Federal Reserve Bank of Dallas
Globalization and Monetary Policy Institute
Working Paper No. 327
https://doi.org/10.24149/gwp327

A Bias-Corrected Method of Moments Approach to Estimation of Dynamic Short-T Panels*

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
USC Dornsife INET, University of Southern California,
and Trinity College, Cambridge, UK

September 2017

Abstract

This paper contributes to the GMM literature by introducing the idea of self-instrumenting target variables instead of searching for instruments that are uncorrelated with the errors, in cases where the correlation between the target variables and the errors can be derived. The advantage of the proposed approach lies in the fact that, by construction, the instruments have maximum correlation with the target variables and the problem of weak instrument is thus avoided. The proposed approach can be applied to estimation of a variety of models such as spatial and dynamic panel data models. In this paper we focus on the latter and consider both univariate and multivariate panel data models with short time dimension. Simple Bias-corrected Methods of Moments (BMM) estimators are proposed and shown to be consistent and asymptotically normal, under very general conditions on the initialization of the processes, individual-specific effects, and error variances allowing for heteroscedasticity over time as well as cross-sectionally. Monte Carlo evidence document BMM's good small sample performance across different experimental designs and sample sizes, including in the case of experiments where the system GMM estimators are inconsistent. We also find that the proposed estimator does not suffer size distortions and has satisfactory power performance as compared to other estimators.

JEL codes: C12, C13, C23

* Alexander Chudik, Federal Reserve Bank of Dallas, Research Department, 2200 N. Pearl Street, Dallas, TX 75201. 214-922-5769. alexander.chudik@dal.frb.org. M. Hashem Pesaran, Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253. pesaran@usc.edu. We would like to thank Seung Ahn, Maurice Bun, Geert Dhaene, Brian Finley, Everett Grant, Kazuhiko Hayakawa, Cheng Hsiao, Vasilis Sarafidis, Vanessa Smith, Ron Smith, Martin Weidner, conference participants at the June 2016 and 2017 IAAE annual conferences, and participants at the Federal Reserve Bank of Dallas research department seminar for helpful comments on earlier versions of this paper. The views in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

Analysis of linear dynamic panel data models where the time dimension (T) is small relative to the cross section dimension (n), plays an important role in applied microeconomic research. The estimation of such panels is carried out predominantly by the application of the Generalized Method of Moments (GMM) after first-differencing.[1] This approach utilizes instruments that are uncorrelated with the errors but are potentially correlated with the target variables (the included regressors). A number of well-known GMM estimation methods have been proposed, including Anderson and Hsiao (1981 and 1982), Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), and Hayakawa (2012), among others. Unlike the likelihood-based methods in the literature (Hsiao et al., 2002, and Hayakawa and Pesaran, 2015), the GMM methods apply to autoregressive (AR) panels as well as to AR panels augmented with strictly or weakly exogenous regressors. However, the GMM approach is subject to a number of drawbacks. Specifically, the first-difference GMM methods by Arellano and Bond (1991) can suffer from the weak instrument problem when the dependent variable is close to being a unit root process or when the variance of individual effects is relatively large. To overcome this problem the system GMM approach by Blundell and Bond (1998) utilizes additional moment conditions, but these additional conditions are valid only under strong requirements on the initialization of the dynamic processes. In particular, as shown in Section 2, the system GMM approach does not allow for initial observations to differ from the long-run means in a systematic manner.

This paper contributes to the GMM literature by introducing the idea of self-instrumenting target variables instead of searching for instruments that are uncorrelated with the errors, in cases where the correlation between the target variables and the errors can be derived. This idea has wide-ranging applications for robust estimation and inference in a number settings, including dynamic short-T panels. It differs from the wide variety of the bias-corrected estimation methods in the literature, which correct a first-stage estimator for small-T bias (see, for example, methods based on exact analytical bias formula or its approximation, Bruno, 2005, Bun, 2003, Bun and Carree, 2005 and 2006, Bun and Kiviet, 2003, Hahn and Kuersteiner, 2002, Hahn and Moon, 2006, Juodis, 2013, and Kiviet, 1995 and 1999; simulation-based bias-correction methods by Everaert and Ponzi, 2007,

[1]Other approaches in the literature include the likelihood-based methods (Hsiao et al., 2002, and Hayakawa and Pesaran, 2015), X-differencing method (Han et al., 2014), factor-analytical method (Bai, 2013), and bias-correction methods mentioned below.

and Phillips and Sul, 2003 and 2007; the jackknife bias corrections by Dhaene and Jochmans, 2015, and Chudik, Pesaran, and Yang, 2016; or the recursive mean adjustment correction procedures, Choi et al., 2010).[2] In contrast to the bias-correction methods, our approach is not based on correcting for a bias of an estimator, but instead it is based on correcting the 'bias' of the moment conditions *before* estimation. One could also consider the application of bias-correction techniques to BMM estimators, but our Monte Carlo results show that such post estimation bias-corrections are not required.

The advantage of the proposed approach lies in the fact that, by construction, the instruments have maximum correlation with the target variables and the problem of weak instrument is thus avoided. The proposed approach can be applied to estimation of a variety of models, where the underlying model is sufficiently specified so that the correlation between the instruments and errors can be derived, such as spatial and dynamic panel data models. In this paper we focus on the latter and consider both univariate and multivariate panel data models with short time dimension.

Simple Bias-corrected Methods of Moments (BMM) estimators are proposed and shown to be consistent and asymptotically normal, under very general conditions on the initialization of the processes, individual-specific effects, with (possibly) heteroscedastic error variances over time as well as cross-sectionally. We refer to the proposed estimators as BMM to distinguish them from traditional GMM estimators, which are based on moment conditions derived from instruments that are orthogonal to the errors. Monte Carlo experiments document BMM's good small sample performance in comparison with a number of GMM alternatives. The inference based on the BMM estimator appears more reliable compared with any of the GMM alternatives considered. In addition, the BMM estimator is valid also in designs where the stricter requirements of the system-GMM approach are not satisfied, albeit it is less efficient asymptotically in designs where such requirements hold. However, in practice it is not known whether conditions regarding the initialization of dynamic processes are satisfied, and it seems desirable to consider estimation procedures that are robust to violation of such restrictive assumptions.

The remainder of this paper is organized as follows. Section 2 sets up the baseline panel AR(1) model. Section 3 presents the main idea, proposes a BMM estimator of AR(1) panels, and establishes

[2] Most of these bias-correction techniques do not apply to short-T type panels where the error variances are heteroskedastic (over i and t), with the exception of Juodis (2013), and possibly the simulation-based bias-correction method of Everaert and Ponzi (2007). A comparative analysis of BMM and bias correction estimators is a welcome addition to the literature but lies beyond the scope of the present paper.

its consistency and asymptotic normality when T is fixed and $n \to \infty$. Section 4 briefly discusses identification of AR(1) coefficient under alternative GMM estimators in the literature. Section 5 extends the BMM estimator to panel VAR(1) models, and to panel data models with higher order lags. Section 6 presents Monte Carlo (MC) evidence, and the last section concludes and discusses avenues for future research. Some of the mathematical proofs are provided in an appendix. Further theoretical and Monte Carlo results are presented in an online supplement.

Notations: Generic positive finite small and large constants that do not depend on the cross section dimension are denoted by c and K, respectively. All vectors are column vectors denoted by bold lowercase letters. Matrices are denoted by bold uppercase letters. Let $\mathbf{A}$ be a $p \times q$ matrix, then $\|\mathbf{A}\| = [tr(\mathbf{A}'\mathbf{A})]^{1/2}$ is the Frobenius norm of matrix $\mathbf{A}$, $Vec(\mathbf{A})$ is a $pq \times 1$ vector formed from stacking the q columns of $\mathbf{A}$. $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively, and $\overset{a}{\sim}$ denotes asymptotic equivalence in distribution for a fixed T, and as $n \to \infty$.

