	0.042	0.017	0.064	0.059	0.030	0.017	0.026	0.025	0.034	0.052	0.034	0.032

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 6: Simulation results: heteroskedasticity, $N = 200$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 5$												
Bias	-0.071	0.003	-0.011	0.010	0.018	-0.002	-0.002	0.001	-0.003	0.005	0.000	-0.025
RMSE	0.084	0.037	0.056	0.040	0.044	0.028	0.048	0.047	0.063	0.055	0.054	0.053
$T = 10$												
Bias	-0.032	0.001	-0.006	0.012	0.019	-0.007	0.014	0.000	-0.003	0.000	-0.012	-0.022
RMSE	0.037	0.013	0.025	0.026	0.027	0.014	0.034	0.030	0.039	0.037	0.039	0.039
$T = 25$												
Bias	-0.013	0.000	-0.002	0.018	0.033	-0.005	0.010	0.000	0.001	0.004	-0.010	-0.011
RMSE	0.015	0.007	0.012	0.026	0.036	0.009	0.021	0.018	0.022	0.024	0.024	0.022
$T = 50$												
Bias	-0.006	0.000	-0.001	0.029	0.057	-0.003	0.006	0.000	0.001	0.005	-0.010	-0.006
RMSE	0.008	0.005	0.007	0.035	0.059	0.006	0.014	0.013	0.016	0.028	0.020	0.014
	$\alpha = 0.9$											
$T = 5$												
Bias	-0.407	-0.092	-0.198	-0.050	0.005	-0.077	-0.048	-0.012	-0.028	-0.012	-0.003	-0.037
RMSE	0.435	0.244	0.326	0.187	0.038	0.238	0.070	0.057	0.073	0.066	0.059	0.062
$T = 10$												
Bias	-0.215	-0.022	-0.097	-0.006	0.007	-0.040	-0.015	-0.001	-0.013	-0.002	-0.001	-0.037
RMSE	0.225	0.097	0.148	0.038	0.018	0.087	0.035	0.031	0.044	0.036	0.035	0.048
$T = 25$												
Bias	-0.083	0.002	-0.029	-0.001	0.009	-0.017	0.002	0.000	-0.001	0.001	0.002	-0.027
RMSE	0.085	0.026	0.046	0.035	0.014	0.023	0.019	0.018	0.024	0.021	0.020	0.033
$T = 50$												
Bias	-0.039	0.000	-0.013	0.003	0.011	-0.010	0.004	0.001	0.002	0.005	0.007	-0.018
RMSE	0.040	0.009	0.022	0.029	0.014	0.012	0.013	0.012	0.017	0.021	0.018	0.022

Note: Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 7: Simulation results: uniform cross-sectional dependence, $N = 50$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 10$												
Bias	-0.037	-0.004	-0.070	-0.015	0.120	-0.015	0.014	0.001	-0.018	-0.031	-0.057	-0.032
RMSE	0.065	0.055	0.135	0.182	0.143	0.054	0.068	0.067	0.106	0.180	0.122	0.078
Size	0.584	0.415	0.501	0.492	0.578	0.449	0.086	0.412	0.154	0.189	0.111	0.121
rob-Size	0.186	0.065	n.a.	n.a.	n.a.	n.a.	0.099	0.065	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.015	-0.002	-0.033	-0.001	0.135	-0.008	0.012	0.002	0.018	0.013	-0.035	-0.013
RMSE	0.036	0.033	0.062	0.137	0.143	0.033	0.045	0.043	0.057	0.127	0.071	0.046
Size	0.564	0.478	0.551	0.530	0.909	0.511	0.105	0.477	0.127	0.289	0.101	0.123
rob-Size	0.101	0.066	n.a.	n.a.	n.a.	n.a.	0.078	0.066	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.007	-0.001	-0.025	-0.009	0.136	-0.005	0.007	0.001	0.025	0.012	-0.040	-0.006
RMSE	0.024	0.023	0.042	0.123	0.141	0.023	0.032	0.032	0.044	0.106	0.060	0.033
Size	0.540	0.496	0.568	0.535	0.961	0.518	0.139	0.496	0.173	0.298	0.143	0.147
rob-Size	0.071	0.054	n.a.	n.a.	n.a.	n.a.	0.071	0.054	n.a.	n.a.	n.a.	n.a.
	$\alpha = 0.9$											
$T = 10$												
Bias	-0.266	-0.092	-0.492	-0.292	-0.085	-0.080	-0.020	-0.006	-0.056	-0.029	0.016	-0.089
RMSE	0.313	0.197	0.582	0.479	0.146	0.223	0.063	0.061	0.107	0.155	0.101	0.115
Size	0.949	0.527	0.924	0.807	0.503	0.865	0.082	0.517	0.271	0.149	0.071	0.432
rob-Size	0.508	0.078	n.a.	n.a.	n.a.	n.a.	0.153	0.085	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.111	-0.032	-0.303	-0.163	-0.038	-0.050	0.002	0.002	-0.006	0.008	0.024	-0.061
RMSE	0.137	0.094	0.352	0.332	0.072	0.100	0.036	0.036	0.063	0.111	0.065	0.090
Size	0.929	0.594	0.945	0.847	0.569	0.796	0.064	0.539	0.190	0.227	0.114	0.342
rob-Size	0.342	0.039	n.a.	n.a.	n.a.	n.a.	0.101	0.048	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.056	-0.017	-0.217	-0.117	-0.016	-0.033	0.005	0.002	0.024	0.023	0.017	-0.037
RMSE	0.075	0.057	0.246	0.276	0.044	0.058	0.025	0.024	0.049	0.100	0.045	0.053
Size	0.898	0.713	0.972	0.829	0.429	0.818	0.065	0.614	0.218	0.306	0.105	0.324
rob-Size	0.224	0.040	n.a.	n.a.	n.a.	n.a.	0.059	0.045	n.a.	n.a.	n.a.	n.a.

