

Bias-corrected method of moments estimators for dynamic panel data models

Jörg Breitung^{a,*}, Sebastian Kripfganz^b, Kazuhiko Hayakawa^c

^a *University of Cologne, Institute of Econometrics, Albertus-Magnus-Platz, 50923 Cologne, Germany*

^b *University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, UK*

^c *Hiroshima University, Department of Economics, 1-2-3 Kagamiyama, Hiroshima, Japan*

Abstract

A computationally simple bias correction for linear dynamic panel data models is proposed and its asymptotic properties are studied when the number of time periods is fixed or tends to infinity with the number of panel units. The approach can accommodate both fixed-effects and random-effects assumptions, heteroskedastic errors, as well as higher-order autoregressive models. Panel-corrected standard errors are proposed that allow for robust inference in dynamic models with cross-sectionally correlated errors. Monte Carlo experiments suggest that under the assumption of strictly exogenous regressors the bias-corrected method of moment estimator outperforms popular GMM estimators in terms of efficiency and correctly sized tests.

Keywords: Bias correction, Moment conditions, Autoregressive model, Panel data, Fixed effects, Random Effects

JEL classification: C13, C23, C63

1. Introduction

Dynamic panel data models are now used in a wide area of empirical applications. Since the work of Anderson and Hsiao (1981), instrumental variables and generalized method of moments (GMM) estimators have been extensively applied in the estimation of linear dynamic panel data models. However, it is known that the GMM estimator by Holtz-Eakin et al. (1988) and Arellano and

*Corresponding author

Email address: breitung@statistik.uni-koeln.de (Jörg Breitung)

7 Bond (1991) suffers from the weak-instruments problem when the persistency
8 of the data is strong, as demonstrated by Blundell and Bond (1998). They also
9 showed that the GMM estimator for models in levels with first-differenced in-
10 struments mitigates that problem and they proposed the so-called system GMM
11 estimator that combines moment conditions for the models in first differences
12 and in levels. Nowadays, the system GMM estimator is most frequently used in
13 practice, albeit Bun and Windmeijer (2010) showed that it still suffers from the
14 weak-instruments problem when the variance of the individual-specific effects is
15 larger than that of the idiosyncratic errors.

16 As alternatives to the GMM approach, maximum likelihood (ML) estima-
17 tors and bias-corrected within-groups (WG) estimators were proposed. Hsiao
18 et al. (2002) suggested a transformed ML estimator that adapts the ML ap-
19 proach to the differenced variables. Hayakawa and Pesaran (2015) extended
20 this transformed ML estimator to allow for cross-sectional heteroskedasticity
21 and proposed robust standard errors. With regard to bias-corrected WG es-
22 timators, Kiviet (1995) and Judson and Owen (1999) demonstrate that they
23 are attractive alternatives to GMM estimators. Although the bias-corrected
24 WG estimator of Kiviet (1995) is based on a higher-order expansion of the bias
25 term, the analytical results are based on the unknown parameters that have to
26 be estimated by some consistent initial estimator. Accordingly, the asymptotic
27 distribution of this estimator is unknown. Bun and Carree (2005) proposed
28 an alternative bias-corrected WG estimator which iteratively solves a nonlinear
29 equation with regard to unknown parameters. Dhaene and Jochmans (2016)
30 obtain an adjusted profile likelihood function by integrating a bias-corrected
31 profile score. Kruiniger (2018) proposes a generalized version of the modified
32 ML estimator that addresses identification problems when the autoregressive
33 parameter equals unity or when $T = 2$.

34 In this paper, we demonstrate that a bias-corrected estimator can be ob-
35 tained as a method of moments estimator. The adjusted profile score is trans-
36 formed into nonlinear moment conditions that can be easily solved with standard
37 numerical methods. Asymptotic results are readily available. The underlying
38 estimating equations are equivalent to those of the Dhaene and Jochmans (2016)
39 estimator when we adopt a fixed-effects assumption for the exogenous regres-
40 sors. For the first-order autoregressive model, they are also equivalent to those
41 of Bun and Carree (2005) and Kruiniger (2018). Furthermore, we re-emphasize
42 an earlier finding by Bun et al. (2017) that the ML estimators of Hsiao et al.
43 (2002) and Bai (2013) are based on a modified log-likelihood function that leads

44 to an asymptotically equivalent bias correction of the first-order condition. Yet,
 45 these ML estimators rely on additional assumptions about the initial observa-
 46 tions that are not required for our approach.

47 Within our method of moments framework, we can easily differentiate be-
 48 tween regressors that are correlated with the individual-specific effects and those
 49 that are uncorrelated with them. All it requires is a slight modification of the
 50 respective moment conditions. This allows for the estimation of dynamic fixed-
 51 effects and dynamic random-effects models, as well as hybrid versions. Under
 52 appropriate orthogonality assumptions, time-invariant regressors can be incor-
 53 porated as well.

54 Moreover, the model allows for individual-specific heteroskedasticity in the
 55 large- N , fixed- T framework. When both N and T are large, the estimator is
 56 also robust to time series heteroskedasticity. Furthermore, we propose cluster-
 57 robust/panel-corrected standard errors that account for cross-sectional depen-
 58 dence, and we extend our bias-corrected method of moments approach to higher-
 59 order autoregressive models. Monte Carlo experiments suggest that these esti-
 60 mators perform well in terms of efficiency and correctly sized tests, relative to
 61 uncorrected WG and GMM approaches.

62 2. Bias-corrected method of moments estimation

63 To motivate the bias correction approach, we initially consider the pure
 64 first-order autoregressive model

$$y_{it} = \alpha y_{i,t-1} + \mu_i + u_{it}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N,$$

65 where $u_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$. The log-likelihood function conditional on the initial
 66 observations is given by

$$\ell(\alpha, \sigma^2, \boldsymbol{\mu}) = -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T u_{it}(\alpha, \mu_i)^2, \quad (1)$$

67 Profiling out the nuisance parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)'$ yields

$$\ell(\alpha, \sigma^2) = -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T [e_{it}(\alpha) - \bar{e}_i(\alpha)]^2,$$

68 where $e_{it}(\alpha) = y_{it} - \alpha y_{i,t-1}$ and $\bar{e}_i(\alpha) = T^{-1} \sum_{t=1}^T e_{it}(\alpha)$. For estimating
 69 the parameter α in such an AR(1) model, Dhaene and Jochmans (2016) further

70 profile out the variance parameter σ^2 to obtain the profile log-likelihood function

$$\ell(\alpha) = -\frac{NT}{2} \left[\ln(2\pi) + 1 + \ln \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [e_{it}(\alpha) - \bar{e}_i(\alpha)] e_{it}(\alpha) \right) \right].$$

71 Maximization with respect to α yields the profile score (divided by N and T)

$$s_{NT}(\alpha) = \frac{1}{NT} \frac{\partial \ell(\alpha)}{\partial \alpha} = \frac{\sum_{i=1}^N \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\alpha)}{\sum_{i=1}^N \sum_{t=1}^T [e_{it}(\alpha) - \bar{e}_i(\alpha)] e_{it}(\alpha)},$$

72 where $\bar{y}_{-1,i} = T^{-1} \sum_{t=1}^T y_{i,t-1}$. The bias-adjusted profile likelihood estima-
73 tor $\hat{\alpha}_{al}$ results from correcting the inconsistency in the profile score by solving
74 $s_{NT}(\hat{\alpha}_{al}) - \text{plim}_{N \rightarrow \infty} s_{NT}(\hat{\alpha}_{al}) = 0$.

75 Instead, we simplify the bias correction by initially assuming that the vari-
76 ance σ^2 is known such that the profile score becomes

$$\tilde{s}_{NT}(\alpha) = \frac{1}{NT} \frac{\partial \ell(\alpha, \sigma^2)}{\partial \alpha} = \frac{1}{\sigma^2 NT} \sum_{i=1}^N \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\alpha).$$

77 This allows us to restrict the focus on the numerator in the profile score. Let
78 $b_T(\alpha)$ be a polynomial in α as presented further below such that

$$b_T(\alpha_0) = \mathbb{E} \left[\frac{1}{\sigma^2 T} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\alpha_0) \right],$$

79 when evaluated at the true parameter value α_0 . A bias-corrected method of
80 moments estimator $\hat{\alpha}_{bc}$ then solves the moment equation

$$m_{NT}(\alpha) = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\hat{\alpha}_{bc}) - b_T(\hat{\alpha}_{bc}) \hat{\sigma}_T^2(\hat{\alpha}_{bc}) \right) = 0,$$

81 where the unknown variance parameter σ^2 was replaced with an estimator
82 $\hat{\sigma}_T^2(\alpha)$.

83 Let us now look in detail at the first-order dynamic model with strictly
84 exogenous regressors given by

$$y_{it} = \alpha_0 y_{i,t-1} + \beta_0' \mathbf{x}_{it} + \mu_i + u_{it}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N,$$

85 where α_0 and the $K \times 1$ vector β_0 denote the true values of the parameters
 86 of interest. For the $K \times 1$ vector of regressors \mathbf{x}_{it} and the error term u_{it} , the
 87 following set of assumptions is imposed that is standard in the literature on
 88 dynamic panel data models:

89 **Assumption 1:** (i) The errors u_{it} are independent across i and t with $\mathbb{E}[u_{it}] =$
 90 0 and $\mathbb{E}[u_{it}^2] = \sigma_i^2 < C$ for some constant $C < \infty$. (ii) The regressors are
 91 strictly exogenous with $\mathbb{E}[\mathbf{x}_{it}u_{is}] = \mathbf{0}$ and $\mathbb{E}[|u_{it}u_{is}\mathbf{x}_{it}\mathbf{x}'_{is}|] < \infty$ for all $t, s \in$
 92 $\{1, 2, \dots, T\}$ and $i \in \{1, 2, \dots, N\}$. (iii) $\mathbb{E}|u_{it}|^{4+\delta} < \infty$ for all i and t and
 93 some $\delta > 0$. (iv) For the initial values, we assume $\mathbb{E}[y_{i0}^2] < \infty$ for all i and
 94 $\mathbb{E}[y_{i0}u_{it}] = 0$ for all i and $t \in \{1, 2, \dots, T\}$.

95 Note that we do not impose any stationarity restriction on the initial values.
 96 The process is allowed to start at any fixed or random level in the finite past.
 97 If T is fixed, Assumption 1 does not rule out nonstationary regressors or un-
 98 stable processes with $|\alpha_0| \geq 1$, although identification of the parameters might
 99 require further refinements as discussed by Kruiniger (2018). If T tends to in-
 100 finity, additional assumptions would be required for the limiting distribution
 101 of the estimator. It should also be noted that we allow for individual-specific
 102 heteroskedasticity. Robustness to time series heteroskedasticity can also be
 103 achieved when $T \rightarrow \infty$.

104 Treating y_{i0} as a fixed constant, the first-order conditions of the (quasi-)ML
 105 or least-squares WG estimator $\hat{\boldsymbol{\theta}}_{ml} = (\hat{\boldsymbol{\alpha}}_{ml}, \hat{\boldsymbol{\beta}}'_{ml})'$ are

$$\mathbf{g}_{NT}(\hat{\boldsymbol{\theta}}_{ml}) = \begin{pmatrix} g_{\alpha, NT}(\hat{\boldsymbol{\theta}}_{ml}) \\ \mathbf{g}_{\beta, NT}(\hat{\boldsymbol{\theta}}_{ml}) \end{pmatrix} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \begin{pmatrix} y_{i,t-1} - \bar{y}_{-1,i} \\ \mathbf{x}_{it} - \bar{\mathbf{x}}_i \end{pmatrix} e_{it}(\hat{\boldsymbol{\theta}}_{ml}) = \mathbf{0},$$

106 where $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$, $e_{it}(\boldsymbol{\theta}) = y_{i,t-1} - \alpha y_{i,t-1} - \boldsymbol{\beta}' \mathbf{x}_{it}$, and $\bar{e}_i(\boldsymbol{\theta}) =$
 107 $T^{-1} \sum_{t=1}^T e_{it}(\boldsymbol{\theta})$. Following Nickell (1981) and Moon et al. (2015), we obtain

$$\begin{aligned} \mathbb{E}[g_{\alpha, NT}(\boldsymbol{\theta}_0)] &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\boldsymbol{\theta}_0) \right] \\ &= \frac{1}{N} \sum_{i=1}^N \left(-\frac{\sigma_i^2}{T^2} [(T-1) + (T-2)\alpha_0 + (T-3)\alpha_0^2 + \dots + 2\alpha_0^{T-3} + \alpha_0^{T-2}] \right) \\ &= b_T(\alpha_0) \frac{1}{N} \sum_{i=1}^N \sigma_i^2, \end{aligned}$$

108 where $b_T(\alpha) = -T^{-2} \sum_{t=0}^{T-2} \sum_{s=0}^t \alpha^s$ is a negative-valued monotonously de-
 109 creasing function that simplifies to

$$b_T(\alpha) = \begin{cases} -\frac{1}{(1-\alpha)T} \left(1 - \frac{1-\alpha^T}{T(1-\alpha)}\right) = -\frac{1}{(1-\alpha)T} + O(T^{-2}) & \text{for } |\alpha| < 1, \\ -\frac{1}{2} + \frac{1}{2T} & \text{for } \alpha = 1. \end{cases}$$

110 Since $\text{plim}_{N \rightarrow \infty} g_{\alpha, NT}(\boldsymbol{\theta}_0) = b_T(\alpha_0)\sigma^2$, where $\sigma^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$, the
 111 WG estimator is inconsistent for fixed T . Using

$$\hat{\sigma}_{T_i}^2(\boldsymbol{\theta}) = \frac{1}{T-1} \sum_{t=1}^T [e_{it}(\boldsymbol{\theta}) - \bar{e}_i(\boldsymbol{\theta})] e_{it}(\boldsymbol{\theta}) \quad (2)$$

112 such that $\mathbb{E}[\hat{\sigma}_{T_i}^2(\boldsymbol{\theta}_0)] = \sigma_i^2$, we obtain the moment conditions

$$\mathbb{E}[\mathbf{m}_{T_i}(\boldsymbol{\theta}_0)] = \mathbb{E} \left[\begin{pmatrix} m_{\alpha, T_i}(\boldsymbol{\theta}_0) \\ \mathbf{m}_{\beta, T_i}(\boldsymbol{\theta}_0) \end{pmatrix} \right] = \mathbf{0}, \quad (3)$$

113 where

$$m_{\alpha, T_i}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\boldsymbol{\theta}) - b_T(\alpha) \hat{\sigma}_{T_i}^2(\boldsymbol{\theta}),$$

$$\mathbf{m}_{\beta, T_i}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) e_{it}(\boldsymbol{\theta}).$$

114 The bias-corrected method of moments estimator $\hat{\boldsymbol{\theta}}_{bc}$ is obtained by solving the
 115 sample moment conditions $\mathbf{m}_{NT}(\hat{\boldsymbol{\theta}}_{bc}) = N^{-1} \sum_{i=1}^N \mathbf{m}_{T_i}(\hat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$.