2 Panel AR(1) model and assumptions

We begin with a simple panel AR(1) model to set out the main idea behind the BMM estimator. Specifically, consider the following dynamic panel data model

$$y_{it} = \alpha_i + \phi y_{i,t-1} + u_{it}, \text{ for } i = 1, 2, ..., n, \tag{1}$$

where $\{\alpha_i, 1 \leq i \leq n\}$ are unobserved unit-specific effects, u_{it} is the idiosyncratic error term, and y_{it} are generated from the initial values, $y_{i,-m_i}$ for $m_i \geq 0$, and $t = -m_i + 1, -m_i + 2..., 1, 2, ..., T$. Using (1) to solve for the initial observations y_{i0}, we obtain

$$y_{i0} = \phi^{m_i} y_{i,-m_i} + \alpha_i \left(\frac{1 - \phi^{m_i}}{1 - \phi} \right) + \sum_{\ell=0}^{m_i - 1} \phi^{\ell} u_{i,-\ell}. \tag{2}$$

It is assumed that available observations for estimation and inference are y_{it}, for $i = 1, 2, ..., n$, and $t = 0, 1, 2, ..., T$. For the implementation of the proposed estimator we require $T \geq 3$, although under mean and variance stationarity identification of ϕ could be achieved even if $T = 2$, namely if the panel covers three time periods.

ASSUMPTION 1 *(Parameter of interest) The true value of ϕ, denoted by ϕ_0, is the parameter* *of interest, and it is assumed that* $\phi \in \Theta$*, where* $\Theta \subset (-1, 1]$ *is a compact set.*[3]

In the case where $|\phi| < 1$, and $m_i \to \infty$, then $E(y_{it}) = E(\alpha_i) / (1 - \phi)$ for all t. We set $\mu_i = \alpha_i / (1 - \phi)$ and refer to μ_i as the long-run mean of y_{it}, even if m_i is finite. However in the unit-root case ($\phi = 1$), μ_i is not defined and to avoid incidental linear trends we set $\alpha_i = 0$.

Taking first differences of (1), we obtain

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \Delta u_{it}, \tag{3}$$

for $t = 2, 3, ..., T$, and $i = 1, 2, ..., n$; but Δy_{i1} is given by

$$\Delta y_{i1} = b_i - (1 - \phi) \sum_{\ell=0}^{m_i - 1} \phi^{\ell} u_{i,-\ell} + u_{i1}, \tag{4}$$

where

$$b_i = -\phi^{m_i} \left[(1 - \phi) y_{i,-m_i} - \alpha_i \right] = -\phi^{m_i} (1 - \phi) (y_{i,-m_i} - \mu_i). \tag{5}$$

The relations (4) and (5) show how the deviations of starting values from the long-run means, given by $(y_{i,-m_i} - \mu_i)$, affect Δy_{i1}.

The contribution of the first term in (4) to Δy_{i1} is given by b_i, and consequently it is clear that the initialization of the process will be unimportant for $|\phi| < 1$, $E|y_{i,-m_i} - \mu_i| < K$, and m_i large. We aim for a minimal set of assumptions on the starting values and individual effects, since in practice such assumptions are difficult to ascertain, and they could have important consequences for estimation and inference when m_i and T are both small.

We consider the following assumptions on the errors, u_{it}, and the starting values, $y_{i,-m_i}$.

ASSUMPTION 2 *(Idiosyncratic errors) For each* $i = 1, 2, ..., n$*, the process* $\{u_{it},\ t = -m_i + 1, -m_i + 2,\ ...,\ 1, 2, ..., T\}$ *is distributed with mean* 0*,* $E\left(u_{it}^2\right) = \sigma_{it}^2$*, and there exist positive constants* c *and* K *such that* $0 < c < \sigma_{it}^2 < K$*. Moreover,* $\bar{\sigma}_{tn}^2 \equiv n^{-1} \sum_{i=1}^{n} \sigma_{it}^2 \to \bar{\sigma}_t^2$ *as* $n \to \infty$*, and* $\sup_{it} E|u_{it}|^{4+\epsilon} < K$ *for some* $\epsilon > 0$*. For each* t*,* u_{it} *is independently distributed over* i*. For each* i*,* u_{it} *is serially uncorrelated over* t*.*

ASSUMPTION 3 *(Initialization and individual effects) Let* $b_i \equiv -\phi^{m_i} \left[(1 - \phi) y_{i,-m_i} - \alpha_i \right]$*. It*

[3] Our theory applies for all finite values of ϕ so long as T and m_i are fixed as $n \to \infty$. We focus on $-1 < \phi \leq 1$, since we believe these values are most relevant in empirical applications.

is assumed that $\bar{\varsigma}_n^2 \equiv n^{-1}\sum_{i=1}^n \varsigma_i^2 \to \bar{\varsigma}^2$ *as* $n \to \infty$*, where* $\varsigma_i^2 = E\left(b_i^2\right)$*, for* $i = 1, 2, ..., n$*, and* $\sup_i E\left|b_i\right|^{4+\epsilon} < K$ *for some* $\epsilon > 0$*. In addition,* b_i *is independently distributed of* $(b_j, u_{jt})'$ *for all* $i \neq j$*,* $i, j = 1, 2, ..., n$*, and all* $t = -m_j + 1, -m_j + 2, ..., 1, 2, ..., T$*, and the following conditions hold:*

$$E\left(\Delta u_{it} b_i\right) = 0, \text{ for } i = 1, 2, ..., n, \text{ and } t = 2, 3, ..., T. \tag{6}$$

Remark 1 *Assumption 2 does not allow the errors,* u_{it}*, to be cross-sectionally dependent, as is customary in the GMM short-T panel data literature, and together with Assumption 3 ensures also that* Δy_{it} *is cross-sectionally independent. When errors are weakly cross-sectionally correlated, in the sense defined in Chudik, Pesaran, and Tosetti (2011), then the BMM estimators proposed in this paper remain consistent, but the inference based on them will no longer be valid. See Section S.1 in the online supplement for further discussion.*

Remark 2 *Assumption 2 allows errors to be unconditionally heteroskedastic across both* i *and* t*.*

Remark 3 *Assumption 3 allows for* $E(b_i)$ *to vary across* i*, and therefore, in view of (3)-(4),* $E(\Delta y_{it})$ *can vary across both* i *and* t*.*

2.1 Assumptions underlying GMM estimators

It is important to compare our assumptions on the individual effects and the starting values with those maintained in the GMM literature. Under Assumptions 2 and 3, initial first-differences, Δy_{i1}, given by (4) have fourth-order moments and the following moment conditions, which are key to our estimation method, hold

$$E\left(\Delta y_{is} \Delta u_{it}\right) = 0, \text{ for } i = 1, 2, ..., n,\ s = 1, 2, ..., t-2, \text{ and } t = 3, 4, ..., T. \tag{7}$$

The same moment conditions are also utilized by Anderson and Hsiao (1981, 1982). However, the subsequent GMM estimators advanced by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) require stronger conditions on the initial values and the individual effects as compared to (7). The first-difference GMM approach considered by Arellano and Bond (1991) assumes

$$E\left(y_{is} \Delta u_{it}\right) = 0, \text{ for } i = 1, 2, ..., n, s = 0, 1, 2, ..., t-2, \text{ and } t = 2, 3, ..., T, \tag{8}$$

which imply (7) but are not required for the moment conditions in (7) to hold. It is clear that estimator based on (8) will depend on the distributional assumptions regarding the individual effects, whereas an estimator based on (7) need not depend on the distributional assumptions regarding the individual effects.[4]

In addition to (8), the system GMM approach considered by Arellano and Bover (1995) and Blundell and Bond (1998) also requires that[5]

$$E\left[\Delta y_{i,t-1}\left(\alpha_i + u_{it}\right)\right] = 0, \text{ for } i = 1,2,...,n; \text{ and } t = 2,3,...,T. \tag{9}$$