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size of the Wald statistics for the hypothesis $\beta = \beta_0$. 'rob-Size' refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 8: Simulation results: uniform cross-sectional dependence, $N = 200$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 10$												
Bias	-0.037	-0.003	-0.064	-0.053	0.024	-0.014	0.016	0.002	-0.015	-0.066	-0.089	-0.031
RMSE	0.061	0.049	0.122	0.199	0.066	0.048	0.040	0.036	0.078	0.132	0.112	0.052
Size	0.775	0.670	0.686	0.712	0.318	0.691	0.138	0.670	0.319	0.365	0.376	0.246
rob-Size	0.175	0.061	n.a.	n.a.	n.a.	n.a.	0.131	0.061	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.014	-0.001	-0.030	-0.038	0.044	-0.008	0.011	0.001	0.017	-0.010	-0.037	-0.014
RMSE	0.034	0.031	0.060	0.158	0.060	0.031	0.032	0.030	0.041	0.083	0.053	0.034
Size	0.733	0.689	0.759	0.796	0.631	0.705	0.282	0.686	0.270	0.304	0.198	0.280
rob-Size	0.103	0.043	n.a.	n.a.	n.a.	n.a.	0.092	0.043	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.006	0.000	-0.020	-0.025	0.050	-0.003	0.004	-0.001	0.022	0.017	-0.009	-0.009
RMSE	0.022	0.022	0.038	0.132	0.058	0.021	0.023	0.022	0.033	0.074	0.026	0.024
Size	0.709	0.688	0.761	0.799	0.827	0.688	0.271	0.688	0.368	0.523	0.169	0.305
rob-Size	0.060	0.055	n.a.	n.a.	n.a.	n.a.	0.061	0.055	n.a.	n.a.	n.a.	n.a.
	$\alpha = 0.9$											
$T = 10$												
Bias	-0.267	-0.092	-0.516	-0.365	-0.179	-0.085	-0.018	-0.005	-0.057	-0.045	0.005	-0.087
RMSE	0.311	0.191	0.605	0.558	0.244	0.216	0.036	0.032	0.087	0.111	0.060	0.104
Size	0.982	0.614	0.974	0.919	0.873	0.917	0.116	0.594	0.497	0.202	0.027	0.691
rob-Size	0.524	0.083	n.a.	n.a.	n.a.	n.a.	0.188	0.083	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.112	-0.034	-0.321	-0.190	-0.092	-0.052	0.002	0.001	-0.009	0.002	0.013	-0.063
RMSE	0.138	0.094	0.370	0.384	0.129	0.100	0.018	0.018	0.044	0.074	0.038	0.087
Size	0.959	0.693	0.973	0.931	0.904	0.903	0.057	0.613	0.328	0.223	0.078	0.664
rob-Size	0.356	0.043	n.a.	n.a.	n.a.	n.a.	0.090	0.051	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.052	-0.012	-0.206	-0.130	-0.044	-0.029	0.004	0.001	0.024	0.023	0.010	-0.037
RMSE	0.071	0.054	0.235	0.311	0.073	0.053	0.014	0.013	0.037	0.062	0.027	0.046
Size	0.946	0.809	0.991	0.942	0.899	0.899	0.081	0.715	0.434	0.486	0.217	0.684
rob-Size	0.210	0.022	n.a.	n.a.	n.a.	n.a.	0.073	0.028	n.a.	n.a.	n.a.	n.a.