116 Notice that the adjusted profile likelihood estimator of Dhaene and Jochmans
 117 (2016) is based on the estimating equations

$$\left(\frac{T}{T-1} \right) \frac{\mathbf{m}_{NT}(\hat{\boldsymbol{\theta}}_{al})}{\hat{\sigma}_{NT}^2(\hat{\boldsymbol{\theta}}_{al})} = \mathbf{0},$$

118 with $\hat{\sigma}_{NT}^2(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^N \hat{\sigma}_{T_i}^2(\boldsymbol{\theta})$. Consequently, the estimators $\hat{\boldsymbol{\theta}}_{bc}$ and $\hat{\boldsymbol{\theta}}_{al}$ are
 119 equivalent, but $\hat{\boldsymbol{\theta}}_{bc}$ is simpler to compute. We can furthermore demonstrate
 120 that also the iterative estimator proposed by Bun and Carree (2005) is based
 121 on equivalent estimating equations. Define

$$\begin{aligned}
s_{11} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) y_{i,t-1}, & s_{10} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) y_{it}, \\
\mathbf{s}_{x1} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) y_{i,t-1}, & \mathbf{s}_{x0} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) y_{it}, \\
\mathbf{S}_{xx} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \mathbf{x}'_{it}.
\end{aligned}$$

122 From $\mathbf{m}_{\beta,NT}(\hat{\boldsymbol{\theta}}_{bc}) = N^{-1} \sum_{i=1}^N \mathbf{m}_{\beta,Ti}(\hat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$, we obtain the closed-form so-
123 lution

$$\hat{\boldsymbol{\beta}}_{bc} = \mathbf{S}_{xx}^{-1} (\mathbf{s}_{x0} - \hat{\alpha}_{bc} \mathbf{s}_{x1}) = \hat{\boldsymbol{\beta}}_{ml} - (\hat{\alpha}_{bc} - \hat{\alpha}_{ml}) \mathbf{S}_{xx}^{-1} \mathbf{s}_{x1}, \quad (4)$$

124 with the uncorrected WG estimators

$$\hat{\alpha}_{ml} = \frac{s_{10} - \mathbf{s}'_{x0} \mathbf{S}_{xx}^{-1} \mathbf{s}_{x1}}{s_{11}^2 - \mathbf{s}'_{x1} \mathbf{S}_{xx}^{-1} \mathbf{s}_{x1}}, \quad \hat{\boldsymbol{\beta}}_{ml} = \mathbf{S}_{xx}^{-1} (\mathbf{s}_{x0} - \hat{\alpha}_{ml} \mathbf{s}_{x1}).$$

125 Using (4), after a few algebraic manipulations of $m_{\alpha,NT}(\hat{\boldsymbol{\theta}}_{bc}) = N^{-1} \sum_{i=1}^N m_{\alpha,Ti}(\hat{\boldsymbol{\theta}}_{bc}) =$
126 0 we obtain

$$\hat{\alpha}_{bc} = \hat{\alpha}_{ml} - \frac{b_T(\hat{\alpha}_{bc}) \hat{\sigma}_{NT}^2(\hat{\boldsymbol{\theta}}_{bc})}{s_{11} - \mathbf{s}'_{x1} \mathbf{S}_{xx}^{-1} \mathbf{s}_{x1}}. \quad (5)$$

127 Equations (4) and (5) are identical to those in Bun and Carree (2005, equation
128 20). Finally, we note that our estimator is equivalent to a particular version
129 of the modified ML estimator proposed by Kruiniger (2018). Strictly speaking,
130 the equivalence holds if there exists a local maximum of the objective function
131 and $\alpha \neq 1$. In order to allow for the possibility that the objective function
132 fails to possess a maximum in the relevant parameter space, Kruiniger (2018)
133 introduces a weight matrix and suggests to minimize the weighted norm of the
134 gradient vector.

135 Our formulation of the profile ML estimator as a method of moments esti-
136 mator has the advantage that standard numerical optimization procedures are
137 readily applicable and the derivation of the asymptotic properties is straight-
138 forward. Moreover, as we will outline further below, extending the estimator
139 to higher-order autoregressive models or to a dynamic random-effects model is

140 easily done within our framework by adjusting the moment conditions accord-
 141 ingly.

142 Due to the nonlinearity of the moment function $m_{\alpha,NT}(\boldsymbol{\theta})$, the solution to
 143 $\mathbf{m}_{NT}(\hat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$ needs to be obtained numerically. This can be done with the
 144 recursive Gauss-Newton algorithm, solving $\hat{\boldsymbol{\theta}}_{bc} = \arg \min_{\boldsymbol{\theta}} \mathbf{m}_{NT}(\boldsymbol{\theta})' \mathbf{m}_{NT}(\boldsymbol{\theta})$.
 145 Starting with an initial guess $\boldsymbol{\theta}^{(0)}$, $\hat{\boldsymbol{\theta}}_{bc}$ results upon convergence using the generic
 146 iteration step

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} - \left(\nabla \mathbf{m}_{NT}(\boldsymbol{\theta}^{(s)}) \right)^{-1} \mathbf{m}_{NT}(\boldsymbol{\theta}^{(s)}),$$

147 with the gradient

$$\nabla \mathbf{m}_{NT}(\boldsymbol{\theta}^{(s)}) = \frac{1}{N} \sum_{i=1}^N \nabla \mathbf{m}_{T_i}(\boldsymbol{\theta}^{(s)}) = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial \mathbf{m}_{T_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(s)}}.$$

148 The detailed entries of $\nabla \mathbf{m}_{T_i}(\boldsymbol{\theta})$ are provided in Appendix A.1. The numerical
 149 optimization can be simplified by profiling out $\hat{\boldsymbol{\beta}}_{bc} = \hat{\boldsymbol{\beta}}(\hat{\alpha}_{bc})$ using the closed-
 150 form solution (4), thus solving $\hat{\alpha}_{bc} = \arg \min_{\alpha} \tilde{m}_{NT}(\alpha)^2$, where $\tilde{m}_{NT}(\alpha) =$
 151 $m_{\alpha,NT} \left((\alpha, \hat{\boldsymbol{\beta}}(\alpha))' \right)$.

152 Dhaene and Jochmans (2016) point out that integrating the adjusted profile
 153 score leads to an adjusted likelihood function that is asymptotically unbounded
 154 from above. Moreover, while $\text{plim}_{N \rightarrow \infty} \mathbf{m}_{NT}(\boldsymbol{\theta}_0) = \mathbf{0}$, the adjusted profile
 155 score has multiple zeros. However, they provide evidence that the adjusted like-
 156 lihood function asymptotically has at most one local maximum for the most
 157 relevant situation of $\alpha_0 \geq -1$. If $T = 2$ (plus a strong condition on the ini-
 158 tial observations) or if $\alpha_0 = 1$, it may have an inflection point at $\boldsymbol{\theta}_0$. For
 159 practical purposes, a verification is needed that the numerical algorithm indeed
 160 converged to a local maximum, which requires $\nabla \tilde{m}_{NT}(\hat{\alpha}_{bc}) < 0$. If this condi-
 161 tion is violated, we suggest to re-initialize the search algorithm with a different
 162 initial guess $\alpha^{(0)}$. If necessary, the process should be repeated until a solution
 163 is found that satisfies the negativity condition for the gradient. In our simu-
 164 lations, this approach proves successful. In finite samples, we cannot exclude
 165 the possibility that there does not exist a unique local maximizer in the interior
 166 of $\Theta = \{\boldsymbol{\theta} : \alpha \in [-1, 1], \boldsymbol{\beta} \in \mathbb{R}^K\}$. In order to protect against a situation of
 167 multiple solutions, a grid search may be carried out over a reasonable range
 168 of $\alpha \in [-1, 1]$. To single out a point estimate, Dhaene and Jochmans (2016)
 169 suggest to choose the solution closest to $\hat{\alpha}_{ml}$. Besides the potential for multi-
 170 ple solutions, Kruiniger (2018) points out the possibility that no local optimum

171 exists in samples with finite T . However, we did not encounter such a case in
 172 our simulation exercises. With $T \rightarrow \infty$, these concerns vanish because $b_T(\alpha)$
 173 is $O(T^{-1})$ for $|\alpha| < 1$ and the bias-corrected estimator converges to the WG
 174 estimator.

175 Assuming that a unique maximizer has been established, the limiting distri-
 176 bution of the bias-corrected method of moments estimator $\widehat{\boldsymbol{\theta}}_{bc}$ can be obtained
 177 in a straightforward way. Here, we restrict our attention to the stationary re-
 178 gion. For $\alpha_0 = 1$, the asymptotic distribution is non-Gaussian, as dicussed by
 179 Kruiniger (2018).

180 **Theorem 1:** (i) Under Assumption 1 and with $|\alpha_0| < 1$, the limiting distribu-
 181 tion of $\widehat{\boldsymbol{\theta}}_{bc}$ for fixed T and as $N \rightarrow \infty$ is given by

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, [\boldsymbol{\Sigma}_T + \sigma^2 \mathbf{B}_T(\alpha_0)]^{-1} \mathbf{S}_T(\boldsymbol{\theta}_0) [\boldsymbol{\Sigma}_T + \sigma^2 \mathbf{B}_T(\alpha_0)]^{-1}),$$

182 with $\mathbf{S}_T(\boldsymbol{\theta}) = \text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{m}_{T_i}(\boldsymbol{\theta}) \mathbf{m}_{T_i}(\boldsymbol{\theta})'$,

$$\boldsymbol{\Sigma}_T = \text{plim}_{N \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \begin{pmatrix} y_{i,t-1} - \bar{y}_{-1,i} \\ \mathbf{x}_{it} - \bar{\mathbf{x}}_i \end{pmatrix} \begin{pmatrix} y_{i,t-1} & \mathbf{x}'_{it} \end{pmatrix},$$

$$\mathbf{B}_T(\alpha) = \begin{pmatrix} \nabla_{\alpha} b_T(\alpha) - \frac{2T}{T-1} b_T(\alpha)^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

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

184 (ii) Under Assumption 1 and with $|\alpha_0| < 1$, the limiting distribution of $\widehat{\boldsymbol{\theta}}_{bc}$
 185 as $T \rightarrow \infty$ and $N/T \rightarrow \kappa$ for $0 \leq \kappa < \infty$, is given by

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

186 with $\boldsymbol{\Sigma} = \lim_{T \rightarrow \infty} \boldsymbol{\Sigma}_T$.

187 A consistent fixed- T estimate of the covariance matrix of $\widehat{\boldsymbol{\theta}}_{bc}$ can be obtained
 188 with the finite-sample analog

$$\mathbf{V}_{NT}(\widehat{\boldsymbol{\theta}}_{bc}) = \frac{1}{N} \left(\nabla \mathbf{m}_{NT}(\widehat{\boldsymbol{\theta}}_{bc}) \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{m}_{T_i}(\widehat{\boldsymbol{\theta}}_{bc}) \mathbf{m}_{T_i}(\widehat{\boldsymbol{\theta}}_{bc})' \right) \left(\nabla \mathbf{m}_{NT}(\widehat{\boldsymbol{\theta}}_{bc})' \right)^{-1}. \quad (6)$$

189 The usual test statistics can be employed for inference. For instance, we may
 190 examine the linear hypothesis $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ with the Wald statistic $(\mathbf{R}\widehat{\boldsymbol{\theta}}_{bc} -$
 191 $\mathbf{r})' \mathbf{V}_{NT}(\widehat{\boldsymbol{\theta}}_{bc})^{-1}(\mathbf{R}\widehat{\boldsymbol{\theta}}_{bc} - \mathbf{r})$.

192 **Remark 1:** Although Theorem 1(i) already implies that the estimator is con-
 193 sistent for any T , we nevertheless provide the limiting distribution for $T \rightarrow \infty$.
 194 First, we note that the estimator is consistent for $\kappa = 0$, which includes the fixed-
 195 N case. Second, as shown by Hahn and Kuersteiner (2002) and Bai (2013), the
 196 asymptotic variance in Theorem 1(ii) equals the lower variance bound, which
 197 is equivalent to the asymptotic variance in the case of no individual-specific ef-
 198 fects μ_i . Therefore, the estimator is efficient when $T \rightarrow \infty$ at a rate not slower
 199 than N . Importantly, the estimator does not involve an asymptotic bias and,
 200 in contrast to the WG and GMM estimators, inference is valid for all values of
 201 κ . This finding suggests that the estimator is particularly attractive in macro
 202 panels, where N and T are of a similar magnitude (cf. Breitung, 2015).