These additional restrictions impose further requirements on the errors and the initial values. To see this, first note that iterating (3) from $t = 1$ and using (4) we have

$$\Delta y_{it} = \phi^{t-1}\left[b_i + u_{i1} - (1-\phi)\sum_{\ell=0}^{m_i-1}\phi^{\ell}u_{i,-\ell}\right] + \sum_{\ell=0}^{t-2}\phi^{\ell}\Delta u_{i,t-\ell}. \tag{10}$$

Since for all i, u_{it}'s are assumed to be serially uncorrelated, then condition (9) is met if

$$\phi^{t-2}E\left[b_i\left(\alpha_i + u_{it}\right)\right] + \phi^{t-2}E\left(u_{i1}\alpha_i\right) + (\phi-1)\phi^{t-2}\sum_{\ell=0}^{m_i-1}\phi^{\ell}E\left(\alpha_i u_{i,-\ell}\right) + \sum_{\ell=0}^{t-3}\phi^{\ell}E\left(\alpha_i\Delta u_{i,t-\ell-1}\right) = 0,$$

for $i = 1,2,...,n$; and $t = 2,3,...,T$. In the case where $m_i \to \infty$, the first term vanishes and the moment conditions (9) will be satisfied if $E\left(u_{it}\alpha_i\right) = 0$, for all i and $t \leq T-1$. If m_i is finite it is further required that $E\left[b_i\left(\alpha_i + u_{it}\right)\right] = 0$, unless $\phi = 0$. Now using (5) and noting that $|\phi| < 1$, we have[6]

$$\begin{aligned} E\left[b_i\left(\alpha_i + u_{it}\right)\right] &= -\phi^{m_i}(1-\phi)E\left[\left(y_{i,-m_i} - \mu_i\right)\left(\alpha_i + u_{it}\right)\right] \\ &= -\phi^{m_i}(1-\phi)E\left[\left(y_{i,-m_i} - \mu_i\right)\alpha_i\right]. \end{aligned}$$

[4]Suppose that $|\phi| < 1$, and consider the case where m_i is finite, namely, $0 \leq m_i < K$, and consider the following initial values $y_{i,-m_i} = \mu_i + v_i$, where $E(v_i) = 0$, and $E(v_i\Delta u_{it}) = 0$, for $i = 1,2,...,n$, and $t = 3,4,...,T$. v_i measures the extent to which the initial values $y_{i,-m_i}$ deviate from the long-run means, μ_i. Under this specification of initial values, Δy_{it}, for $t = 0,1,...,T$ and all i does not depend on μ_i, and estimator based on (7) will not depend on the distributional assumptions about μ_i.

[5]The complete set of moment conditions is $E\left[\Delta y_{is}\left(\alpha_i + u_{it}\right)\right] = 0$, for $i = 1,2,...,n$, $s = 1,2,...,t-1$, and $t = 2,3,...,T$. The set of conditions in (9) contains the $T-2$ moment conditions in the system GMM approach that are not redundant.

[6]Note that by assumption $E\left(u_{it}\alpha_i\right) = 0 = E\left(u_{it}y_{i,-m_i}\right)$, for $t = 2,3,....$

Therefore, when m_i is finite for the validity of the moment conditions (9) it is also required that

$$E\left[\mu_i\left(y_{i,-m_i}-\mu_i\right)\right]=0, \text{ for } i=1,2,...,n. \tag{11}$$

This condition requires that for each i, individual effects are uncorrelated with the deviations of initial values from their equilibrium values (long-run means μ_i). These restrictions might not hold in practice. For example, condition (11) is invalidated if some processes start from zero ($y_{i,-m_i}=0$), but the individual effects differ from zero ($\mu_i \neq 0$).

It is true that by imposing additional conditions on individual effects and starting values it might be possible to obtain more efficient estimator of ϕ. However, it is also desirable to seek estimators of ϕ that are consistent under reasonably robust set of assumptions on starting values, individual effects, and error variances. Seen from this perspective, Assumption 3 is more general than the moment conditions assumed in the existing GMM literature.

When comparing GMM and BMM estimators, it is also worth noting from (10) that if $|\phi|<1$ and $\{y_{it}\}$ are initialized in a distant past (with $m_i \to \infty$), then Δy_{it} will no longer depend on α_i and renders the BMM and Anderson-Hsiao IV estimators invariant to the individual effects. However, this is not the case for the GMM estimators that make use of lagged values of y_{it} in construction of their moment conditions. As a result, the performance of such GMM estimators can be affected by the size of $Var\left(\alpha_i\right)$ relative to the other parameters of the model, in particular $Var\left(u_{it}\right)$, see Blundell and Bond (1998) and Binder et al. (2005) for further discussions.

3 BMM estimation of short-T AR(1) panels

Following the GMM approach we consider the first-differenced version of the panel AR model (3), but instead of using (valid) instruments for $\Delta y_{i,t-1}$ that are uncorrelated with the error terms, Δu_{it}, we propose a self-instrumenting procedure whereby $\Delta y_{i,t-1}$ is 'instrumented' for itself, but the population bias due to the non-zero correlation between $\Delta y_{i,t-1}$ and Δu_{it} is corrected accordingly. The advantage of using $\Delta y_{i,t-1}$ as an instrument lies in the fact that by construction it has maximum correlation with the target variable (itself), so long as we are able to correct for the bias that arises due to $Cov\left(\Delta y_{i,t-1}, \Delta u_{it}\right) \neq 0$. To summarize, GMM searches for instruments that are uncorrelated with the errors but are sufficiently correlated with the target variables. Instead, we propose using the target variables as instruments but correct the moment conditions for the non-zero correlations between the errors and the instruments. Both approaches employ method of moments, but differ in the way the moments are constructed.

Using $\Delta y_{i,t-1}$ as an instrument, we obtain under Assumptions 2 and 3,

$$E\left(\Delta u_{it}\Delta y_{i,t-1}\right)=-\sigma_{i,t-1}^2,\text{ for } i=1,2,...,n,\text{ and } t=2,3,...,T-1. \tag{12}$$

To solve for σ_{it}^2, we note that $E\left(\Delta u_{it}\right)^2=\sigma_{i,t-1}^2+\sigma_{it}^2$ and $E\left(\Delta u_{i,t+1}\Delta y_{it}\right)=-\sigma_{it}^2$. Hence, $\sigma_{i,t-1}^2=E\left(\Delta u_{it}\right)^2+E\left(\Delta u_{i,t+1}\Delta y_{it}\right)$, and we obtain the following quadratic moment (QM) condition,

$$E\left(\Delta u_{it}\Delta y_{i,t-1}\right)+E\left(\Delta u_{it}\right)^2+E\left(\Delta u_{i,t+1}\Delta y_{it}\right)=0, \tag{13}$$

for $i=1,2,...,n$, and $t=2,3,...,T-1$. It is useful to note that the solution $\sigma_{i,t-1}^2=E\left(\Delta u_{it}\right)^2+E\left(\Delta u_{i,t+1}\Delta y_{it}\right)$ depends on the set of assumptions considered, and different solutions could be obtained under different (stricter) conditions. In this paper, we focus on the general set of conditions summarized by Assumptions 2 and 3, although other conditions can be obtained if one is prepared to make stronger assumptions such as $\sigma_{it}^2=\sigma_{i,t-1}^2=\sigma_i^2$. Another possibility is to assume covariance stationarity of y_{it}, which will lead to a linear moment condition solution, discussed in Remark 5 below.[7]

We use the QM condition (13) alone to obtain an estimator of ϕ. We propose averaging (13) over i and t, which will deliver an exactly identified moment estimator. It is clearly possible to use other weights, as done in the GMM literature, to combine individual moment conditions in (13). But to keep the analysis simple and to focus on the main contribution of the paper, we shall not consider optimally weighting the moment conditions in (13), or augmenting them with Anderson-Hsiao type moment conditions.