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size of the Wald statistics for the hypothesis $\beta = \beta_0$. 'rob-Size' refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 9: Simulation results: interactive random effects, $N = 50$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
						$\alpha = 0.4$						
$T = 10$												
Bias	-0.035	-0.001	-0.052	0.017	0.123	-0.010	0.017	0.002	-0.009	0.005	-0.042	-0.023
RMSE	0.056	0.045	0.106	0.159	0.142	0.044	0.063	0.061	0.092	0.157	0.104	0.066
Size	0.484	0.323	0.365	0.336	0.581	0.345	0.070	0.322	0.093	0.121	0.074	0.086
rob-Size	0.240	0.080	n.a.	n.a.	n.a.	n.a.	0.114	0.080	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.014	-0.001	-0.024	0.024	0.138	-0.006	0.009	0.000	0.012	0.014	-0.037	-0.012
RMSE	0.029	0.026	0.048	0.125	0.145	0.026	0.042	0.042	0.053	0.124	0.069	0.044
Size	0.417	0.332	0.388	0.419	0.915	0.350	0.104	0.332	0.102	0.214	0.104	0.113
rob-Size	0.100	0.058	n.a.	n.a.	n.a.	n.a.	0.084	0.058	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.008	-0.002	-0.019	0.018	0.135	-0.004	0.006	0.001	0.017	0.020	-0.039	-0.005
RMSE	0.020	0.019	0.034	0.105	0.140	0.019	0.030	0.029	0.038	0.103	0.057	0.030
Size	0.414	0.367	0.446	0.381	0.960	0.375	0.106	0.367	0.116	0.223	0.142	0.113
rob-Size	0.082	0.060	n.a.	n.a.	n.a.	n.a.	0.068	0.060	n.a.	n.a.	n.a.	n.a.
$T = 10$												
Bias	-0.249	-0.063	-0.470	-0.205	-0.034	-0.070	-0.018	-0.004	-0.053	-0.017	0.007	-0.058
RMSE	0.277	0.150	0.554	0.362	0.093	0.165	0.062	0.059	0.099	0.136	0.085	0.086
Size	0.943	0.431	0.882	0.625	0.278	0.700	0.080	0.401	0.206	0.103	0.045	0.241
rob-Size	0.666	0.056	n.a.	n.a.	n.a.	n.a.	0.145	0.059	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.098	-0.016	-0.268	-0.137	-0.006	-0.036	0.003	0.002	-0.005	0.008	0.015	-0.037
RMSE	0.112	0.066	0.309	0.285	0.043	0.066	0.037	0.036	0.056	0.103	0.057	0.058
Size	0.922	0.455	0.924	0.726	0.256	0.664	0.074	0.411	0.121	0.179	0.083	0.204
rob-Size	0.478	0.022	n.a.	n.a.	n.a.	n.a.	0.091	0.029	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.045	-0.005	-0.162	-0.071	0.005	-0.018	0.003	0.000	0.018	0.018	0.010	-0.027
RMSE	0.058	0.042	0.189	0.242	0.033	0.040	0.025	0.025	0.042	0.094	0.040	0.040
Size	0.832	0.591	0.908	0.718	0.214	0.699	0.064	0.513	0.138	0.254	0.075	0.246
rob-Size	0.279	0.025	n.a.	n.a.	n.a.	n.a.	0.075	0.034	n.a.	n.a.	n.a.	n.a.