203 **Remark 2:** Under Assumption 1, the bias-corrected method of moments esti-
 204 mator is robust against heteroskedasticity across panel units. When we relax
 205 the assumption of constant error variances over time within panel units, we
 206 notice that the unaccounted bias becomes less severe as T becomes large. Let
 207 $\mathbb{E}[u_{it}^2] = \sigma_{it}^2 < C$ for some $C < \infty$. Then,

$$\begin{aligned} \mathbb{E}[g_{\alpha, NT}(\boldsymbol{\theta}_0)] &= \frac{1}{N} \sum_{i=1}^N \left[-\frac{1}{T^2} \left(\sum_{t=1}^{T-1} \sigma_{it}^2 + \alpha_0 \sum_{t=1}^{T-2} \sigma_{it}^2 + \dots + \alpha^{T-3}(\sigma_{i1}^2 + \sigma_{i2}^2) + \alpha^{T-2} \sigma_{i1}^2 \right) \right] \\ &= b_T(\alpha_0) \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sigma_{it}^2 - \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T^2} \sum_{t=0}^{T-2} \left(\sigma_{i, T-t-1}^2 - \frac{1}{T} \sum_{s=1}^T \sigma_{is}^2 \right) \sum_{s=0}^t \alpha_0^s \right] \\ &= b_T(\alpha_0) \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sigma_{it}^2 - O(T^{-2}), \end{aligned}$$

208 since $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=0}^{T-2} (\sigma_{i, T-t-1}^2 - T^{-1} \sum_{s=1}^T \sigma_{is}^2) = 0$ and $\sum_{s=0}^t \alpha_0^s \leq \max\{1, (1 -$
 209 $\alpha_0)^{-1}\}$, provided that $|\alpha_0| < 1$. Because the remainder term vanishes at a faster
 210 rate than $b_T(\alpha_0)$, the bias correction remains valid under temporal heteroskedas-
 211 ticity when T is large.

212 **Remark 3:** Assumption 1 remains silent about the relationship between the
 213 strictly exogenous regressors \mathbf{x}_{it} and the individual-specific effects μ_i . The setup
 214 thus corresponds to a dynamic fixed-effects panel data model. An advantage

215 of the method of moments estimator is its adaptability to a random-effects
 216 framework by imposing an additional assumption that the individual-specific
 217 intercept μ_i is a random variable with $\mathbb{E}[\mu_i] = 0$, $\mathbb{E}[\mu_i^2] = \sigma_\mu^2$, and $\mathbb{E}[\mathbf{x}_{it}\mu_i] = \mathbf{0}$
 218 for all i and t . The mean zero assumption is without loss of generality if a
 219 constant intercept is included in \mathbf{x}_{it} . We can then replace $\mathbf{m}_{\beta, T_i}(\boldsymbol{\theta})$ in the
 220 moment conditions (3) by

$$\mathbf{m}_{\beta, T_i}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it} e_{it}(\boldsymbol{\theta}).$$

221 This yields a more efficient estimator. In practice, the random-effects assump-
 222 tion may be imposed on all or a subset of the regressors in \mathbf{x}_{it} . It also allows
 223 the regressors to be invariant over time. Dropping the mean adjustment in the
 224 definition of $\boldsymbol{\Sigma}_T$, the limiting distribution of $\hat{\boldsymbol{\theta}}_{bc}$ for fixed T is a straightforward
 225 corollary of Theorem 1(i).

226 3. Relationship to maximum likelihood estimation

227 In this section, we show that the ML estimation procedures proposed by
 228 Hsiao et al. (2002) and Bai (2013) can be seen in a similar light as our bias-
 229 corrected method of moments estimator. In particular, we demonstrate that
 230 their first-order conditions with respect to α can be decomposed into two terms,
 231 $g_{\alpha, NT}(\boldsymbol{\theta})$ and a bias correction term. Accordingly, the main differences between
 232 those approaches are the assumptions on the initial conditions that result in
 233 variations of the bias correction term. An important advantage of our method
 234 of moments approach is that we do not need to impose any assumption on the
 235 initial observations y_{i0} other than its finite variance and exogeneity with respect
 236 to all subsequent errors $u_{i1}, u_{i2}, \dots, u_{iT}$.

237 To simplify the discussion, we focus on the simple AR(1) process with-
 238 out exogenous regressors and with homoskedastic errors. The (Gaussian) log-
 239 likelihood function (conditional on the initial observations) in equation (1) can
 240 be rewritten as

$$\begin{aligned}
\ell(\alpha, \sigma^2, \boldsymbol{\mu}) &= -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T [u_{it}(\alpha, \mu_i) - \bar{u}_i(\alpha, \mu_i)]^2 - \frac{T}{2\sigma^2} \sum_{i=1}^N \bar{u}_i(\alpha, \mu_i)^2, \\
&= -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T [e_{it}(\alpha) - \bar{e}_i(\alpha)]^2 - \frac{T}{2\sigma^2} \sum_{i=1}^N \bar{u}_i(\alpha, \mu_i)^2,
\end{aligned} \tag{7}$$

241 where $\bar{u}_i(\alpha, \mu_i) = T^{-1} \sum_{t=1}^T u_{it}(\alpha, \mu_i)$. Profiling out $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_N)'$ is
242 equivalent to dropping the last term in the log-likelihood function, which then
243 results in the WG estimator that is known to be biased. The first derivative
244 with respect to α (divided by N and T) is given by

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

245 Since $\mathbb{E}[\bar{y}_{-1,i} \bar{u}_i] = -b_T(\alpha_0)\sigma^2$, replacing the last term with $-b_T(\alpha)$ yields the
246 nonlinear moment conditions from the previous section.

247 Hsiao et al. (2002) treat the initial observations as a random variable with
248 $\mathbb{E}[\Delta y_{i1}] = b$ and $\mathbb{E}[(\Delta y_{i1} - b)^2] = \omega\sigma^2$, where Δ is the first-difference operator.
249 From $y_{i1} - y_{i0} = u_{i1} - u_{i0} + b$ and $\mathbb{E}[u_{i0}u_{i1}] = 0$, it follows that $\mathbb{E}[u_{i0}^2] = (\omega - 1)\sigma^2$.
250 The log-likelihood function becomes

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

251 where $\mathbf{e}_i^* = (e_{i0}, e_{i1}, \dots, e_{iT})'$, $\mathbf{D} = (\mathbf{0}, \mathbf{I}_T) - (\mathbf{I}_T, \mathbf{0})$ is a $T \times (T + 1)$ transfor-
252 mation matrix, $\boldsymbol{\Omega} = \mathbf{D}\mathbf{D}' - (2 - \omega)\boldsymbol{\varphi}\boldsymbol{\varphi}'$ is a $T \times T$ matrix with determinant
253 $|\boldsymbol{\Omega}| = T(\omega - 1) + 1$, and $\boldsymbol{\varphi}$ is the first column of the $T \times T$ identity matrix \mathbf{I}_T .
254 For notational convenience, we are using the short-hand notation $\mathbf{e}_i^* = \mathbf{e}_i^*(\alpha, b)$
255 and $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\omega)$. With

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

256 and $\mathbf{D}'(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D} = \mathbf{I}_{T+1} + (T + 1)^{-1}\boldsymbol{\nu}_{T+1}\boldsymbol{\nu}_{T+1}'$, where $\boldsymbol{\nu}_{T+1}$ is a $(T + 1) \times 1$
257 vector of ones, the log-likelihood function can be rewritten as

$$\begin{aligned}
\ell(\alpha, b, \sigma^2, \omega) &= -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=0}^T (u_{it} - \bar{u}_i^*)^2 \\
&\quad - \frac{N}{2} \ln(|\mathbf{\Omega}|) - \frac{(2-\omega)(T+1)}{2\sigma^2|\mathbf{\Omega}|} \sum_{i=1}^N (u_{i0} - \bar{u}_i^*)^2 \\
&= -\frac{NT}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (e_{it} - \bar{e}_i)^2 \\
&\quad - \frac{N}{2} \ln(|\mathbf{\Omega}|) - \frac{T}{2\sigma^2|\mathbf{\Omega}|} \sum_{i=1}^N (u_{i0} - \bar{u}_i)^2, \tag{8}
\end{aligned}$$

258 using the relationship $(T+1)(u_{i0} - \bar{u}_i^*) = T(u_{i0} - \bar{u}_i)$ between the within-
259 group average $\bar{u}_i^* = (T+1)^{-1} \sum_{t=0}^T u_{it}$ that includes the initial value u_{i0} and
260 $\bar{u}_i = T^{-1} \sum_{t=1}^T u_{it}$ that does not.

261 Hsiao et al. (2002) effectively replace the last term in the log-likelihood
262 function (7) with the adjustment terms in the last line of equation (8). The first
263 derivative with respect to α becomes

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

264 In Appendix A.2, we show that $\mathbb{E}[(y_{i0} - \bar{y}_{-1,i})(u_{i0} - \bar{u}_i)] = -b_T(\alpha_0)\sigma^2|\mathbf{\Omega}|$. Thus,
265 the bias corrections of the Hsiao et al. (2002) ML estimator and our method
266 of moments estimator are asymptotically equivalent, in line with observations
267 made earlier by Bun et al. (2017). However, the ML estimator has the important
268 drawback that it requires estimation of the additional nuisance parameters b
269 and ω , even though the bias does not involve these parameters. Moreover, a
270 violation of the initial-observations condition, for instance $\mathbb{E}[\Delta y_{i1}] = b_i \neq b$,
271 would turn this ML estimator inconsistent for fixed T . Not surprisingly given
272 the similarity of the bias correction, Bun et al. (2017) and Juodis (2018) find
273 that this transformed ML approach also exhibits multiple solutions and the
274 estimator may converge to a solution different from the global maximum. This
275 issue can be nonnegligible when T is small.

276 A similar logic applies to the ML framework of Bai (2013). In the AR(1)
277 model without exogenous variables, the individual-specific effects μ_i can be
278 treated as random effects with $\mathbb{E}[\mu_i] = \mu$ and $\mathbb{E}[(\mu_i - \mu)^2] = \zeta$. As an ini-

279 tial condition, Bai (2013) assumes $y_{i0} = 0$ for all i . After profiling out μ , the
 280 first derivative of the Gaussian log-likelihood function results as

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

281 Since $\mathbb{E}[\bar{y}_{-1,i} \bar{e}_i] = -b_T(\alpha_0)(\zeta T + \sigma^2)$, the second term can again be interpreted
 282 as an asymptotically equivalent bias correction term for the first-order condi-
 283 tion, provided that the initial condition $y_{i0} = 0$ holds. As discussed by Bai
 284 (2013, Supplementary Appendix), arbitrary initial conditions can be accommo-
 285 dated by rewriting the model in terms of deviations from the initial observation,
 286 $y_{it} - y_{i0}$. This resembles taking differences as in the approach by Hsiao et al.
 287 (2002). A less restrictive initial condition is underlying the random-effects ML
 288 estimator of Alvarez and Arellano (2003), which can also be regarded as a bias-
 289 corrected estimator as emphasized by Bun et al. (2017). The latter authors
 290 further highlight that the Hsiao et al. (2002) fixed-effects ML estimator is a
 291 restricted version of the random-effects ML estimator.

292 4. Higher-order dynamics

293 Similar to Dhaene and Jochmans (2016), we can extend the bias-corrected
 294 method of moments estimator to an autoregressive model of order p . To sim-
 295 plify the exposition, strictly exogenous regressors are initially neglected. The
 296 respective moment functions $\mathbf{m}_{\beta, T_i}(\boldsymbol{\theta})$ would remain the same as in the AR(1)
 297 model. Consider the AR(p) model

$$y_{it} = \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \dots + \alpha_p y_{i,t-p} + \mu_i + u_{it}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N.$$

298 For notational convenience, we suppress the subscript 0 in denoting the true
 299 coefficient values in this section. Assumption 1 is slightly amended to account
 300 for the additional initial values:

301 **Assumption 2:** Assumption 1 (i)–(iii) continue to hold. (iv) For all $s \in \{1 -$
 302 $p, \dots, -1, 0\}$, we assume $\mathbb{E}[y_{is}^2] < \infty$ for all i and $\mathbb{E}[y_{is} u_{it}] = 0$ for all i and
 303 $t \in \{1, 2, \dots, T\}$.

304 It is convenient to write the model in companion form

$$\mathbf{A}_T(\boldsymbol{\alpha})\mathbf{y}_i = \mathbf{C}_T(\boldsymbol{\alpha})\mathbf{y}_i^0 + \mu_i\boldsymbol{\nu}_T + \mathbf{u}_i,$$

305 where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$, $\mathbf{y}_i^0 = (y_{i,1-p}, \dots, y_{i,-1}, y_{i0})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{iT})'$,
 306 and

$$\mathbf{A}_T(\boldsymbol{\alpha}) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -\alpha_1 & 1 & \ddots & & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & & \vdots \\ -\alpha_p & -\alpha_{p-1} & & \ddots & \ddots & & \vdots \\ 0 & -\alpha_p & & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & 1 & 0 \\ 0 & \cdots & 0 & -\alpha_p & \cdots & -\alpha_1 & 1 \end{pmatrix}, \quad \mathbf{C}_T(\boldsymbol{\alpha}) = \begin{pmatrix} \alpha_p & \alpha_{p-1} & \cdots & \alpha_1 \\ 0 & \alpha_p & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{p-1} \\ \vdots & & \ddots & \alpha_p \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix},$$

307 with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$. The dimensions of the matrices $\mathbf{A}_T(\boldsymbol{\alpha})$ and $\mathbf{C}_T(\boldsymbol{\alpha})$
 308 are $T \times T$ and $T \times p$, respectively. Accordingly,

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

309 With the $T \times T$ matrix

$$\mathbf{L}_T^{(l)} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{T-l} & \mathbf{0} \end{pmatrix},$$

310 we obtain for the $T \times 1$ vector of lags $\mathbf{y}_{-l,i} = \mathbf{L}_T^{(l)}\mathbf{y}_i = (0, \dots, 0, y_{i1}, y_{i2}, \dots, y_{i,T-l})'$
 311 that

$$\mathbb{E} \left[\frac{1}{T} \mathbf{y}'_{-l,i} \mathbf{M}_T \mathbf{u}_i \right] = \frac{\sigma_i^2}{T} \text{tr} \left(\mathbf{L}_T^{(l)} \mathbf{A}_T(\boldsymbol{\alpha})^{-1} \mathbf{M}_T \right) = b_T^{(l)}(\boldsymbol{\alpha}) \sigma_i^2, \quad l \geq 1,$$

312 where $\mathbf{M}_T = \mathbf{I}_T - T^{-1}\boldsymbol{\nu}_T\boldsymbol{\nu}_T'$ is the $T \times T$ projection matrix that creates de-
 313 viations from within-group means, and $b_T^{(l)}(\boldsymbol{\alpha}) = -T^{-2}\boldsymbol{\nu}_T'\mathbf{L}_T^{(l)}\mathbf{A}_T(\boldsymbol{\alpha})^{-1}\boldsymbol{\nu}_T$. For
 314 $|\sum_{j=1}^p \alpha_j| < 1$, $b_T^{(l)}(\boldsymbol{\alpha}) = -(T \sum_{j=1}^p \alpha_j)^{-1} + O(T^{-2})$. For the AR(1) model, it
 315 is easy to see that

$$\mathbf{A}_T(\alpha)^{-1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \alpha & 1 & \ddots & & \vdots \\ \alpha^2 & \alpha & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \alpha^{T-1} & \alpha^{T-2} & \cdots & \alpha & 1 \end{pmatrix}, \quad \mathbf{L}_T^{(1)} \mathbf{A}_T(\alpha)^{-1} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & & & \vdots \\ \alpha & 1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha^{T-2} & \alpha^{T-3} & \cdots & 1 & 0 \end{pmatrix},$$

316 such that $\boldsymbol{\nu}_T' \mathbf{L}_T^{(1)} \mathbf{A}_T(\alpha)^{-1} \boldsymbol{\nu}_T = \sum_{t=0}^{T-2} \sum_{s=0}^t \alpha^s$.