Averaging moment condition (13) over t, and substituting (3) for Δu_{it} and $\Delta u_{i,t+1}$, we obtain

$$E\left[M_{iT}\left(\phi\right)\right]=0,\text{ for } i=1,2,...,n, \tag{14}$$

[7] Covariance stationarity requires strong restrictions on the initialization of the dynamic processes, in addition to time-invariant error variances.

where

$$M_{iT}(\phi) = \frac{1}{T-2}\sum_{t=2}^{T-1}\Big[\left(\Delta y_{it} - \phi\Delta y_{i,t-1}\right)\Delta y_{i,t-1} + \left(\Delta y_{it} - \phi\Delta y_{i,t-1}\right)^2 + \left(\Delta y_{i,t+1} - \phi\Delta y_{it}\right)\Delta y_{it}\Big]. \tag{15}$$

The BMM estimator is then given by

$$\hat{\phi}_{nT} = \arg\min_{\phi\in\Theta}\left\|\bar{M}_{nT}(\phi)\right\|, \tag{16}$$

where $\|.\|$ denotes the Euclidean norm, $\Theta \subset (-1, 1]$ is a compact set for the admissible values of ϕ defined by Assumption 1, and

$$\bar{M}_{nT}(\phi) = \frac{1}{n}\sum_{i=1}^{n} M_{iT}(\phi). \tag{17}$$

To derive the asymptotic properties of $\tilde{\phi}_{nT}$, let ϕ_0 denote the true value of ϕ, assumed to lie inside Θ, and note that under $\phi = \phi_0$, (3) yields $\Delta y_{it} = \phi_0\Delta y_{i,t-1} + \Delta u_{it}$, and (15) can be written as

$$\begin{aligned} M_{iT}(\phi) &= \frac{1}{T-2}\sum_{t=2}^{T-1}\left\{\begin{array}{l} \left[\Delta u_{it} - (\phi-\phi_0)\Delta y_{i,t-1}\right]\Delta y_{i,t-1} \\ +\left[\Delta u_{it} - (\phi-\phi_0)\Delta y_{i,t-1}\right]^2 \\ +\left[\Delta u_{i,t+1} - (\phi-\phi_0)\Delta y_{it}\right]\Delta y_{it} \end{array}\right\} \\ &= \Lambda_{iT} + V_{iT}, \end{aligned} \tag{18}$$

where

$$V_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right), \tag{19}$$

and

$$\Lambda_{iT} = (\phi-\phi_0)^2 Q_{iT} - (\phi-\phi_0)\left(Q_{iT} + Q_{iT}^{+} + 2H_{iT}\right), \tag{20}$$

in which

$$Q_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta y_{i,t-1}^2,\ Q_{iT}^{+} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta y_{it}^2, \text{ and } H_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta u_{it}\Delta y_{i,t-1}. \tag{21}$$

We have one unknown parameter ϕ and one moment condition (14). Suppose there exists $\hat{\phi}_{nT}$ such that $\bar{M}_{nT}\left(\hat{\phi}_{nT}\right)=0$. Then (18) evaluated at $\phi=\hat{\phi}_{nT}$ yields

$$\left(\hat{\phi}_{nT}-\phi_0\right)\left[\left(\hat{\phi}_{nT}-\phi_0\right)\bar{Q}_{nT}-\bar{B}_{nT}\right]=-\bar{V}_{nT}, \tag{22}$$

where

$$\bar{V}_{nT}=\frac{1}{n}\sum_{i=1}^{n}V_{iT}. \tag{23}$$

$$\bar{Q}_{nT}=\frac{1}{n}\sum_{i=1}^{n}Q_{iT}, \tag{24}$$

and

$$\bar{B}_{nT}=\frac{1}{n}\sum_{i=1}^{n}\left(Q_{iT}+Q_{iT}^{+}+2H_{iT}\right). \tag{25}$$

Using results (A.5)-(A.6) of Lemma A.1 in the appendix, under Assumptions 1-3, we have (for a fixed T)

$$\bar{Q}_{nT}=E\left(\bar{Q}_{nT}\right)+O_p\left(n^{-1/2}\right), \text{ and } \bar{B}_{nT}=E\left(\bar{B}_{nT}\right)+O_p\left(n^{-1/2}\right), \tag{26}$$

where

$$E\left(\bar{Q}_{nT}\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(Q_{iT}\right)>0. \tag{27}$$

In addition, using result (A.7) of Lemma A.2 in the appendix, we have

$$\bar{V}_{nT}=O_p\left(n^{-1/2}\right). \tag{28}$$

We now use (22) to show that there exists a unique $\sqrt{n}$-consistent estimator of ϕ. Suppose that $\hat{\phi}_{nT}$ is a $\sqrt{n}$-consistent estimator of ϕ. Then we establish that such an estimator is in fact unique. Using (22), we have

$$\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)^2\bar{Q}_{nT}-\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)\bar{B}_{nT}=-\sqrt{n}\bar{V}_{nT}. \tag{29}$$

But, if there exists a $\sqrt{n}$-consistent estimator, then $\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)^2\bar{Q}_{nT}=O_p\left(n^{-1/2}\right)$, and hence

$$\bar{B}_{nT}\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)=-\sqrt{n}\bar{V}_{nT}+O_p\left(n^{-1/2}\right). \tag{30}$$

Also, using (26) the above can be written as

$$E\left(\bar{B}_{nT}\right)\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)=-\sqrt{n}\bar{V}_{nT}+O_p\left(n^{-1/2}\right).$$

where by (28), $\sqrt{n}\bar{V}_{nT}=O_p(1)$. If

$$\bar{B}_T=\lim_{n\to\infty}\frac{1}{n}E\left(\bar{B}_{nT}\right)\neq 0, \tag{31}$$

it then follows that the $\sqrt{n}$-consistent estimator, $\hat{\phi}_{nT}$, must be unique. It also follows that

$$\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)\overset{a}{\sim}\bar{B}_T^{-1}\sqrt{n}\bar{V}_{nT}.$$

Finally, using result (A.8) of Lemma A.2 in the appendix, we have $\sqrt{n}\bar{V}_{nT}\to_d N(0,S_T)$, where $S_T=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}E\left(V_{iT}^2\right)$, and it follows that $\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right)\to_d N(0,\Sigma_T)$ with $\Sigma_T=\bar{B}_T^{-2}S_T$.

The key condition for the existence of a $\sqrt{n}$-consistent estimator of ϕ is given by $\bar{B}_T\neq 0$, which can be written more fully as

$$\bar{B}_T=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(Q_{iT}+Q_{iT}^{+}+2H_{iT}\right),$$

where

$$\begin{aligned}Q_{iT} &= (T-2)^{-1}\sum_{t=2}^{T-1}\Delta y_{i,t-1}^2,\ Q_{iT}^{+}=(T-2)^{-1}\sum_{t=2}^{T-1}\Delta y_{it}^2\\ H_{iT} &= (T-2)^{-1}\sum_{t=2}^{T-1}\Delta u_{it}\Delta y_{i,t-1}.\end{aligned}$$

It is easily seen that condition $\bar{B}_T\neq 0$ is satisfied when Δy_{it} is a stationary process (for $m_i\to\infty$, $\sigma_{it}=\sigma_i^2$ and $|\phi|<1$). In this case we have

$$\bar{B}_T=2\left(\frac{1-\phi}{1+\phi}\right)\bar{\sigma}_n^2>0,$$

where $\bar{\sigma}_n^2=n^{-1}\sum_{i=1}^{n}\sigma_i^2$. In the non-stationary case (with m finite) $\bar{B}_T\neq 0$ even if $\phi=1$ so long as σ_{it} is sufficiently variable over the observed sample.

The following theorem summarizes the main results established above.