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size of the Wald statistics for the hypothesis $\beta = \beta_0$. 'rob-Size' refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 10: Simulation results: interactive random effects, $N = 200$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 10$												
Bias	-0.035	-0.002	-0.044	-0.018	0.027	-0.010	0.014	0.000	-0.009	-0.041	-0.052	-0.025
RMSE	0.053	0.040	0.095	0.133	0.053	0.039	0.036	0.033	0.066	0.089	0.073	0.044
Size	0.754	0.562	0.561	0.533	0.250	0.583	0.105	0.562	0.231	0.201	0.175	0.166
rob-Size	0.278	0.064	n.a.	n.a.	n.a.	n.a.	0.135	0.064	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.014	-0.001	-0.021	-0.010	0.041	-0.006	0.010	0.001	0.010	-0.008	-0.026	-0.011
RMSE	0.027	0.024	0.043	0.116	0.050	0.024	0.027	0.025	0.033	0.064	0.040	0.028
Size	0.660	0.577	0.620	0.617	0.646	0.603	0.200	0.577	0.186	0.211	0.139	0.200
rob-Size	0.124	0.058	n.a.	n.a.	n.a.	n.a.	0.097	0.058	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.007	0.000	-0.015	-0.014	0.052	-0.003	0.006	0.000	0.015	0.018	-0.010	-0.006
RMSE	0.018	0.017	0.029	0.120	0.058	0.017	0.020	0.019	0.026	0.079	0.024	0.020
Size	0.656	0.605	0.641	0.673	0.890	0.610	0.203	0.605	0.239	0.461	0.128	0.211
rob-Size	0.077	0.051	n.a.	n.a.	n.a.	n.a.	0.073	0.051	n.a.	n.a.	n.a.	n.a.
$T = 10$												
Bias	-0.246	-0.059	-0.463	-0.266	-0.087	-0.071	-0.018	-0.004	-0.052	-0.029	-0.004	-0.058
RMSE	0.271	0.143	0.546	0.429	0.139	0.155	0.035	0.031	0.078	0.078	0.045	0.071
Size	0.976	0.565	0.951	0.866	0.623	0.843	0.105	0.540	0.445	0.115	0.020	0.516
rob-Size	0.676	0.050	n.a.	n.a.	n.a.	n.a.	0.180	0.050	n.a.	n.a.	n.a.	n.a.
$T = 25$												
Bias	-0.097	-0.014	-0.256	-0.155	-0.043	-0.035	0.002	0.001	-0.005	0.002	0.005	-0.039
RMSE	0.112	0.067	0.300	0.328	0.077	0.066	0.018	0.018	0.036	0.058	0.029	0.049
Size	0.957	0.634	0.961	0.879	0.728	0.842	0.050	0.566	0.216	0.176	0.042	0.536
rob-Size	0.456	0.027	n.a.	n.a.	n.a.	n.a.	0.081	0.033	n.a.	n.a.	n.a.	n.a.
$T = 50$												
Bias	-0.046	-0.006	-0.157	-0.094	-0.023	-0.019	0.004	0.000	0.018	0.019	0.008	-0.026
RMSE	0.058	0.041	0.183	0.265	0.052	0.039	0.013	0.013	0.029	0.056	0.024	0.033
Size	0.928	0.770	0.964	0.909	0.785	0.843	0.067	0.634	0.296	0.389	0.146	0.515
rob-Size	0.301	0.025	n.a.	n.a.	n.a.	n.a.	0.074	0.033	n.a.	n.a.	n.a.	n.a.