317 **Remark 4:** An interpretation of the elements of matrix $\mathbf{A}_T(\alpha)^{-1}$ can be ob-
318 tained from the moving-average representation of the AR(p) model. Let

$$y_{it} - \mathbb{E}[y_{it} | \mu_i, \mathbf{x}_{it}, \mathbf{x}_{i,t-1}, \dots] = \phi_0 u_{it} + \phi_1 u_{i,t-1} + \phi_2 u_{i,t-2} + \dots,$$

319 with $\phi_0 = 1$. For $j \geq k$, the element in the j -th row and k -th column of
320 $\mathbf{A}_T(\alpha)^{-1}$ equals ϕ_{j-k} . All other elements equal zero.

321 The moment functions for the bias-corrected estimator in the AR(p) model
322 result as

$$m_{\alpha_l, T_i}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T (y_{i,t-l} - \bar{y}_{-l,i}) e_{it}(\boldsymbol{\theta}) - b_T^{(l)}(\boldsymbol{\alpha}) \hat{\sigma}_{T_i}^2(\boldsymbol{\theta}), \quad l = 1, 2, \dots, p,$$

323 with $\hat{\sigma}_{T_i}^2(\boldsymbol{\theta})$ given in equation (2). Let $\mathbf{m}_{\alpha, T_i}(\boldsymbol{\theta}) = (m_{\alpha_1, T_i}(\boldsymbol{\theta}), m_{\alpha_2, T_i}(\boldsymbol{\theta}), \dots, m_{\alpha_p, T_i}(\boldsymbol{\theta}))'$
324 and $\mathbf{b}_T(\boldsymbol{\alpha}) = (b_T^{(1)}(\boldsymbol{\alpha}), b_T^{(2)}(\boldsymbol{\alpha}), \dots, b_T^{(p)}(\boldsymbol{\alpha}))'$. For the model with strictly ex-
325 ogenous regressors \mathbf{x}_{it} , the complete set of moment conditions can be written
326 compactly as

$$\mathbb{E}[\mathbf{m}_{T_i}(\boldsymbol{\theta}_0)] = \mathbb{E} \left[\begin{pmatrix} \mathbf{m}_{\alpha, T_i}(\boldsymbol{\theta}_0) \\ \mathbf{m}_{\beta, T_i}(\boldsymbol{\theta}_0) \end{pmatrix} \right] = \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T [\mathbf{z}_{it}(\boldsymbol{\theta}_0) - \bar{\mathbf{z}}_i(\boldsymbol{\theta}_0)] e_{it}(\boldsymbol{\theta}_0) \right] = \mathbf{0},$$

327 with

$$\mathbf{z}_{it}(\boldsymbol{\theta}) = \mathbf{w}_{it} - \frac{T}{T-1} \begin{pmatrix} \mathbf{b}_T(\boldsymbol{\alpha}) \\ \mathbf{0} \end{pmatrix} e_{it}(\boldsymbol{\theta}),$$

328 $\bar{\mathbf{z}}_i(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \mathbf{z}_{it}(\boldsymbol{\theta})$, and $\mathbf{w}_{it} = (y_{i,t-1}, y_{i,t-2}, \dots, y_{i,t-p}, \mathbf{x}'_{it})'$. The bias-
 329 corrected method of moments estimator solves $N^{-1} \sum_{i=1}^N \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$. The
 330 computational details are similar to the AR(1) case in Section 2.

331 5. Cross-sectional dependence

332 In many macroeconomic applications, it is reasonable to assume that the
 333 elements of the error vector $\mathbf{u}_t = (u_{1t}, u_{2t}, \dots, u_{Nt})'$ are correlated:

334 **Assumption 3:** (i) The errors u_{it} are independent across t but dependent across
 335 i with $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}$ and $\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \boldsymbol{\Sigma}_{u,t}$ for all $t \in \{1, 2, \dots, T\}$. The largest
 336 eigenvalue of the positive-definite matrices $\boldsymbol{\Sigma}_{u,1}, \boldsymbol{\Sigma}_{u,2}, \dots, \boldsymbol{\Sigma}_{u,T}$ is bounded as
 337 $N \rightarrow \infty$. Assumption 1 (ii)–(iv) continue to hold.

338 For notational convenience, we focus on the AR(1) model. The results con-
 339 tinue to hold for the AR(p) model. Although the bias-corrected estimator re-
 340 mains consistent under cross correlation, the estimator of the covariance matrix
 341 in equation (6) is biased as, in general,

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

342 We need to estimate the expression on the left-hand side consistently. We can
 343 rewrite the moment functions as

$$\mathbf{m}_{NT}(\boldsymbol{\theta}) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbf{z}_{it}(\boldsymbol{\theta}) [u_{it}(\boldsymbol{\theta}) - \bar{u}_i(\boldsymbol{\theta})] = \frac{1}{NT} \sum_{t=1}^T \mathbf{Z}_t(\boldsymbol{\theta})' [\mathbf{u}_t(\boldsymbol{\theta}) - \bar{\mathbf{u}}(\boldsymbol{\theta})],$$

344 with $\mathbf{Z}_t(\boldsymbol{\theta}) = (\mathbf{z}_{1t}(\boldsymbol{\theta}), \mathbf{z}_{2t}(\boldsymbol{\theta}), \dots, \mathbf{z}_{Nt}(\boldsymbol{\theta}))'$ and $\bar{\mathbf{u}}(\boldsymbol{\theta}) = (\bar{u}_1(\boldsymbol{\theta}), \bar{u}_2(\boldsymbol{\theta}), \dots, \bar{u}_N(\boldsymbol{\theta}))'$.
 345 Furthermore, let $\mathbf{W}_t = (\mathbf{w}_{1t}, \mathbf{w}_{2t}, \dots, \mathbf{w}_{Nt})'$. Since \mathbf{u}_t is independent across t ,
 346 we can obtain the following asymptotic result.

347 **Theorem 2:** Under Assumption 3 and with $|\alpha_0| < 1$, the limiting distribution
 348 of $\hat{\boldsymbol{\theta}}_{bc}$ as $N, T \rightarrow \infty$, $N/T \rightarrow \kappa$ for $0 \leq \kappa < \infty$, is given by

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{-1} \mathbf{S}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}^{-1}),$$

349 with $\boldsymbol{\Sigma}$ defined in Theorem 1 and

$$\mathbf{S}(\boldsymbol{\theta}_0) = \text{plim}_{N,T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T \mathbf{Z}_t(\boldsymbol{\theta}_0)' \boldsymbol{\Sigma}_{u,t} \mathbf{Z}_t(\boldsymbol{\theta}_0) = \text{plim}_{N,T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T \mathbf{W}_t' \boldsymbol{\Sigma}_{u,t} \mathbf{W}_t.$$

350 Under weakly cross-sectionally dependent errors and large- T asymptotics,
 351 $\mathbf{S}(\boldsymbol{\theta}_0)$ can be consistently estimated by the cluster-robust finite-sample analog

$$\mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}_{bc}) = \frac{1}{NT} \sum_{t=1}^T \mathbf{Z}_t(\hat{\boldsymbol{\theta}}_{bc})' \left(\mathbf{e}_t(\hat{\boldsymbol{\theta}}_{bc}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}}_{bc}) \right) \left(\mathbf{e}_t(\hat{\boldsymbol{\theta}}_{bc}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}}_{bc}) \right)' \mathbf{Z}_t(\hat{\boldsymbol{\theta}}_{bc}),$$

352 where $\mathbf{e}_t(\hat{\boldsymbol{\theta}}_{bc}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}}_{bc}) = \mathbf{u}_t(\hat{\boldsymbol{\theta}}_{bc}) - \bar{\mathbf{u}}(\hat{\boldsymbol{\theta}}_{bc})$ is an $N \times 1$ vector of mean-adjusted
 353 regression residuals.

354 **Remark 5:** The cluster-robust approach to the estimation of the covariance
 355 matrix runs into difficulties if the cross-sectional dependence is due to common
 356 factors. Assume that $u_{it} = \lambda_i f_t + \epsilon_{it}$, where f_t and ϵ_{it} are i.i.d. sequences
 357 with $\mathbb{E}[f_t^2] = \sigma_f^2$ and $\mathbb{E}[\epsilon_{it}^2] = \sigma_\epsilon^2$. Accordingly, $\boldsymbol{\Sigma}_{u,t} = \sigma_f^2 \boldsymbol{\lambda} \boldsymbol{\lambda}' + \sigma_\epsilon^2 \mathbf{I}_N$, where
 358 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_N)$. Since $(NT)^{-1} \sum_{t=1}^T \mathbf{Z}_t(\boldsymbol{\theta})' \boldsymbol{\lambda} \boldsymbol{\lambda}' \mathbf{Z}_t(\boldsymbol{\theta}) = O_p(N)$, in general,
 359 the estimator $\hat{\boldsymbol{\theta}}_{bc}$ is no longer \sqrt{NT} -consistent. This is not surprising as the
 360 least-squares estimator (as well as the bias-corrected estimator) is inconsistent
 361 for T fixed and may even be inconsistent for $T \rightarrow \infty$; see Phillips and Sul
 362 (2007) and Everaert and De Groote (2016). Following a similar approach as
 363 Bun and Carree (2005), De Vos and Everaert (2021) propose a bias-corrected
 364 common correlated effects estimator. Applying the same logic as in our paper,
 365 their estimating equation could also be used to construct a method of moments
 366 estimator.

367 Summing up, our results imply that it is straightforward to adopt the cluster-
 368 robust statistical inference whenever the errors are weakly dependent across the
 369 panel units. Since Theorem 2 requires $T \rightarrow \infty$, the number of time periods
 370 should be “sufficiently large” to ensure reliable asymptotic inference, although
 371 the bias-corrected estimator remains consistent for fixed T and $N \rightarrow \infty$ under
 372 weak cross-sectional dependence. The reason is that the number of clusters, T ,
 373 needs to go to infinity for valid inference. Similar conclusions were obtained
 374 for the ordinary within-group estimator for the static panel data model (cf.
 375 Breitung, 2015, Sec. 15.4.3).

376 **6. Small-sample properties**

377 To assess the small-sample properties of the bias-corrected method of mo-
 378 ments estimator in comparison to alternative estimators that have been sug-
 379 gested in the literature, we perform some Monte Carlo experiments. In the
 380 baseline scenario, the data are generated from a simplified homoskedastic ver-
 381 sion of the dynamic panel data model considered by Kiviet et al. (2017) in their
 382 simulation exercise:

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \sigma_\mu \mu_i + \sigma_u u_{it},$$

$$x_{it} = \gamma x_{i,t-1} + \pi_\mu \mu_i + \pi_\lambda \lambda_i + \sigma_\epsilon \epsilon_{it}.$$

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

389 We distinguish between a process with moderate persistence, $\alpha = 0.4$, and
 390 high persistence, $\alpha = 0.9$. The process is initialized at $t = -50$ with $y_{i,-50} =$
 391 $x_{i,-50} = 0$, and the first 50 observations are discarded. As a robustness check,
 392 we also consider the initialization without burn-in period, $y_{i0} = x_{i0} = 0$, which
 393 implies that the observed process starts off its stationary path.

394 For the experiments with higher-order dynamics, we modify the above data-
 395 generating process as follows:

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

396 and set $(\alpha_1, \alpha_2, \alpha_3) = (0.48, -0.2, 0.12)$ to achieve $\sum_{j=1}^3 \alpha_j = 0.4$ and $(\alpha_1, \alpha_2, \alpha_3) =$
 397 $(1.08, -0.45, 0.27)$ to obtain $\sum_{j=1}^3 \alpha_j = 0.9$. All other parameter values are left
 398 unchanged.

399 To analyze the estimators' performance under heteroskedasticity or cross-
 400 sectional error dependence, we consider the following modifications of the data-
 401 generating process, where the parameterizations ensure that $Var(u_{it}) = 1$ to
 402 keep the signal-to-noise ratio unaffected. A data-generating process with het-
 403 eroskedasticity across both dimensions is obtained by replacing $u_{it} = \sqrt{3/4} \delta_i \tau_t v_{it}$.

404 Uniform cross-sectional dependence is introduced by modifying $u_{it} = \sqrt{3/(4N)} \sum_{j=1}^N \omega_{ij} v_{jt}$,
 405 and interactive random effects are modeled by letting $u_{it} = \sqrt{3/7}(\delta_i \tau_t + v_{it})$.
 406 In all specifications, v_{it} and τ_t are independent standard normally distributed,
 407 and δ_i and ω_{ij} are uniformly distributed over the interval $(0, 2)$.

408 While our bias-correction approach assumes that the regressor x_{it} is strictly
 409 exogenous, we also analyse its robustness when the regressor is instead pre-
 410 determined. In the empirical practice, the hypothesis of strictly exogenous
 411 regressors can be tested by adopting the tests considered in Su et al. (2016)
 412 and Mayer (2020). To generate a predetermined regressor, we replace ϵ_{it} in the
 413 data-generating process by $\rho u_{i,t-1} + \sqrt{1 - \rho^2} \epsilon_{it}$, and we set $\rho = 0.4/\sigma_\epsilon$.