Theorem 1 *Suppose y_{it}, for $i = 1, 2, ..., n$, and $t = -m+1, -m+2, ..., 1, 2, ..., T$, are generated by (1) with starting values $y_{i,-m}$, and the true value of the parameter of interest ϕ_0. Let Assumptions 1-3 hold, and suppose $\bar{B}_T \neq 0$ and $n^{-1}\sum_{i=1}^{n} E\left(V_{iT}^2\right) \rightarrow S_T > 0$, where $\bar{B}_T$ is given by (31) and V_{iT} is defined in (19). Consider the BMM estimator $\hat{\phi}_{nT}$ given by (16). Let T be fixed and $n \rightarrow \infty$. Then, the unique $\sqrt{n}$-consistent estimator $\hat{\phi}_{nT}$ satisfies*

$$\sqrt{n}\left(\hat{\phi}_{nT} - \phi_0\right) \rightarrow_d N\left(0, \Sigma_T\right),$$

where

$$\Sigma_T = \bar{B}_T^{-2} S_T. \tag{32}$$

Remark 4 *When $\bar{B}_T = 0$, from (22) we have,*

$$\left(\hat{\phi}_{nT} - \phi_0\right)^2 \bar{Q}_{nT} = \bar{V}_{nT} + \left(\hat{\phi}_{nT} - \phi_0\right) O_p\left(n^{-1/2}\right), \tag{33}$$

and, given that $\bar{Q}_{nT} \rightarrow \bar{Q}_T > 0$ as $n \rightarrow \infty$, there exists a unique $n^{1/4}$-consistent estimator $\hat{\phi}_{nT}$. As noted earlier a leading case when $\bar{B}_T = 0$, is the unit root case ($\phi = 1$) under error variance homogeneity over t.

It is illustrative to consider $\bar{B}_T$ for $T = 3$. In the appendix, we derive under Assumptions 2 and 3,

$$\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2 + (1-\phi)^2 \bar{\sigma}_1^2 + \left(1+\phi^2\right)(1-\phi)\psi_0. \tag{34}$$

where

$$\psi_0 = (1-\phi)\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n} E\left(y_{i0} - \mu_i\right)^2 - 2\lim_{n\rightarrow\infty}\frac{1}{n}\sum_{i=1}^{n} E\left[u_{i1}\left(y_{i0} - \mu_i\right)\right]. \tag{35}$$

If $\phi = 1$, then $\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2 \neq 0$ if $\bar{\sigma}_1^2 \neq \bar{\sigma}_2^2$. In general, $\bar{B}_3 \neq 0$ if $\bar{\sigma}_1^2 \neq \bar{\sigma}_2^2$, for all values of $|\phi| \leq 1$, unless $|\phi| < 1$, and

$$(1-\phi)\left(1+\phi^2\right)\psi_0 = \phi(2-\phi)\bar{\sigma}_1^2 - \bar{\sigma}_2^2.$$

Therefore, time variations in the average error variances, $\bar{\sigma}_t^2$, can help identification under the BMM quadratic moment condition, particularly if ϕ is close to unity. Identification conditions for the GMM estimators are discussed in Section 4 below, with exact conditions in the case of $T = 3$ derived in Section S.3 of the online supplement. As can be seen from these results the GMM estimators do not benefit from time variations in $\bar{\sigma}_t^2$.

The variance term in (32), Σ_T, can be estimated consistently by

$$\hat{\Sigma}_{nT} = \hat{\overline{B}}_{nT}^{-2} \left(\frac{1}{n} \sum_{i=1}^{n} \hat{V}_{i,nT}^2 \right), \tag{36}$$

where

$$\hat{\overline{B}}_{nT} = \frac{1}{n} \sum_{i=1}^{n} \left(Q_{iT} + Q_{iT}^{+} + 2\hat{H}_{i,nT} \right), \tag{37}$$

$\hat{H}_{i,nT} = (T-2)^{-1} \sum_{t=2}^{T-1} \Delta\hat{u}_{it} \Delta y_{i,t-1}$, $\Delta\hat{u}_{it} = \Delta y_{it} - \hat{\phi}_{nT} \Delta y_{i,t-1}$, ($\Delta\hat{u}_{it}$ depends on n and T, but we omit subscripts n, T to simplify the notations), and

$$\hat{V}_{i,nT} = -\frac{1}{T-2} \sum_{t=2}^{T-1} \left(\Delta\hat{u}_{it} \Delta y_{i,t-1} + \Delta\hat{u}_{it}^2 + \Delta\hat{u}_{i,t+1} \Delta y_{it} \right). \tag{38}$$

Consistency of $\hat{\Sigma}_{nT}$ is established in Proposition 1 in the appendix.

Remark 5 *In the case of covariance stationary panels ($\phi < 1$ and $m_i \to \infty$), we have $\Delta y_{it} = \sum_{\ell=0}^{\infty} \phi^{\ell} \Delta u_{i,t-\ell}$, where $E\left(u_{it}^2\right) = \sigma_i^2$ and therefore $E\left(\Delta y_{it}^2\right) = 2\sigma_i^2 / (1+\phi)$ is time-invariant. Under these restrictions we have $\sigma_i^2 = (1+\phi) E\left(\Delta y_{i,t-1}^2\right)/2$, $E\left(\Delta u_{it} \Delta y_{i,t-1}\right) = E\left(\Delta u_{i,t+1} \Delta y_{it}\right)$, and using (12) the quadratic moment condition, (13), simplifies to the following linear moment condition:*

$$E\left(\Delta y_{it} \Delta y_{i,t-1}\right) + \frac{1}{2}(1-\phi) E\left(\Delta y_{i,t-1}^2\right) = 0,$$

which yields the associated BMM estimator

$$\hat{\phi}_n = \frac{\sum_{i=1}^{n} \sum_{t=2}^{T} \left(2\Delta y_{it} \Delta y_{i,t-1} + \Delta y_{i,t-1}^2 \right)}{\sum_{i=1}^{n} \sum_{t=2}^{T} \Delta y_{i,t-1}^2}. \tag{39}$$

Note that in this case ϕ is identified even $T = 2$. Interestingly enough, the above linear BMM estimator is identical to the first difference least square (FDLS) estimator proposed by Han and Phillips (2010).[8] *As discussed by Han and Phillips (2010), $\hat{\phi}_n$ given by (39) has standard Gaussian asymptotics for all values of $\phi \in (-1, 1]$ and does not suffer from the weak instrument problem.*

[8] We are grateful to Kazuhiko Hayakawa for drawing our attention to this fact.

Hence the BMM estimator reduces to FDLS estimator under covariance stationarity. However, when T is fixed the covariance stationarity assumption is rather restrictive for most empirical applications in economics, where typically not much is known about the initialization of the dynamic processes over i, and the heteroskedasticity of error variances over t.