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size of the Wald statistics for the hypothesis $\beta = \beta_0$. 'rob-Size' refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 11: Simulation results: nonstationary initialization, $N = 50$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
						$\alpha = 0.4$						
$T = 5$												
Bias	-0.041	0.000	-0.056	0.031	0.386	0.000	-0.001	-0.003	-0.041	0.009	0.071	0.009
RMSE	0.048	0.027	0.175	0.101	0.388	0.027	0.098	0.099	0.172	0.156	0.187	0.099
$T = 10$												
Bias	-0.023	0.001	-0.037	0.087	0.364	0.003	0.010	0.000	-0.018	0.043	0.003	0.010
RMSE	0.029	0.018	0.083	0.157	0.366	0.019	0.062	0.062	0.090	0.138	0.118	0.062
$T = 25$												
Bias	-0.011	0.000	-0.009	0.084	0.346	0.001	0.007	-0.001	0.002	0.036	-0.026	0.003
RMSE	0.016	0.012	0.026	0.107	0.347	0.012	0.038	0.037	0.045	0.090	0.070	0.037
$T = 50$												
Bias	-0.006	0.000	-0.006	0.073	0.294	0.001	0.005	0.000	0.005	0.017	-0.040	0.002
RMSE	0.010	0.008	0.015	0.083	0.297	0.008	0.025	0.025	0.031	0.061	0.061	0.025
						$\alpha = 0.9$						
$T = 5$												
Bias	-0.328	0.011	-0.448	-0.013	0.114	0.032	-0.016	-0.002	-0.064	-0.012	-0.001	0.028
RMSE	0.334	0.119	0.621	0.149	0.124	0.145	0.098	0.102	0.132	0.129	0.136	0.100
$T = 10$												
Bias	-0.152	0.003	-0.308	-0.004	0.089	0.029	-0.001	0.000	-0.047	0.003	0.008	0.032
RMSE	0.156	0.052	0.379	0.094	0.092	0.073	0.062	0.061	0.091	0.089	0.088	0.066
$T = 25$												
Bias	-0.058	0.000	-0.265	0.006	0.055	0.011	0.002	-0.001	-0.036	0.009	0.011	0.020
RMSE	0.060	0.017	0.299	0.059	0.063	0.025	0.036	0.036	0.059	0.057	0.052	0.040
$T = 50$												
Bias	-0.030	0.000	-0.094	0.004	0.041	0.005	0.003	0.000	-0.003	0.007	0.006	0.013
RMSE	0.031	0.010	0.108	0.043	0.050	0.012	0.024	0.024	0.034	0.040	0.036	0.027

Note: The estimators in the comparison are the within-groups estimator (WG), our bias-corrected estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).

Table 12: Simulation results: nonstationary initialization, $N = 200$

	α						β					
	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML	WG	BC	AB-GMM	AS-GMM	BB-GMM	QML
	$\alpha = 0.4$											
$T = 5$												
Bias	-0.041	0.000	-0.014	0.003	0.379	0.000	0.000	-0.002	-0.011	-0.004	0.101	0.010
RMSE	0.043	0.014	0.088	0.028	0.380	0.014	0.048	0.048	0.086	0.064	0.135	0.049
$T = 10$												
Bias	-0.023	0.000	-0.010	0.006	0.348	0.002	0.013	0.003	-0.003	-0.001	0.019	0.012
RMSE	0.025	0.009	0.041	0.034	0.348	0.010	0.033	0.030	0.045	0.047	0.065	0.032
$T = 25$												
Bias	-0.011	0.000	-0.003	0.012	0.346	0.001	0.009	0.001	0.001	0.000	-0.010	0.004
RMSE	0.012	0.006	0.012	0.018	0.346	0.006	0.020	0.018	0.024	0.029	0.037	0.018
$T = 50$												
Bias	-0.006	0.000	-0.001	0.024	0.360	0.001	0.005	0.000	0.001	0.007	-0.010	0.002
RMSE	0.007	0.004	0.007	0.027	0.360	0.004	0.014	0.013	0.016	0.025	0.028	0.013
	$\alpha = 0.9$											
$T = 5$												
Bias	-0.321	0.011	-0.210	0.020	0.115	0.022	-0.015	-0.001	-0.029	-0.001	0.009	0.028
RMSE	0.323	0.069	0.369	0.099	0.117	0.077	0.049	0.047	0.074	0.062	0.063	0.053
$T = 10$												
Bias	-0.149	0.000	-0.130	0.014	0.091	0.020	0.001	0.003	-0.018	0.001	0.008	0.035
RMSE	0.150	0.024	0.187	0.066	0.092	0.036	0.030	0.030	0.048	0.044	0.044	0.045
$T = 25$												
Bias	-0.056	0.000	-0.139	0.013	0.070	0.010	0.004	0.001	-0.020	0.002	0.009	0.022
RMSE	0.057	0.008	0.164	0.041	0.070	0.014	0.018	0.017	0.034	0.029	0.030	0.028
$T = 50$												
Bias	-0.030	0.000	-0.027	0.015	0.065	0.005	0.003	0.000	-0.002	0.003	0.010	0.012
RMSE	0.030	0.005	0.036	0.026	0.065	0.007	0.013	0.012	0.017	0.019	0.021	0.018

Note: The estimators in the comparison are the within-groups estimator (WG), our bias-corrected estimator (BC), the one-step Arellano and Bond (1991) GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995) GMM estimator (AS-GMM), the two-step Blundell and Bond (1998) GMM estimator (BB-GMM), and the Hsiao et al. (2002) QML estimator. Reported are the average bias of the estimates and the root mean square error (RMSE).