414 We compare the performance of the uncorrected within-groups estimator
 415 (WG), our bias-corrected estimator (BC), the one-step Arellano and Bond (1991)
 416 difference GMM estimator (AB-GMM), the two-step Ahn and Schmidt (1995)
 417 GMM estimator (AS-GMM) with additional nonlinear moment conditions that
 418 are valid under the absence of serial error correlation, the two-step Blundell
 419 and Bond (1998) system GMM estimator (BB-GMM) with additional linear
 420 moment conditions valid under mean stationarity, and the Hsiao et al. (2002)
 421 QML estimator. In addition to the average bias and root mean square error
 422 (RMSE), we report the empirical size of Wald tests given a nominal size of 5%.
 423 We consider a fixed- T robust variance-covariance estimator clustered at the in-
 424 dividual level and, for the WG and BC estimators in the simulation designs
 425 with cross-sectional dependence, a large- T robust variance-covariance estimator
 426 clustered at the time periods. For the AS-GMM and the BB-GMM estimators,
 427 the finite-sample Windmeijer (2005) correction is applied. In our baseline speci-
 428 fication, we also add the Kiviet (1995) bias-corrected estimator (K-BC), but for
 429 computational simplicity we refrain from obtaining bootstrap-based standard
 430 errors. Detailed information can be found in the Supplementary Appendix.

431 We consider all sample size combinations of $T \in \{5, 10, 25, 50\}$ and $N \in$
 432 $\{50, 200\}$. The results are based on 1,000 replications for each simulation design.
 433 As discussed in Section 2, for the BC estimator the numerical algorithm might
 434 converge to a local minimum of the adjusted likelihood function. We observe a
 435 small fraction of such incorrect solutions, especially when T is as small as 5 or
 436 10. The problem disappears when T becomes large. If the negativity condition
 437 for the gradient is violated, we re-initialize the algorithm with a random draw for
 438 $\alpha^{(0)}$ from the uniform distribution over the interval $(0, 1)$. If necessary, we repeat
 439 this process until a solution is found that satisfies the negativity condition. In
 440 our experience, this procedure effectively prevents inappropriate solutions. An

441 additional grid search was never necessary in any of our simulations.

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

	WG	BC	K-BC	AB-GMM	AS-GMM	BB-GMM	QML
$\alpha = 0.4$							
$T = 5$							
Bias	-0.077	0.001	-0.003	-0.038	0.067	0.097	-0.003
RMSE	0.086	0.041	0.042	0.117	0.150	0.140	0.041
Size	0.549	0.078	NA	0.065	0.157	0.272	0.076
$T = 10$							
Bias	-0.034	0.000	0.000	-0.020	0.109	0.129	-0.006
RMSE	0.041	0.023	0.023	0.055	0.145	0.143	0.023
Size	0.351	0.062	NA	0.069	0.287	0.624	0.072
$T = 25$							
Bias	-0.014	-0.001	-0.001	-0.008	0.094	0.142	-0.004
RMSE	0.019	0.013	0.013	0.023	0.108	0.147	0.013
Size	0.193	0.058	NA	0.066	0.452	0.956	0.074
$T = 50$							
Bias	-0.006	0.000	0.000	-0.005	0.078	0.134	-0.002
RMSE	0.011	0.009	0.009	0.014	0.087	0.138	0.009
Size	0.124	0.066	NA	0.064	0.481	0.968	0.065
$\alpha = 0.9$							
$T = 5$							
Bias	-0.433	-0.034	-0.227	-0.442	-0.044	0.018	-0.003
RMSE	0.438	0.124	0.247	0.561	0.205	0.091	0.156
Size	1.000	0.103	NA	0.283	0.112	0.089	0.267
$T = 10$							
Bias	-0.223	-0.004	-0.091	-0.249	0.007	0.031	-0.022
RMSE	0.226	0.067	0.104	0.321	0.084	0.053	0.073
Size	1.000	0.073	NA	0.265	0.053	0.180	0.189
$T = 25$							
Bias	-0.085	0.000	-0.018	-0.078	0.009	0.027	-0.018
RMSE	0.087	0.025	0.028	0.103	0.046	0.037	0.026
Size	0.996	0.052	NA	0.208	0.048	0.135	0.156
$T = 50$							
Bias	-0.039	-0.001	-0.004	-0.033	0.006	0.023	-0.009
RMSE	0.041	0.012	0.013	0.044	0.033	0.035	0.014
Size	0.961	0.067	NA	0.188	0.014	0.065	0.163

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the Kiviet (1995) bias-corrected estimator (K-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 for a Wald test with nominal size of 5%.

442 In the following, we sketch the main findings from our simulation exercise.
 443 For the AR(1) model with stationary initial conditions and i.i.d. errors (Tables
 444 1–4), our BC estimator performs very similar in terms of bias and RMSE to the
 445 transformed QML estimator. This is not surprising given that the latter can
 446 be seen as an estimator that applies a similar bias correction to the first-order

Table 2: Simulation results for β : baseline model (IID), $N = 50$

	WG	BC	K-BC	AB-GMM	AS-GMM	BB-GMM	QML
$\alpha = 0.4$							
$T = 5$							
Bias	-0.005	-0.001	-0.002	-0.022	0.019	-0.025	-0.025
RMSE	0.093	0.093	0.093	0.134	0.148	0.142	0.094
Size	0.060	0.078	NA	0.061	0.054	0.061	0.060
$T = 10$							
Bias	0.015	0.001	0.001	-0.008	0.038	-0.042	-0.018
RMSE	0.062	0.060	0.060	0.075	0.117	0.103	0.062
Size	0.055	0.062	NA	0.047	0.067	0.071	0.068
$T = 25$							
Bias	0.013	0.003	0.003	0.008	0.025	-0.038	-0.006
RMSE	0.037	0.035	0.035	0.044	0.080	0.067	0.035
Size	0.059	0.058	NA	0.043	0.041	0.089	0.051
$T = 50$							
Bias	0.006	0.000	0.000	0.006	0.015	-0.043	-0.004
RMSE	0.026	0.025	0.025	0.031	0.058	0.059	0.025
Size	0.067	0.066	NA	0.058	0.041	0.166	0.064
$\alpha = 0.9$							
$T = 5$							
Bias	-0.048	-0.001	-0.033	-0.058	-0.005	-0.008	-0.033
RMSE	0.102	0.095	0.097	0.124	0.125	0.122	0.094
Size	0.102	0.079	NA	0.088	0.056	0.049	0.071
$T = 10$							
Bias	-0.016	0.002	-0.013	-0.031	0.002	-0.009	-0.039
RMSE	0.064	0.063	0.064	0.086	0.090	0.086	0.071
Size	0.072	0.063	NA	0.094	0.050	0.062	0.139
$T = 25$							
Bias	0.004	0.003	0.001	0.002	0.013	0.003	-0.024
RMSE	0.034	0.033	0.034	0.046	0.055	0.049	0.041
Size	0.042	0.053	NA	0.060	0.045	0.049	0.116
$T = 50$							
Bias	0.004	0.001	0.000	0.005	0.007	-0.002	-0.017
RMSE	0.025	0.024	0.024	0.033	0.041	0.035	0.029
Size	0.063	0.068	NA	0.055	0.046	0.040	0.122

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the Kiviet (1995) bias-corrected estimator (K-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 for a Wald test with nominal size of 5%.

447 condition, as highlighted in Section 3. One should bear in mind, however, that
448 the BC estimator does not require a specific assumption on the initial values
449 of the dynamic process. When $\alpha = 0.4$, its performance is also remarkably
450 similar to the estimator proposed by Kiviet (1995). Under high persistence, the
451 latter is less precise when T is very small. Our BC estimator furthermore has
452 the advantage that it does not depend on the choice of a preliminary estimator

Table 3: Simulation results for α : baseline model (IID), $N = 200$

	WG	BC	K-BC	AB-GMM	AS-GMM	BB-GMM	QML
$\alpha = 0.4$							
$T = 5$							
Bias	-0.079	-0.001	-0.003	-0.008	0.026	0.024	-0.004
RMSE	0.081	0.021	0.022	0.058	0.065	0.054	0.021
Size	0.989	0.054	NA	0.056	0.065	0.081	0.057
$T = 10$							
Bias	-0.035	-0.001	-0.001	-0.005	0.017	0.022	-0.007
RMSE	0.037	0.011	0.012	0.025	0.030	0.034	0.013
Size	0.878	0.052	NA	0.050	0.068	0.135	0.096
$T = 25$							
Bias	-0.013	0.000	0.000	-0.002	0.022	0.036	-0.004
RMSE	0.015	0.007	0.007	0.011	0.027	0.039	0.008
Size	0.511	0.054	NA	0.064	0.327	0.719	0.099
$T = 50$							
Bias	-0.006	0.000	0.000	-0.001	0.032	0.060	-0.002
RMSE	0.008	0.004	0.004	0.007	0.035	0.061	0.005
Size	0.305	0.046	NA	0.059	0.753	1.000	0.050
$\alpha = 0.9$							
$T = 5$							
Bias	-0.430	-0.006	-0.171	-0.191	0.000	0.001	-0.002
RMSE	0.432	0.082	0.183	0.306	0.122	0.058	0.102
Size	1.000	0.087	NA	0.127	0.145	0.076	0.234
$T = 10$							
Bias	-0.221	0.004	-0.062	-0.083	-0.001	0.008	-0.029
RMSE	0.222	0.044	0.069	0.124	0.060	0.033	0.040
Size	1.000	0.054	NA	0.118	0.089	0.090	0.213
$T = 25$							
Bias	-0.084	0.000	-0.010	-0.020	0.006	0.015	-0.016
RMSE	0.084	0.012	0.015	0.035	0.026	0.022	0.019
Size	1.000	0.043	NA	0.095	0.055	0.176	0.396
$T = 50$							
Bias	-0.039	0.000	-0.002	-0.009	0.010	0.022	-0.009
RMSE	0.040	0.006	0.006	0.017	0.019	0.024	0.010
Size	1.000	0.069	NA	0.091	0.118	0.623	0.395

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the Kiviet (1995) bias-corrected estimator (K-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 for a Wald test with nominal size of 5%.

453 and that standard errors are readily available. All GMM estimators perform
454 substantially worse under this baseline data-generating process.

455 Under higher-order dynamics with 3 lags of the dependent variable (Tables
456 5–6 in the Supplementary Appendix), the results suggest that the BC estima-
457 tor effectively removes the bias and yields estimates with the lowest RMSE
458 compared to the GMM estimators. The only exception is the case with high

Table 4: Simulation results for β : baseline model (IID), $N = 200$

	WG	BC	K-BC	AB-GMM	AS-GMM	BB-GMM	QML
$\alpha = 0.4$							
$T = 5$							
Bias	-0.002	0.000	0.000	-0.004	0.011	-0.002	-0.023
RMSE	0.046	0.046	0.046	0.063	0.066	0.060	0.051
Size	0.037	0.054	NA	0.044	0.039	0.035	0.078
$T = 10$							
Bias	0.014	0.000	0.000	-0.002	-0.001	-0.012	-0.020
RMSE	0.033	0.030	0.030	0.038	0.041	0.042	0.036
Size	0.085	0.052	NA	0.058	0.044	0.062	0.099
$T = 25$							
Bias	0.011	0.001	0.001	0.002	0.001	-0.014	-0.008
RMSE	0.021	0.018	0.018	0.022	0.028	0.030	0.019
Size	0.098	0.054	NA	0.044	0.046	0.082	0.053
$T = 50$							
Bias	0.005	0.000	0.000	0.001	0.005	-0.013	-0.005
RMSE	0.014	0.013	0.013	0.016	0.025	0.023	0.014
Size	0.066	0.046	NA	0.052	0.059	0.119	0.070
$\alpha = 0.9$							
$T = 5$							
Bias	-0.048	0.000	-0.024	-0.025	-0.002	-0.004	-0.033
RMSE	0.066	0.047	0.053	0.068	0.060	0.059	0.055
Size	0.179	0.068	NA	0.087	0.043	0.039	0.118
$T = 10$							
Bias	-0.016	0.000	-0.009	-0.010	-0.004	-0.007	-0.038
RMSE	0.034	0.030	0.032	0.041	0.042	0.042	0.047
Size	0.082	0.053	NA	0.059	0.048	0.043	0.272
$T = 25$							
Bias	0.002	0.001	0.000	0.001	0.001	-0.003	-0.026
RMSE	0.018	0.017	0.018	0.023	0.027	0.026	0.031
Size	0.049	0.042	NA	0.038	0.039	0.042	0.307
$T = 50$							
Bias	0.003	0.000	0.000	0.001	0.003	0.000	-0.018
RMSE	0.012	0.012	0.012	0.017	0.019	0.018	0.022
Size	0.053	0.069	NA	0.060	0.057	0.054	0.305

Note: The comparison includes the within-groups estimator (WG), the bias-corrected method of moments estimator (BC), the Kiviet (1995) bias-corrected estimator (K-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 for a Wald test with nominal size of 5%.

459 persistence and $T \leq 10$, where the BB-GMM estimator performs best.

460 When the data-generating process exhibits cross-sectional and time-dependent
461 heteroskedasticity (Tables 7–8 in the Supplementary Appendix), our findings
462 suggest that the BC estimator is robust under moderate persistence even if T
463 is as small as 10. For processes with high persistence, a larger number of time
464 periods, $T \geq 25$, is required to successfully remove the bias. Overall, the BC

465 estimator again performs similar to the QML estimator, and both estimators
466 clearly outperform the GMM estimators in most cases. The BB-GMM estimator
467 again has some advantage under high persistence and small T .

468 The results under cross-sectional dependence (Tables 9–12 in the Supplemen-
469 tary Appendix) suggest that inference based on the i.i.d. assumption may be
470 severely biased when it is violated. For the model with uniform cross-sectional
471 dependence, the Wald tests with nominal size of 5% reject in more than 50%
472 of the cases, and sometimes even more than 90%. On the other hand, the
473 cluster-robust standard errors proposed in Section 5 effectively correct for cross-
474 sectional dependence and yield empirical sizes close to the nominal size. The
475 robustification even works well in models with a common-factor structure, al-
476 though the underlying asymptotic theory underlying Theorem 2 does not apply
477 to models with strong cross-sectional dependence.