Remark 6 *It is possible to augment the QM condition (13) with additional moment conditions to improve asymptotic efficiency. In addition, considering the moment conditions for individual time points t separately as opposed to averaging them across t can also lead to an improved asymptotic efficiency, but it will result in a larger number of moment conditions. It is, however, unclear whether this will necessarily lead to improved performance in finite samples of interest. How to choose the set of moments or how best to combine a possibly large set of moment conditions are both very important ongoing research problems in the literature. Solving these problems is not within the scope of the present paper, which focuses on simple estimation procedures that perform well for all values of n and T and is not subject to the weak instrument problem.*

Remark 7 *When the AR panel data model (1) is augmented with strictly exogenous regressors, namely*

$$y_{it} = \alpha_i + \phi y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}, \text{ for } i = 1, 2, ..., n, \tag{40}$$

where $\mathbf{x}_{it}$ is a $k-1 \times 1$ vector of strictly exogenous regressors, and y_{it} are generated from the initial values, $y_{i,-m_i}$ for $m_i \geq 0$, and $t = -m_i + 1, -m_i + 2..., 1, 2, ..., T$, then it is possible to augment the QM moment condition (13) with standard orthogonality conditions for the strictly exogenous regressors $\mathbf{x}_{it}$. In particular, condition (15), which in the context of ARX model (40) is given by

$$M_{iT}^{(1)}(\phi, \boldsymbol{\beta}) = \frac{1}{T-2} \sum_{t=2}^{T-1} \left[\begin{array}{l} \left(\Delta y_{it} - \phi \Delta y_{i,t-1} - \boldsymbol{\beta}'\Delta\mathbf{x}_{it}\right) \Delta y_{i,t-1} \\ + \left(\Delta y_{it} - \phi \Delta y_{i,t-1} - \boldsymbol{\beta}'\Delta\mathbf{x}_{it}\right)^2 \\ + \left(\Delta y_{i,t+1} - \phi \Delta y_{it} - \boldsymbol{\beta}'\Delta\mathbf{x}_{i,t+1}\right) \Delta y_{it} \end{array} \right], \tag{41}$$

can be augmented with the following $k-1$ standard orthogonality conditions given by (self-instrumenting $\Delta\mathbf{x}_{it}$)

$$\mathbf{M}_{iT}^{(2)}(\phi, \boldsymbol{\beta}) = \frac{1}{T-2} \sum_{t=2}^{T-1} \left(\Delta y_{it} - \phi \Delta y_{i,t-1} - \boldsymbol{\beta}'\Delta\mathbf{x}_{it}\right) \Delta\mathbf{x}_{it}'. \tag{42}$$

There are k unknown parameters, ϕ and $\boldsymbol{\beta}$, and k moment conditions in (41)-(42) which can be *used to derive BMM estimators of ϕ and $\boldsymbol{\beta}$.*

Remark 8 *When $\mathbf{x}_{it}$ are weakly exogenous and the objective of the analysis is impulse-response analysis or forecasting, then one could employ a panel VAR model in $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$, which we will consider below in Section (5). It is also possible to derive the conditional model (40) from the joint distribution of y_{it} and $\mathbf{x}_{it}$. In cases where the joint distribution is given by a VAR model, then the conditional model (40) can be obtained only under very restrictive conditions derived in Section S.2 of the online supplement. Specifically, $\theta_{it} = \boldsymbol{\Omega}_{xx,it}^{-1}\boldsymbol{\omega}_{xy,it}$ must be time invariant, where $\boldsymbol{\omega}_{xy,it} = E\left(\mathbf{u}_{x,it}u_{y,it}\right)$, $\boldsymbol{\Omega}_{xx,it}^{-1} = E\left(\mathbf{u}_{x,it}\mathbf{u}_{x,it}'\right)$, and $\mathbf{u}_{it} = \left(u_{y,it}, \mathbf{u}_{x,it}'\right)'$ are the idiosyncratic innovations in the panel VAR representation of $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$. Finally, in cases where a VAR specification is considered as too restrictive for the analysis of $\mathbf{z}_{it}$, one could follow the GMM literature and use $\Delta\mathbf{x}_{i,t-s}$ $s = 1, 2...$ as instruments for $\Delta\mathbf{x}_{it}$ and augment the resultant moment conditions with the quadratic moment condition given by (13).*

4 Alternative GMM estimators

In order to better place the proposed BMM method in the GMM literature, we consider the sufficient correlation requirements of three alternative GMM estimators.[9] We begin with Anderson and Hsiao (1981, 1982), who considered an IV estimator, where $\Delta y_{i,t-1}$ is instrumented by $\Delta y_{i,t-2}$. This estimator is based on the following moment conditions

$$\text{AH:} \quad E\left[\left(\Delta y_{it} - \phi\Delta y_{i,t-1}\right)\Delta y_{i,t-2}\right] = 0, \text{ for } t = 3, 4, ..., T. \tag{43}$$

A sufficient and necessary condition for ϕ to be identified from the population moment conditions in (43) is $E\left(\Delta y_{i,t-1}\Delta y_{i,t-2}\right) \neq 0$, for some $t \in \{3, 4, ..., T\}$. This condition holds when $|\phi| < 1$, but not if $\phi = 1$.[10] In contrast to the BMM estimator, the AH estimator does not exploit the heteroskedasticity of errors over t for identification of ϕ.

Consider next the moment conditions proposed by Arellano and Bond (1991), where $\Delta y_{i,t-1}$ is

[9] See Bun and Kleibergen (2013) for a related discussion.

[10] Exact conditions for identification of ϕ from the moment conditions that underlie alternative GMM estimators are provided in Section S.3 of the online supplement, in the case where $T = 3$.

instrumented by the levels $y_{i,t-s}$, for $s < t-1$, namely

$$\text{AB:} \quad E\left[\left(\Delta y_{it} - \phi \Delta y_{i,t-1}\right) y_{is}\right] = 0 \text{ for } s = 0, 1, ..., t-2, \text{ and } t = 2, 3, ..., T. \tag{44}$$

As discussed in Section 2, AB conditions are stricter than AH conditions. A necessary condition required for ϕ to be identified from the population moment conditions AB is $|\phi| < 1$. Similarly to AH, the AB estimator is not identified when $\phi = 1$, and the AB moment conditions does not take advantage of time variations in σ_{it}^2. To overcome this problem, Arellano and Bover (1995) and Blundell and Bond (1998), considered additional moment conditions given by

$$\text{BB:} \quad E\left[\Delta y_{i,t-1}\left(y_{it} - \phi y_{i,t-1}\right)\right] = 0, \text{ for } t = 2, 3, ..., T, \tag{45}$$

which do identify ϕ even if $\phi = 1$, regardless of the values of $\sigma_{it}^2 > 0$. The better identification of ϕ is achieved at the expense of more restrictive conditions on the initialization of the AR processes discussed in Section 2. See condition (11), in particular. Note that AB and BB estimators can be implemented for $T \geq 2$.

5 VAR panel data models

5.1 VAR(1) panel data models

The analysis of Section 3 can be readily extended to panel VAR models. Consider the $k \times 1$ vector of variables $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$ and suppose that it is generated by the panel VAR(1) model,

$$\mathbf{z}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\Phi}\mathbf{z}_{i,t-1} + \mathbf{u}_{it}, \tag{46}$$

for $t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T$, and $i = 1, 2, ..., n$, with the starting values given by $\mathbf{z}_{i,-m}$ for $m \geq 0$, where $\boldsymbol{\alpha}_i$ is a $k \times 1$ vector of individual effects, $\boldsymbol{\Phi}$ is a $k \times k$ matrix of slope coefficients, $\mathbf{u}_{it} = (u_{i1t}, u_{i2t}, ..., u_{ikt})'$ is a $k \times 1$ vector of idiosyncratic errors, k is finite and does not depend on n. Similarly, to the univariate case, it is assumed that available observations are $\mathbf{z}_{it}$, for $i = 1, 2, ..., n$, and $t = 0, 1, 2, ..., T$; $T \geq 3$. We consider the following assumptions for the multivariate case which are direct extensions of Assumptions 1-3:

ASSUMPTION 4 *(Parameters of interest) The true values of the $k \times k$ elements of $\mathbf{\Phi} = (\phi_{rs})$, denoted by $\mathbf{\Phi}_0 = \left(\phi_{rs}^0\right)$, are the parameters of interest, and it is assumed that $\mathbf{\Phi} \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a compact set of real-valued $k \times k$ matrices with the largest eigenvalue within or on the unit circle.*