478 It is well known that the BB-GMM estimator under a nonstationary initial-
479 ization can be severely biased. This is confirmed by our simulations (Tables
480 13–14 in the Supplementary Appendix). All other estimators are virtually un-
481 biased whenever N is sufficiently large. While the BC and QML estimators
482 perform similarly under a stationary initialization, the BC estimator turns out
483 to be more efficient than the QML estimator of Hsiao et al. (2002) under the
484 nonstationary initialization.

485 We finally study the small-sample properties of the estimators when the re-
486 gressor is predetermined (Tables 15–16 in the Supplementary Appendix). While
487 our bias-corrected estimator and the QML estimator are no longer consistent
488 in this case, in terms of bias and RMSE they are remarkably robust to such a
489 violation of the strict exogeneity assumption. For the QML estimator, a similar
490 observation has been made by Moral-Benito (2013). While there are some no-
491 ticeable size distortions for the BC and QML estimators, the GMM estimators
492 also cannot fully convince even though they can be flexibly modified to accomo-
493 date predetermined regressors. In principle, the bias correction approach can
494 be extended to accomodate predetermined regressors by imposing additional
495 assumptions on the data-generating process of these regressors. This could be
496 achieved by assuming a panel vector autoregressive model as in Juodis (2013),
497 but it would come at the cost of additional parameters to be estimated. An
498 advantage of the latter approach would be that it allows to easily test the strict
499 exogeneity assumption.

500 **7. Conclusion**

501 We proposed an estimator that is based on a simple set of moment condi-
502 tions that can be easily solved with standard numerical optimization procedures.
503 It is straightforward to generalize the estimator to higher-order autoregressive
504 models or dynamic random-effects models. An estimator of the asymptotic co-
505 variance matrix is readily available, as are robust standard errors that effectively
506 adjust for cross-sectional dependence, which is a relevant feature in the analysis
507 of macroeconomic panel data.

508 Besides the added flexibility, the bias-corrected method of moments esti-
509 mator is easier to implement than likelihood-based estimators that implicitly
510 employ a similar bias correction. In our Monte Carlo simulations, our bias-
511 corrected estimator also performs favorably compared to maximum likelihood
512 and generalized method of moments estimators under the assumption of strictly
513 exogenous regressors.

514 **References**

- 515 Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel
516 data. *Journal of Econometrics* 68, 5–27. doi:10.1016/0304-4076(94)01641-C.
- 517 Alvarez, J., Arellano, M., 2003. The time series and cross-section asymp-
518 totics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.
519 doi:10.1111/1468-0262.00441.
- 520 Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error
521 components. *Journal of the American Statistical Association* 76, 598–606.
522 doi:10.1080/01621459.1981.10477691.
- 523 Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data:
524 Monte Carlo evidence and an application to employment equations. *Review*
525 *of Economic Studies* 58, 277–297. doi:10.2307/2297968.
- 526 Bai, J., 2013. Fixed-effects dynamic panel models, a factor analytical method.
527 *Econometrica* 81, 285–314. doi:10.3982/ECTA9409.
- 528 Blundell, R., Bond, S.R., 1998. Initial conditions and moment restrictions
529 in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
530 doi:10.1016/S0304-4076(98)00009-8.

- 531 Breitung, J., 2015. The analysis of macroeconomic panel data, in: Baltagi, B.H.
532 (Ed.), *The Oxford Handbook of Panel Data*. Oxford University Press, Oxford.
533 chapter 15, pp. 453–493. doi:10.1093/oxfordhb/9780199940042.013.0015.
- 534 Bun, M.J.G., Carree, M.A., 2005. Bias-corrected estimation in dynamic
535 panel data models. *Journal of Business & Economic Statistics* 23, 200–210.
536 doi:10.1198/073500104000000532.
- 537 Bun, M.J.G., Carree, M.A., Juodis, A., 2017. On maximum likelihood esti-
538 mation of dynamic panel data models. *Oxford Bulletin of Economics and*
539 *Statistics* 79, 463–494. doi:10.1111/obes.12156.
- 540 Bun, M.J.G., Windmeijer, F., 2010. The weak instrument problem of the system
541 GMM estimator in dynamic panel data models. *Econometrics Journal* 13, 95–
542 126. doi:10.1111/j.1368-423X.2009.00299.x.
- 543 De Vos, I., Everaert, G., 2021. Bias-corrected common correlated effects pooled
544 estimation in dynamic panels. *Journal of Business & Economic Statistics* 39,
545 294–306. doi:10.1080/07350015.2019.1654879.
- 546 Dhaene, G., Jochmans, K., 2016. Likelihood inference in an au-
547 toregression with fixed effects. *Econometric Theory* 32, 1178–1215.
548 doi:10.1017/S0266466615000146.
- 549 Everaert, G., De Groote, T., 2016. Common correlated effects estimation of
550 dynamic panels with cross-sectional dependence. *Econometric Reviews* 35,
551 428–463. doi:10.1080/07474938.2014.966635.
- 552 Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dy-
553 namic panel model with fixed effects when both n and T are large. *Econo-*
554 *metrica* 70, 1639–1657. doi:10.1111/1468-0262.00344.
- 555 Hayakawa, K., Pesaran, M.H., 2015. Robust standard errors in transformed
556 likelihood estimation of dynamic panel data models. *Journal of Econometrics*
557 188, 111–134. doi:10.1016/j.jeconom.2015.03.042.
- 558 Holtz-Eakin, D., Newey, W., Rosen, H.S., 1988. Estimating vector autoregres-
559 sions with panel data. *Econometrica* 56, 1371–1395. doi:10.2307/1913103.
- 560 Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood esti-
561 mation of fixed effects dynamic panel data models covering short time periods.
562 *Journal of Econometrics* 109, 107–150. doi:10.1016/S0304-4076(01)00143-9.

- 563 Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a
564 guide for macroeconomists. *Economics Letters* 65, 9–15. doi:10.1016/S0165-
565 1765(99)00130-5.
- 566 Juodis, A., 2013. A note on bias-corrected estimation in dynamic panel data
567 models. *Economics Letters* 92, 220–227. doi:10.1016/j.econlet.2012.12.013.
- 568 Juodis, A., 2018. First difference transformation in panel VAR models: Ro-
569 bustness, estimation, and inference. *Econometric Reviews* 37, 650–693.
570 doi:10.1080/07474938.2016.1139559.
- 571 Kiviet, J.F., 1995. On bias, inconsistency, and efficiency of various estima-
572 tors in dynamic panel data models. *Journal of Econometrics* 68, 53–78.
573 doi:10.1016/0304-4076(94)01643-E.
- 574 Kiviet, J.F., Pleus, M., Poldermans, R., 2017. Accuracy and efficiency of various
575 GMM inference techniques in dynamic micro panel data models. *Economet-*
576 *rics* 5, 14. doi:10.3390/econometrics5010014.
- 577 Kruiniger, H., 2018. A further look at modified ML estimation of the panel
578 AR(1) model with fixed effects and arbitrary initial conditions. MPRA Paper
579 88623.
- 580 Mayer, A., 2020. (Consistently) testing strict exogeneity against the alternative
581 of predeterminedness in linear time-series models. *Economics Letters* 193.
582 doi:10.1016/j.econlet.2020.109335.
- 583 Moon, H.R., Perron, B., Phillips, P.C.B., 2015. Incidental parameters and
584 dynamic panel modeling, in: Baltagi, B.H. (Ed.), *The Oxford Handbook*
585 *of Panel Data*. Oxford University Press, Oxford. chapter 4, pp. 111–148.
586 doi:10.1093/oxfordhb/9780199940042.013.0004.
- 587 Moral-Benito, E., 2013. likelihood-based estimation of dynamic panels with
588 predetermined regressors. *Journal of Business & Economic Statistics* 31, 451–
589 472. doi:10.1080/07350015.2013.818003.
- 590 Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49,
591 1417–1426. doi:10.2307/1911408.
- 592 Phillips, P.C.B., Sul, D., 2007. Bias in dynamic panel estimation with fixed ef-
593 fects, incidental trends and cross section dependence. *Journal of Econometrics*
594 137, 162–188. doi:10.1016/j.jeconom.2006.03.009.

595 Su, L., Zhang, Y., Wei, J., 2016. A practical test for strict exogeneity in
 596 linear panel data models with fixed effects. *Economics Letters* 147, 27–31.
 597 doi:10.1016/j.econlet.2016.08.012.

598 Windmeijer, F., 2005. A finite sample correction for the variance of linear
 599 efficient two-step GMM estimators. *Journal of Econometrics* 126, 25–51.
 600 doi:10.1016/j.jeconom.2004.02.005.

601 **Appendix A Computational details**

602 *A.1 Gradient of the moment function*

603 For the bias-corrected method of moments estimator presented in Section 2,
 604 the gradient

$$\nabla \mathbf{m}_{NT}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} \nabla_{\alpha} m_{\alpha, Ti}(\boldsymbol{\theta}) & \nabla_{\beta} m_{\alpha, Ti}(\boldsymbol{\theta})' \\ \nabla_{\alpha} \mathbf{m}_{\beta, Ti}(\boldsymbol{\theta}) & \nabla_{\beta} \mathbf{m}_{\beta, Ti}(\boldsymbol{\theta}) \end{pmatrix}$$

605 has the elements

$$\begin{aligned} \nabla_{\alpha} m_{\alpha, Ti}(\boldsymbol{\theta}) &= -\frac{1}{T} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) y_{i,t-1} - \nabla_{\alpha} b_T(\alpha) \hat{\sigma}_{Ti}^2(\boldsymbol{\theta}) - b_T(\alpha) \nabla_{\alpha} \hat{\sigma}_{Ti}^2(\boldsymbol{\theta}), \\ \nabla_{\beta} m_{\alpha, Ti}(\boldsymbol{\theta}) &= -\frac{1}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) y_{i,t-1} - b_T(\alpha) \nabla_{\beta} \hat{\sigma}_{Ti}^2(\boldsymbol{\theta}), \\ \nabla_{\alpha} \mathbf{m}_{\beta, Ti}(\boldsymbol{\theta}) &= -\frac{1}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) y_{i,t-1}, \\ \nabla_{\beta} \mathbf{m}_{\beta, Ti}(\boldsymbol{\theta}) &= -\frac{1}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \mathbf{x}'_{it}, \end{aligned}$$

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

$$\begin{aligned} \nabla_{\alpha} \hat{\sigma}_{Ti}^2(\boldsymbol{\theta}) &= -\frac{2}{T-1} \sum_{t=1}^T (y_{i,t-1} - \bar{y}_{-1,i}) e_{it}(\boldsymbol{\theta}), \\ \nabla_{\beta} \hat{\sigma}_{Ti}^2(\boldsymbol{\theta}) &= -\frac{2}{T-1} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) e_{it}(\boldsymbol{\theta}). \end{aligned}$$

607 Furthermore, $\mathbb{E}[\nabla_{\alpha} \widehat{\sigma}_{T_i}^2(\boldsymbol{\theta}_0)] = 2b_T(\alpha_0)\sigma_i^2 T/(T-1)$ and $\mathbb{E}[\nabla_{\beta} \widehat{\sigma}_{T_i}^2(\boldsymbol{\theta}_0)] = \mathbf{0}$.

608 For the more general AR(p) model considered in Section 4, the gradient can
 609 be expressed compactly as

$$\nabla \mathbf{m}_{NT}(\boldsymbol{\theta}) = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{w}_{it} - \bar{\mathbf{w}}_i) \mathbf{z}_{it}(\boldsymbol{\theta})' + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \nabla \mathbf{z}_{it}(\boldsymbol{\theta}) [e_{it}(\boldsymbol{\theta}) - \bar{e}_i(\boldsymbol{\theta})],$$

610 where

$$\nabla \mathbf{z}_{it}(\boldsymbol{\theta}) = -\frac{T}{T-1} \begin{pmatrix} \nabla_{\alpha} \mathbf{b}_T(\boldsymbol{\alpha}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} e_{it}(\boldsymbol{\theta}) + \frac{T}{T-1} \begin{pmatrix} \mathbf{b}_T(\boldsymbol{\alpha}) \\ \mathbf{0} \end{pmatrix} \mathbf{w}'_{it}.$$

611 The element in the j -th row and k -th column of the $p \times p$ matrix $\nabla_{\alpha} \mathbf{b}_T(\boldsymbol{\alpha})$ equals
 612 $-T^{-2} \boldsymbol{\nu}'_T \mathbf{L}_T^{(k)} [\nabla_{\alpha_j} \mathbf{A}_T(\boldsymbol{\alpha})^{-1}] \boldsymbol{\nu}_T$, with $\nabla_{\alpha_j} \mathbf{A}_T(\boldsymbol{\alpha})^{-1} = -\mathbf{A}_T(\boldsymbol{\alpha}) [\nabla_{\alpha_j} \mathbf{A}_T(\boldsymbol{\alpha})] \mathbf{A}_T(\boldsymbol{\alpha})$.
 613 The matrix $\nabla_{\alpha_j} \mathbf{A}_T(\boldsymbol{\alpha})$ has all elements on the j -th diagonal below the main
 614 diagonal equal to -1 and all other elements equal to 0. Eventually,

$$\mathbb{E}[\nabla \mathbf{m}_{T_i}(\boldsymbol{\theta}_0)] = -\mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T (\mathbf{w}_{it} - \bar{\mathbf{w}}_i) \mathbf{w}'_{it} \right] - \sigma_i^2 \begin{pmatrix} \nabla_{\alpha} \mathbf{b}_T(\boldsymbol{\alpha}_0) - \frac{2T}{T-1} \mathbf{b}_T(\boldsymbol{\alpha}_0) \mathbf{b}_T(\boldsymbol{\alpha}_0)' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

615 A.2 Bias-corrected maximum likelihood estimation

616 In Section 3, we argued that the transformed ML estimator of Hsiao et al.
 617 (2002) leads to a bias-corrected first-order condition similar to that of the bias-
 618 corrected method of moments estimator. This follows from the observation that
 619 for $|\alpha_0| < 1$ we obtain

$$\begin{aligned} \mathbb{E}[(y_{i0} - \bar{y}_{-1,i})(u_{i0} - \bar{u}_i)] &= \left[\frac{\omega - 1}{1 - \alpha_0} - \frac{(\omega - 1)(1 - \alpha_0^T)}{(1 - \alpha_0)^2 T} - b_T(\alpha_0) \right] \sigma^2 \\ &= \left[\frac{|\boldsymbol{\Omega}| - 1}{(1 - \alpha_0)T} \left(1 - \frac{1 - \alpha_0^T}{(1 - \alpha_0)T} \right) - b_T(\alpha_0) \right] \sigma^2 \\ &= [-b_T(\alpha_0)(|\boldsymbol{\Omega}| - 1) - b_T(\alpha_0)] \sigma^2 = -b_T(\alpha_0) \sigma^2 |\boldsymbol{\Omega}|. \end{aligned}$$

620 **Appendix B Proofs**

621 *B.1 Proof of Theorem 1*

622 Let $\widehat{\boldsymbol{\theta}}_{bc} \in \Theta$ be the point-valued minimizer of the objective function $\mathbf{m}_{NT}(\boldsymbol{\theta})' \mathbf{m}_{NT}(\boldsymbol{\theta})$
 623 such that $\text{plim}_{N \rightarrow \infty} \widehat{\boldsymbol{\theta}}_{bc} = \boldsymbol{\theta}_0$. By the mean value theorem,

$$\mathbf{m}_{NT}(\widehat{\boldsymbol{\theta}}_{bc}) = \mathbf{m}_{NT}(\boldsymbol{\theta}_0) + \nabla \mathbf{m}_{NT}(\bar{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0),$$

624 with the mean value $\bar{\boldsymbol{\theta}}$. Due to continuity of the moment function, $\text{plim}_{N \rightarrow \infty} \bar{\boldsymbol{\theta}} =$
 625 $\boldsymbol{\theta}_0$. Thus, with $\mathbf{m}_{NT}(\widehat{\boldsymbol{\theta}}_{bc}) = \mathbf{0}$,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) = -(\nabla \mathbf{m}_{NT}(\bar{\boldsymbol{\theta}}))^{-1} \sqrt{N} \mathbf{m}_{NT}(\boldsymbol{\theta}_0).$$

626 From the results in Appendix A.1 and with continuity of the gradient, it follows
 627 that $\text{plim}_{N \rightarrow \infty} [\nabla \mathbf{m}_{NT}(\bar{\boldsymbol{\theta}})]^{-1} = -[\boldsymbol{\Sigma}_T + \sigma^2 \mathbf{B}_T(\alpha_0)]^{-1}$. Due to the independence
 628 of $\mathbf{m}_{T_i}(\boldsymbol{\theta}_0)$ across i with $\mathbb{E}[\mathbf{m}_{T_i}(\boldsymbol{\theta}_0)] = \mathbf{0}$ and $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mathbf{m}_{T_i}(\boldsymbol{\theta}_0) \mathbf{m}_{T_i}(\boldsymbol{\theta}_0)' =$
 629 $\mathbf{S}_T(\boldsymbol{\theta}_0)$, the central limit theorem yields

$$\sqrt{N} \mathbf{m}_{NT}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}_T(\boldsymbol{\theta}_0))$$

630 as $N \rightarrow \infty$. With Slutsky's theorem, the limiting distribution as in Theorem
 631 1(i) follows.

632 Let $y_{it} = \xi_{it}(\boldsymbol{\theta}) + v_{it}(\boldsymbol{\theta})$, where $v_{it} = \sum_{j=0}^{t-1} \alpha^j u_{i,t-j}(\boldsymbol{\theta})$ and $\xi_{it}(\boldsymbol{\theta})$ is a
 633 function of current and past regressors \mathbf{x}_{it} , individual effects μ_i , and initial
 634 conditions y_{i0} . After some algebra, we can rewrite the elements of the moment
 635 function as

$$\begin{aligned} m_{\alpha, NT}(\boldsymbol{\theta}) &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [y_{i,t-1} - \bar{\xi}_{-1,i}(\boldsymbol{\theta})] u_{it}(\boldsymbol{\theta}) - \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T v_{i,t-1}(\boldsymbol{\theta}) u_{is}(\boldsymbol{\theta}) \\ &\quad - \frac{b_T(\alpha)}{N(T-1)} \sum_{i=1}^N \sum_{t=1}^T [u_{it}(\boldsymbol{\theta}) - \bar{u}_i(\boldsymbol{\theta})] u_{it}(\boldsymbol{\theta}), \\ m_{\beta, NT}(\boldsymbol{\theta}) &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) u_{it}(\boldsymbol{\theta}), \end{aligned}$$

636 where $\bar{\xi}_{-1,i}(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^T \xi_{i,t-1}(\boldsymbol{\theta})$. Moreover, $\mathbb{E}[\bar{\xi}_{-1,i}(\boldsymbol{\theta}_0)] = \mathbb{E}[\bar{y}_{-1,i}]$ and

$$\begin{aligned}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T(y_{i,t-1}-\bar{\xi}_{-1,i}(\boldsymbol{\theta}_0))u_{it}(\boldsymbol{\theta}_0)\right] &= 0, \\ \mathbb{E}\left[\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^Tv_{i,t-1}(\boldsymbol{\theta}_0)u_{is}(\boldsymbol{\theta}_0)\right] &= -b_T(\alpha_0)\sigma_i^2, \\ \mathbb{E}\left[\frac{1}{T-1}\sum_{t=1}^T[u_{it}(\boldsymbol{\theta}_0)-\bar{u}_i(\boldsymbol{\theta}_0)]u_{it}(\boldsymbol{\theta}_0)\right] &= \sigma_i^2.\end{aligned}$$

637 As $N, T \rightarrow \infty$, and with $|\alpha_0| < 1$, it follows that

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

638 and therefore

$$\sqrt{NT}\mathbf{m}_{NT}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}).$$

639 Since $b_T(\alpha_0) = O(T^{-1})$, it follows with the results from Appendix A.1 that

$$\nabla\mathbf{m}_{NT}(\boldsymbol{\theta}_0) = -\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\mathbf{w}_{it}-\bar{\mathbf{w}}_i)\mathbf{w}'_{it} + O_p(T^{-1}),$$

640 where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}'_{it})'$ and $\bar{\mathbf{w}}_i = T^{-1}\sum_{t=1}^T\mathbf{w}_{it}$. Consequently, by continuity
641 of the gradient, $\text{plim}_{N,T\rightarrow\infty}[\nabla\mathbf{m}_{NT}(\bar{\boldsymbol{\theta}})]^{-1} = -\boldsymbol{\Sigma}^{-1}$, and the result in Theorem
642 1(ii) follows.

643 *B.2 Proof of Theorem 2*

644 Define $\mathbf{q}_{Nt}(\boldsymbol{\theta}) = N^{-1/2}\sum_{i=1}^N\mathbf{w}_{it}u_{it}(\boldsymbol{\theta})$. As $N, T \rightarrow \infty$ and with $|\alpha_0| < 1$,

$$\frac{1}{NT}\sum_{t=1}^T\mathbf{Z}_t(\boldsymbol{\theta})'[\mathbf{u}_t(\boldsymbol{\theta})-\bar{\mathbf{u}}(\boldsymbol{\theta})][\mathbf{u}_t(\boldsymbol{\theta})-\bar{\mathbf{u}}(\boldsymbol{\theta})]'\mathbf{Z}_t(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^T\mathbf{q}_{Nt}(\boldsymbol{\theta})\mathbf{q}_{Nt}(\boldsymbol{\theta})' + o_p(1).$$

645 Since $u_{it}(\boldsymbol{\theta})$ is independent of \mathbf{w}_{it} , it follows that $\mathbb{E}[\mathbf{q}_{Nt}(\boldsymbol{\theta}_0)] = \mathbf{0}$. Furthermore,
646 if the eigenvalues of $\boldsymbol{\Sigma}_{u,t}$ are bounded for all t , the norm of $\mathbb{E}[\mathbf{q}_{Nt}(\boldsymbol{\theta}_0)\mathbf{q}_{Nt}(\boldsymbol{\theta}_0)']$
647 is bounded as well and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\mathbf{q}_{Nt}(\boldsymbol{\theta}_0) \mathbf{q}_{Nt}(\boldsymbol{\theta}_0)'] = \text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T \mathbf{W}_t' \boldsymbol{\Sigma}_{u,t} \mathbf{W}_t = \mathbf{S}(\boldsymbol{\theta}_0).$$

648 Then,

$$\sqrt{NT} \mathbf{m}_{NT}(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{S}(\boldsymbol{\theta}_0)).$$

649 The remainder of the proof follows along the lines of Appendix B.1.

Bias-corrected method of moments estimators for dynamic panel data models

Jörg Breitung*

Sebastian Kripfganz[†]

Kazuhiko Hayakawa[‡]

Supplementary Appendix

Appendix C Monte Carlo simulation

C.1 Details of the data-generating process

Following Kiviet et al. (2017), we hold fixed the fraction of the variance of x_{it} that is due to the individual-specific effects ($EV F_x$) and the variance fraction of $\pi_\mu \mu_i$ in the composite individual-specific effects (IEF_x^μ). This allows us to endogenously determine

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

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 obtain

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

For details, see Kiviet et al. (2017, Section 4). Fixing $\gamma = 0.4$, $EV F_x = IEF_x^\mu = 0.3$, $DEN_y^\eta = 4$, and $SNR = 5$ implies $\beta \approx 2.044$ under moderate persistence and $\beta \approx 0.307$

*University of Cologne, Institute of Econometrics and Statistics, Universitätsstraße 24, 50923 Cologne, Germany. E-mail: breitung@statistik.uni-koeln.de

[†]University of Exeter Business School, Department of Economics, Streatham Court, Rennes Drive, Exeter, EX4 4PU, United Kingdom. E-mail: S.Kripfganz@exeter.ac.uk

[‡]Hiroshima University, Department of Economics, 1-2-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8525, Japan. E-mail: kazuhaya@hiroshima-u.ac.jp

in the high-persistence case.

C.2 Construction of the estimators

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

- The one-step Arellano and Bond (1991) GMM estimator (AB-GMM) utilizes the moment conditions¹

$$E[y_{i,t-1-s}\Delta e_{it}(\boldsymbol{\theta}_0)] = 0, \quad 1 \leq s \leq 3, \quad t = s + 1, \dots, T,$$

$$E\left[\sum_{t=2}^T \Delta x_{it}\Delta e_{it}(\boldsymbol{\theta}_0)\right] = 0.$$

For the simulation design with predetermined x_{it} , we replace the last moment condition by

$$E[x_{i,t-s}\Delta e_{it}(\boldsymbol{\theta}_0)] = 0, \quad 1 \leq s \leq 3, \quad t = s + 1, \dots, T.$$

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 $E[e_{iT}(\boldsymbol{\theta}_0)\Delta e_{it}(\boldsymbol{\theta}_0)] = 0$, $t = 2, \dots, T - 1$, 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 $E[\Delta y_{i,t-1}e_{it}(\boldsymbol{\theta}_0)] = 0$, $t = 2, \dots, T$, 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

¹We 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.

likelihood (QML) estimator with the initial-observations parameterization

$$E[\Delta y_{i1} | \Delta x_{i1}, \dots, \Delta x_{iT}] = b + \sum_{s=1}^3 \pi_s \Delta x_{is}.$$

Finally, in our baseline specification we also consider the Kiviet (1995) bias-corrected estimator (K-BC), which we initialize with the AB-GMM estimator.

C.3 Estimation of the variance-covariance matrix

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

$$\begin{aligned} \mathbf{V}_{NT}(\hat{\boldsymbol{\theta}}) &= \frac{1}{N} \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})'^{-1}, \\ \mathbf{V}_{NT}^{csd}(\hat{\boldsymbol{\theta}}) &= \frac{1}{NT} \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{S}_{NT}^{csd}(\hat{\boldsymbol{\theta}}) \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})'^{-1}. \end{aligned}$$

Variance-covariance estimator $\mathbf{V}_{NT}(\hat{\boldsymbol{\theta}})$ is the conventional fixed- T cluster-robust estimator clustered at the individual level. Estimator $\mathbf{V}_{NT}^{csd}(\hat{\boldsymbol{\theta}})$ is a robust large- T estimator clustered at the time periods. For the WG estimator, we have

$$\begin{aligned} \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}}) &= \frac{1}{N} \sum_{i=1}^N \nabla \mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}}), \quad \mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^N \mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}}) \mathbf{g}_{Ti}(\hat{\boldsymbol{\theta}})', \\ \mathbf{S}_{NT}^{csd}(\hat{\boldsymbol{\theta}}) &= \frac{1}{NT} \sum_{t=1}^T \mathbf{W}_t' [\mathbf{e}_t(\hat{\boldsymbol{\theta}}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}})] [\mathbf{e}_t(\hat{\boldsymbol{\theta}}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}})]' \mathbf{W}_t. \end{aligned}$$

Similarly, for the BC estimator we compute

$$\begin{aligned} \mathbf{D}_{NT}(\hat{\boldsymbol{\theta}}) &= \frac{1}{N} \sum_{i=1}^N \nabla_{\boldsymbol{\theta}} \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}), \quad \mathbf{S}_{NT}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^N \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}}) \mathbf{m}_{Ti}(\hat{\boldsymbol{\theta}})', \\ \mathbf{S}_{NT}^{csd}(\hat{\boldsymbol{\theta}}) &= \frac{1}{NT} \sum_{t=1}^T \mathbf{Z}_t(\hat{\boldsymbol{\theta}})' [\mathbf{e}_t(\hat{\boldsymbol{\theta}}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}})] [\mathbf{e}_t(\hat{\boldsymbol{\theta}}) - \bar{\mathbf{e}}(\hat{\boldsymbol{\theta}})]' \mathbf{Z}_t(\hat{\boldsymbol{\theta}}). \end{aligned}$$

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

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

For the two-step AS-GMM and BB-GMM estimators with optimal weighting matrix, the variance matrix is estimated as $\mathbf{V}_{NT}(\hat{\boldsymbol{\theta}}) = N^{-1}\mathbf{D}_{NT}(\hat{\boldsymbol{\theta}})^{-1}$, and subsequently 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.

References

- Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68(1), 5–27.
- 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.
- Blundell, R. and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1), 115–143.
- 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.
- 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.
- 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.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126(1), 25–51.