ASSUMPTION 5 *(Idiosyncratic errors) For each $i = 1, 2, ..., n$, the process $\{\mathbf{u}_{it},\ t = -m_i + 1, -m_i + 2,\ ..., 1, 2, ..., T\}$ is distributed with mean $\mathbf{0}$, $E\left(\mathbf{u}_{it}\mathbf{u}_{it}'\right) = \mathbf{\Omega}_{it}$, and there exist positive constants c and K such that $0 < c < \|\mathbf{\Omega}_{it}\| < K$. Moreover, $\bar{\mathbf{\Omega}}_{tn} \equiv n^{-1}\sum_{i=1}^n \mathbf{\Omega}_{it} \to \bar{\mathbf{\Omega}}_t$ as $n \to \infty$, and $\sup_{i,j,t} E\left|u_{ijt}\right|^{4+\epsilon} < K$ for some $\epsilon > 0$. For each t, $\mathbf{u}_{it}$ is independently distributed over i. For each i, $\mathbf{u}_{it}$ is serially uncorrelated over t.*

ASSUMPTION 6 *(Initialization and individual effects) Let $\mathbf{b}_i \equiv \mathbf{\Phi}^{m_i}\left[\boldsymbol{\alpha}_i - (\mathbf{I} - \mathbf{\Phi})\,\mathbf{z}_{i,-m_i}\right] = (b_{i1}, b_{i2}, ..., b_{ik})'$. It is assumed that $\bar{\mathbf{D}}_{b,n} \equiv n^{-1}\sum_{i=1}^n \mathbf{D}_{b,i} \to \bar{\mathbf{D}}_b$ as $n \to \infty$, where $\mathbf{D}_{b,i} = E\left(\mathbf{b}_i\mathbf{b}_i'\right)$, and $\sup_{i,s} E\left|b_{is}\right|^{4+\epsilon} < K$, for some $\epsilon > 0$. In addition, $\mathbf{b}_i$ is independently distributed of $\left(\mathbf{b}_j', \mathbf{u}_{jt}'\right)'$, for all $i \neq j$, $i, j = 1, 2, ..., n$, and all $t = -m_j + 1, -m_j + 2, ..., 1, 2, ..., T$, and the following conditions hold:*

$$E\left(\Delta\mathbf{u}_{it}\mathbf{b}_i'\right) = \mathbf{0},\ \textit{for}\ i = 1, 2, ..., n,\ \textit{and}\ t = 2, 3, ..., T. \tag{47}$$

Taking first-differences of (46), we have

$$\Delta\mathbf{z}_{it} = \mathbf{\Phi}\Delta\mathbf{z}_{i,t-1} + \Delta\mathbf{u}_{it}. \tag{48}$$

Self-instrumenting $\Delta\mathbf{z}_{i,t-1}$, we obtain

$$E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}_{i,t-1}'\right) = -\mathbf{\Omega}_{i,t-1}. \tag{49}$$

Similarly to Section 3, we use $E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{u}_{it}'\right) = \mathbf{\Omega}_{i,t-1} + \mathbf{\Omega}_{it}$ and $E\left(\Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}_{it}\right) = -\mathbf{\Omega}_{it}$ to obtain the following QM conditions,

$$E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}_{i,t-1}'\right) + E\left[\Delta\mathbf{u}_{it}\Delta\mathbf{u}_{it}'\right] + E\left(\Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}_{it}'\right) = \mathbf{0}, \tag{50}$$

for $i = 1, 2, ..., n$, and $t = 2, 3, ..., T - 1$. (50) is a multivariate version of (13).

Averaging moment conditions (13) over t, we obtain (similarly to (14))

$$E\left[\mathbf{M}_{iT}\left(\mathbf{\Phi}\right)\right] = \mathbf{0},\ \text{for}\ i = 1, 2, ..., n, \tag{51}$$

where

$$\mathbf{M}_{iT}(\mathbf{\Phi}) = \frac{1}{T-2}\sum_{t=2}^{T-1}(\Delta\mathbf{z}_{it}-\mathbf{\Phi}\Delta\mathbf{z}_{i,t-1})\,\Delta\mathbf{z}'_{i,t-1}+\frac{1}{T-2}\sum_{t=2}^{T-1}(\Delta\mathbf{z}_{it}-\mathbf{\Phi}\Delta\mathbf{z}_{i,t-1})\,(\Delta\mathbf{z}_{it}-\mathbf{\Phi}\Delta\mathbf{z}_{i,t-1})' + \frac{1}{T-2}\sum_{t=2}^{T-1}(\Delta\mathbf{z}_{i,t+1}-\mathbf{\Phi}\Delta\mathbf{z}_{it})\,\Delta\mathbf{z}'_{it}. \tag{52}$$

The BMM estimator of $\mathbf{\Phi}$ is given by

$$\hat{\mathbf{\Phi}}_{nT} = \arg\min_{\mathbf{\Phi}\in\mathbf{\Theta}} \left\|\bar{\mathbf{M}}_{nT}(\mathbf{\Phi})\right\|, \tag{53}$$

where $\mathbf{\Theta}$ is a compact set of admissible values of $\mathbf{\Phi}$ defined in Assumption 4, and $\bar{\mathbf{M}}_{nT}(\mathbf{\Phi}) = n^{-1}\sum_{i=1}^{n}\mathbf{M}_{iT}(\mathbf{\Phi})$. To derive the asymptotic properties of $\hat{\mathbf{\Phi}}_{nT}$, let $\mathbf{\Phi}_0$ denote the true value of $\mathbf{\Phi}\in\mathbf{\Theta}$, and note that under $\mathbf{\Phi}=\mathbf{\Phi}_0$, (48) yields $\Delta\mathbf{z}_{it}=\mathbf{\Phi}_0\Delta\mathbf{z}_{i,t-1}+\Delta\mathbf{u}_{it}$, and (52) can be written as

$$\mathbf{M}_{iT}(\mathbf{\Phi}) = \mathbf{\Lambda}_{iT}+\mathbf{V}_{iT}, \tag{54}$$

where

$$\mathbf{V}_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-1}+\Delta\mathbf{u}_{it}\Delta\mathbf{u}'_{it}+\Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}'_{it}\right), \tag{55}$$

$$\mathbf{\Lambda}_{iT} = (\mathbf{\Phi}-\mathbf{\Phi}_0)\,\mathbf{Q}_{iT}\,(\mathbf{\Phi}-\mathbf{\Phi}_0)' - (\mathbf{\Phi}-\mathbf{\Phi}_0)\left(\mathbf{Q}_{iT}+\mathbf{Q}^{+}_{iT}-\mathbf{H}_{iT}\right)+\mathbf{H}_{iT}\,(\mathbf{\Phi}-\mathbf{\Phi}_0)',$$

and

$$\mathbf{Q}_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta\mathbf{z}_{i,t-1}\Delta\mathbf{z}'_{i,t-1},\ \mathbf{Q}^{+}_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta\mathbf{z}_{it}\Delta\mathbf{z}'_{it},\ \mathbf{H}_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-1}.$$

We have k^2 unknown parameters in $\mathbf{\Phi}$ and k^2 moment conditions in (51). Suppose there exists $\hat{\mathbf{\Phi}}_{nT}$ such that $\bar{\mathbf{M}}_{nT}\left(\hat{\mathbf{\Phi}}_{nT}\right)=\mathbf{0}$. Then (54) evaluated at $\mathbf{\Phi}=\hat{\mathbf{\Phi}}_{nT}$ yields the following multivariate version of (22),

$$\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)\bar{\mathbf{Q}}_{nT}\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)' - \bar{\mathbf{H}}_{nT}\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)' - \left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)\left(\bar{\mathbf{H}}'_{nT}+\bar{\mathbf{Q}}_{nT}+\bar{\mathbf{Q}}^{+}_{nT}\right) = -\bar{\mathbf{V}}_{nT}, \tag{56}$$

where $\bar{\mathbf{Q}}_{nT}=n^{-1}\sum_{i=1}^{n}\mathbf{Q}_{iT}$, $\bar{\mathbf{Q}}^{+}_{nT}=n^{-1}\sum_{i=1}^{n}\mathbf{Q}^{+}_{iT}$, $\bar{\mathbf{H}}_{nT}=n^{-1}\sum_{i=1}^{n}\mathbf{H}_{iT}$, and $\bar{\mathbf{V}}_{nT}=n^{-1}\sum_{i=1}^{n}\mathbf{V}_{iT}$. Similarly to the univariate case, we show that there exists unique $\sqrt{n}$-consistent solution. Suppose that $\hat{\mathbf{\Phi}}_{nT}$ is a $\sqrt{n}$-consistent estimator of $\mathbf{\Phi}$. Then the first term (56) is $O_p\left(n^{-1}\right)$, and we obtain, using (56),

$$\sqrt{n}\bar{\mathbf{H}}_{nT}\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)'+\sqrt{n}\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)\left(\bar{\mathbf{H}}'_{nT}+\bar{\mathbf{Q}}_{nT}+\bar{\mathbf{Q}}^{+}_{nT}\right)=\sqrt{n}\bar{\mathbf{V}}_{nT}+O_p\left(n^{-1/2}\right),$$