Table 5: 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
Size	0.674	0.092	0.059	0.218	0.531	0.075	0.092	0.073	0.065	0.059
$T = 10$										
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
Size	0.461	0.079	0.054	0.517	0.869	0.059	0.079	0.076	0.094	0.066
$T = 25$										
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
Size	0.249	0.062	0.073	0.765	0.992	0.063	0.062	0.063	0.117	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
Size	0.155	0.049	0.100	0.700	0.922	0.051	0.049	0.056	0.086	0.040
$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.9$										
$T = 5$										
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
Size	1.000	0.148	0.334	0.094	0.100	0.106	0.148	0.141	0.056	0.060
$T = 10$										
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
Size	0.999	0.147	0.329	0.054	0.250	0.087	0.140	0.120	0.073	0.070
$T = 25$										
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
Size	0.999	0.054	0.290	0.032	0.071	0.057	0.053	0.069	0.045	0.052
$T = 50$										
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
Size	0.983	0.053	0.384	0.009	0.011	0.040	0.055	0.063	0.056	0.042

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), and the two-step Blundell and Bond (1998) GMM estimator (BB-GMM). Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 6: 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
Size	0.995	0.061	0.043	0.176	0.256	0.117	0.061	0.054	0.093	0.062
$T = 10$										
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
Size	0.963	0.055	0.050	0.147	0.410	0.053	0.055	0.049	0.065	0.045
$T = 25$										
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
Size	0.682	0.058	0.065	0.619	0.931	0.060	0.058	0.037	0.086	0.048
$T = 50$										
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
Size	0.405	0.039	0.067	0.980	1.000	0.074	0.039	0.064	0.319	0.347
$\alpha = \alpha_1 + \alpha_2 + \alpha_3 = 0.9$										
$T = 5$										
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
Size	1.000	0.158	0.204	0.160	0.086	0.272	0.158	0.108	0.052	0.053
$T = 10$										
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
Size	1.000	0.183	0.147	0.094	0.125	0.137	0.183	0.053	0.043	0.044
$T = 25$										
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
Size	1.000	0.036	0.148	0.065	0.291	0.049	0.038	0.042	0.046	0.041
$T = 50$										
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
Size	1.000	0.049	0.166	0.190	0.872	0.051	0.047	0.061	0.058	0.057

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), and the two-step Blundell and Bond (1998) GMM estimator (BB-GMM). Reported are the average bias of the estimates, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 7: 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
Size	0.498	0.090	0.048	0.114	0.237	0.090	0.059	0.090	0.053	0.063	0.054	0.069
$T = 10$												
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
Size	0.327	0.050	0.067	0.208	0.612	0.070	0.059	0.050	0.046	0.089	0.061	0.066
$T = 25$												
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
Size	0.182	0.066	0.074	0.335	0.969	0.086	0.067	0.066	0.060	0.081	0.105	0.073
$T = 50$												
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
Size	0.130	0.058	0.079	0.472	0.997	0.074	0.068	0.058	0.055	0.069	0.128	0.063
$\alpha = 0.9$												
$T = 5$												
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
Size	0.942	0.320	0.313	0.119	0.146	0.435	0.110	0.277	0.106	0.071	0.058	0.099
$T = 10$												
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
Size	0.966	0.075	0.302	0.133	0.346	0.163	0.068	0.069	0.080	0.052	0.051	0.147
$T = 50$												
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
Size	0.848	0.079	0.285	0.159	0.463	0.165	0.062	0.075	0.047	0.051	0.056	0.144

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 for a Wald test with nominal size of 5%.

Table 8: 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
Size	0.811	0.201	0.055	0.046	0.064	0.144	0.051	0.201	0.046	0.040	0.045	0.078
$T = 10$												
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
Size	0.749	0.080	0.051	0.072	0.138	0.121	0.077	0.080	0.057	0.047	0.060	0.135
$T = 25$												
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
Size	0.491	0.050	0.061	0.236	0.790	0.114	0.085	0.050	0.048	0.047	0.076	0.085
$T = 50$												
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
Size	0.270	0.047	0.068	0.480	0.993	0.090	0.077	0.047	0.052	0.060	0.079	0.078
$\alpha = 0.9$												
$T = 5$												
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
Size	0.985	0.419	0.192	0.109	0.105	0.591	0.254	0.393	0.108	0.095	0.061	0.186
$T = 10$												
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
Size	0.999	0.348	0.162	0.043	0.098	0.476	0.105	0.328	0.075	0.051	0.061	0.293
$T = 25$												
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
Size	1.000	0.144	0.138	0.076	0.171	0.383	0.055	0.144	0.042	0.046	0.050	0.344
$T = 50$												
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
Size	1.000	0.082	0.119	0.110	0.275	0.333	0.064	0.082	0.056	0.055	0.070	0.312

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, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 9: 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	NA	NA	NA	NA	0.099	0.065	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.078	0.066	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.071	0.054	NA	NA	NA	NA
$\alpha = 0.9$												
$T = 10$												
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	NA	NA	NA	NA	0.153	0.085	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.101	0.048	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.059	0.045	NA	NA	NA	NA

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 for Wald tests with nominal size of 5%. ‘rob-Size’ refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 10: 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	NA	NA	NA	NA	0.131	0.061	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.092	0.043	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.061	0.055	NA	NA	NA	NA
$\alpha = 0.9$												
$T = 10$												
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	NA	NA	NA	NA	0.188	0.083	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.090	0.051	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.073	0.028	NA	NA	NA	NA

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 for Wald tests with nominal size of 5%. ‘rob-Size’ refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 11: 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	NA	NA	NA	NA	0.114	0.080	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.084	0.058	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.068	0.060	NA	NA	NA	NA
$\alpha = 0.9$												
$T = 10$												
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	NA	NA	NA	NA	0.145	0.059	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.091	0.029	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.075	0.034	NA	NA	NA	NA

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 for Wald tests with nominal size of 5%. ‘rob-Size’ refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 12: 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	NA	NA	NA	NA	0.135	0.064	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.097	0.058	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.073	0.051	NA	NA	NA	NA
$\alpha = 0.9$												
$T = 10$												
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	NA	NA	NA	NA	0.180	0.050	NA	NA	NA	NA
$T = 25$												
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	NA	NA	NA	NA	0.081	0.033	NA	NA	NA	NA
$T = 50$												
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	NA	NA	NA	NA	0.074	0.033	NA	NA	NA	NA

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 for Wald tests with nominal size of 5%. ‘rob-Size’ refers to the Wald test employing robust standard errors considered in Theorem 2.

Table 13: 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
Size	0.318	0.060	0.054	0.113	1.000	0.059	0.064	0.060	0.060	0.084	0.091	0.069
$T = 10$												
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
Size	0.247	0.047	0.078	0.212	1.000	0.054	0.068	0.047	0.072	0.099	0.083	0.073
$T = 25$												
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
Size	0.153	0.059	0.081	0.220	1.000	0.059	0.062	0.059	0.054	0.074	0.095	0.067
$T = 50$												
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
Size	0.116	0.052	0.063	0.339	0.999	0.052	0.061	0.052	0.049	0.051	0.164	0.056
$\alpha = 0.9$												
$T = 5$												
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
Size	1.000	0.086	0.212	0.217	0.670	0.237	0.073	0.083	0.099	0.075	0.072	0.079
$T = 10$												
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
Size	0.997	0.054	0.322	0.168	0.958	0.167	0.067	0.054	0.126	0.061	0.063	0.112
$T = 25$												
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
Size	0.976	0.069	0.637	0.044	0.446	0.077	0.060	0.070	0.131	0.044	0.049	0.110
$T = 50$												
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
Size	0.912	0.056	0.497	0.018	0.150	0.072	0.050	0.057	0.063	0.042	0.042	0.093

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, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 14: 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
Size	0.852	0.061	0.041	0.050	1.000	0.067	0.054	0.061	0.061	0.055	0.204	0.053
$T = 10$												
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
Size	0.715	0.054	0.052	0.054	1.000	0.055	0.072	0.054	0.056	0.056	0.055	0.064
$T = 25$												
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
Size	0.444	0.054	0.047	0.147	1.000	0.057	0.074	0.054	0.058	0.064	0.031	0.056
$T = 50$												
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
Size	0.276	0.045	0.050	0.477	1.000	0.049	0.077	0.045	0.049	0.062	0.165	0.054
$\alpha = 0.9$												
$T = 5$												
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
Size	1.000	0.040	0.102	0.213	0.999	0.110	0.066	0.037	0.094	0.068	0.058	0.088
$T = 10$												
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
Size	1.000	0.049	0.136	0.206	1.000	0.070	0.049	0.048	0.092	0.057	0.049	0.255
$T = 25$												
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
Size	1.000	0.042	0.434	0.144	1.000	0.172	0.045	0.043	0.137	0.063	0.059	0.244
$T = 50$												
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
Size	1.000	0.048	0.171	0.085	1.000	0.163	0.061	0.048	0.053	0.050	0.081	0.172

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, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 15: Simulation results: predetermined regressor, $N = 50$

	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	-0.123	-0.047	-0.043	0.025	0.091	-0.048	-0.129	-0.178	-0.090	-0.039	-0.060	-0.163
RMSE	0.129	0.063	0.074	0.110	0.129	0.063	0.160	0.202	0.185	0.215	0.196	0.187
Size	0.902	0.253	0.131	0.118	0.259	0.258	0.262	0.253	0.114	0.076	0.086	0.415
$T = 10$												
Bias	-0.060	-0.025	-0.032	0.075	0.151	-0.024	-0.036	-0.075	-0.045	-0.043	-0.070	-0.071
RMSE	0.064	0.034	0.048	0.110	0.163	0.033	0.074	0.098	0.096	0.138	0.126	0.095
Size	0.777	0.214	0.166	0.185	0.755	0.209	0.074	0.214	0.086	0.039	0.119	0.200
$T = 25$												
Bias	-0.024	-0.011	-0.019	0.074	0.153	-0.011	-0.002	-0.022	-0.011	-0.063	-0.087	-0.023
RMSE	0.027	0.017	0.027	0.087	0.159	0.017	0.039	0.044	0.047	0.106	0.114	0.045
Size	0.447	0.144	0.165	0.323	0.950	0.148	0.050	0.144	0.052	0.046	0.078	0.093
$\alpha = 0.9$												
$T = 5$												
Bias	-0.422	-0.035	-0.418	-0.081	0.007	-0.087	-0.023	-0.196	-0.161	-0.040	-0.012	-0.136
RMSE	0.427	0.129	0.466	0.198	0.087	0.142	0.101	0.227	0.226	0.166	0.135	0.167
Size	1.000	0.131	0.604	0.097	0.081	0.235	0.086	0.134	0.195	0.075	0.058	0.316
$T = 10$												
Bias	-0.217	-0.030	-0.267	-0.023	0.025	-0.037	0.012	-0.088	-0.090	-0.038	-0.010	-0.064
RMSE	0.220	0.072	0.304	0.084	0.044	0.073	0.062	0.114	0.130	0.096	0.075	0.088
Size	1.000	0.151	0.645	0.050	0.126	0.220	0.062	0.153	0.175	0.051	0.043	0.189
$T = 25$												
Bias	-0.083	-0.016	-0.105	-0.008	0.017	-0.017	0.022	-0.026	-0.014	-0.019	-0.004	-0.029
RMSE	0.084	0.025	0.122	0.045	0.034	0.025	0.041	0.045	0.049	0.065	0.058	0.045
Size	0.998	0.138	0.571	0.027	0.032	0.147	0.081	0.142	0.062	0.007	0.011	0.113

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, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.

Table 16: Simulation results: predetermined regressor, $N = 200$

	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	-0.125	-0.049	-0.011	0.006	0.012	-0.049	-0.125	-0.175	-0.023	0.001	-0.006	-0.160
RMSE	0.126	0.053	0.034	0.039	0.040	0.053	0.135	0.182	0.086	0.085	0.086	0.167
Size	1.000	0.684	0.064	0.065	0.074	0.701	0.720	0.684	0.067	0.054	0.063	0.915
$T = 10$												
Bias	-0.061	-0.026	-0.009	0.010	0.024	-0.025	-0.035	-0.073	-0.012	-0.008	-0.013	-0.070
RMSE	0.062	0.029	0.021	0.023	0.034	0.028	0.048	0.080	0.045	0.049	0.048	0.077
Size	1.000	0.625	0.087	0.066	0.172	0.593	0.176	0.625	0.061	0.050	0.047	0.588
$T = 25$												
Bias	-0.023	-0.010	-0.005	0.025	0.055	-0.010	-0.005	-0.024	-0.003	-0.016	-0.014	-0.025
RMSE	0.024	0.012	0.011	0.030	0.058	0.012	0.020	0.031	0.023	0.034	0.031	0.032
Size	0.947	0.333	0.090	0.369	0.919	0.358	0.056	0.333	0.047	0.076	0.086	0.243
$\alpha = 0.9$												
$T = 5$												
Bias	-0.419	-0.013	-0.232	-0.041	-0.006	-0.093	-0.022	-0.201	-0.091	-0.009	0.003	-0.137
RMSE	0.420	0.086	0.283	0.107	0.062	0.105	0.051	0.210	0.132	0.078	0.068	0.144
Size	1.000	0.106	0.354	0.174	0.067	0.465	0.075	0.163	0.192	0.061	0.062	0.832
$T = 10$												
Bias	-0.215	-0.033	-0.107	-0.021	0.004	-0.044	0.012	-0.085	-0.037	-0.007	-0.002	-0.064
RMSE	0.216	0.048	0.130	0.052	0.031	0.051	0.033	0.092	0.061	0.045	0.042	0.071
Size	1.000	0.259	0.328	0.110	0.077	0.427	0.067	0.266	0.123	0.058	0.056	0.572
$T = 25$												
Bias	-0.081	-0.014	-0.030	0.000	0.015	-0.016	0.019	-0.028	-0.004	-0.004	-0.003	-0.031
RMSE	0.082	0.017	0.039	0.023	0.021	0.018	0.027	0.034	0.024	0.025	0.023	0.036
Size	1.000	0.300	0.230	0.053	0.207	0.361	0.176	0.301	0.053	0.046	0.046	0.391

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, the root mean square error (RMSE), and the empirical size for a Wald test with nominal size of 5%.