Also, since by Lemma A.4, $\sqrt{n}\bar{V}_{nT}=O_p(1)$, then the $\sqrt{n}$-consistent estimator is unique, if it exists. Vectorizing above equations, we have[11]

$$\begin{aligned}&\sqrt{n}\left(\mathbf{I}_k\otimes\bar{\mathbf{H}}_{nT}\right)\mathbf{R}Vec\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)+\sqrt{n}\left[\left(\bar{\mathbf{H}}_{nT}+\bar{\mathbf{Q}}_{nT}+\bar{\mathbf{Q}}^{+}_{nT}\right)\otimes\mathbf{I}_k\right]Vec\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)\\=\;&Vec\left(\sqrt{n}\bar{\mathbf{V}}_{nT}\right)+O_p\left(n^{-1/2}\right),\end{aligned}$$

where $\mathbf{R}$ is $k^2\times k^2$ re-ordering matrix uniquely defined by $Vec\left[\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)'\right]=\mathbf{R}Vec\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)$. Let

$$\bar{\mathbf{B}}_{nT}=n^{-1}\sum_{i=1}^{n}\bar{\mathbf{B}}_{iT},\text{ and }\bar{\mathbf{B}}_{iT}=\left(\mathbf{I}_k\otimes\mathbf{H}_{iT}\right)\mathbf{R}+\left(\mathbf{H}_{iT}+\mathbf{Q}_{iT}+\mathbf{Q}^{+}_{iT}\right)\otimes\mathbf{I}_k. \tag{57}$$

Using Lemma A.3 in appendix, we have $\bar{\mathbf{B}}_{nT}=E\left(\bar{\mathbf{B}}_{nT}\right)+O_p\left(n^{-1/2}\right)$. Let

$$\bar{\mathbf{B}}_T=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}E\left(\bar{\mathbf{B}}_{iT}\right). \tag{58}$$

(58) is a multivariate version of (31). Similarly to the univariate case, we require that $\bar{\mathbf{B}}_T$ is invertible for $\sqrt{n}$-consistency. Assuming $\bar{\mathbf{B}}_T$ is invertible, it then follows that the $\sqrt{n}$-consistent estimator, $\hat{\mathbf{\Phi}}_{nT}$, must be unique. Finally, using (A.14) of Lemma A.4, we obtain

$$\sqrt{n}Vec\left(\hat{\mathbf{\Phi}}_{nT}-\mathbf{\Phi}_0\right)\to_d N\left(\mathbf{0},\bar{\mathbf{B}}_T^{-1}\mathbf{S}_T\bar{\mathbf{B}}_T^{-1\prime}\right),$$

where $\mathbf{S}_T=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}E\left[Vec\left(\mathbf{V}_{iT}\right)Vec\left(\mathbf{V}_{iT}\right)'\right]$.

Remark 9 *If $\bar{\mathbf{B}}_T\neq\mathbf{0}$ is a singular matrix, then some elements of $\hat{\mathbf{\Phi}}_{nT}$ can not be $\sqrt{n}$-consistent. If $\bar{\mathbf{B}}_T=\mathbf{0}$ and $\bar{\mathbf{Q}}_T=\lim_{n\to\infty}\bar{\mathbf{Q}}_{nT}$ is positive definite, then there exists a unique $n^{1/4}$-consistent estimator.*

[11] Note that for any $p\times p$ generic matrices $\mathbf{A}$ and $\mathbf{X}$, we have $Vec\left(\mathbf{A}\mathbf{X}'\right)=\left(\mathbf{I}_p\otimes\mathbf{A}\right)Vec\left(\mathbf{X}'\right)$, and $Vec\left(\mathbf{X}\mathbf{A}\right)=\left(\mathbf{A}'\otimes\mathbf{I}_p\right)Vec\left(\mathbf{X}\right)$.

The following theorem summarizes our results for the case where $\bar{\mathbf{B}}_{nT}$ is nonsingular for all n, and as $n \to \infty$.

Theorem 2 *Suppose* $\mathbf{z}_{it}$*, for* $i = 1, 2, ..., n$*, and* $t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T$*, are generated by (46) with starting values* $\mathbf{z}_{i,-m_i}$*, and the true value of the parameters of interest* $\mathbf{\Phi}_0$*. Let Assumptions 4-6 hold, and suppose* $\bar{\mathbf{B}}_T = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\left(\bar{\mathbf{B}}_{iT}\right)$ *is nonsingular, and* $\mathbf{S}_T = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n E\left[Vec\left(\mathbf{V}_{iT}\right) Vec\left(\mathbf{V}_{iT}\right)'\right]$*, where* $\bar{\mathbf{B}}_{iT}$ *is defined in (57) and* $\mathbf{V}_{iT}$ *is defined in (55). Consider the BMM estimator,* $\hat{\mathbf{\Phi}}_{nT}$*, defined by (53), and let* T *be fixed as* $n \to \infty$*. Then, the unique* $\sqrt{n}$*-consistent estimator* $\hat{\mathbf{\Phi}}_{nT}$ *satisfies*

$$\sqrt{n} Vec\left(\hat{\mathbf{\Phi}}_{nT} - \mathbf{\Phi}_0\right) \to_d N\left(\mathbf{0}, \mathbf{\Sigma}_T\right),$$

where

$$\mathbf{\Sigma}_T = \bar{\mathbf{B}}_T^{-1} \mathbf{S}_T \bar{\mathbf{B}}_T^{-1\prime}.$$

Similarly to the univariate case, $\mathbf{\Sigma}_T$ can be consistently estimated by[12]

$$\hat{\mathbf{\Sigma}}_{nT} = \hat{\bar{\mathbf{B}}}_{nT}^{-1} \hat{\mathbf{S}}_{nT} \hat{\bar{\mathbf{B}}}_{nT}^{-1\prime},$$

where $\hat{\bar{\mathbf{B}}}_{nT} = n^{-1} \sum_{i=1}^n \hat{\bar{\mathbf{B}}}_{i,nT}$,

$$\hat{\bar{\mathbf{B}}}_{i,nT} = \left(\mathbf{I}_k \otimes \hat{\mathbf{H}}_{i,nT}\right) \mathbf{R} + \left(\hat{\mathbf{H}}_{i,nT} + \mathbf{Q}_{iT} + \mathbf{Q}_{iT}^{+}\right) \otimes \mathbf{I}_k,$$

$\hat{\mathbf{H}}_{i,nT} = (T-2)^{-1} \sum_{t=2}^{T-1} \Delta\hat{\mathbf{u}}_{it} \Delta\mathbf{z}_{i,t-1}'$, $\Delta\hat{\mathbf{u}}_{it} = \Delta\mathbf{z}_{it} - \hat{\mathbf{\Phi}}_{nT} \Delta\mathbf{z}_{i,t-1}$, ($\Delta\hat{\mathbf{u}}_{it}$ depends on n and T, but we omit subscripts n, T to simplify the notations),

$$\hat{\mathbf{S}}_{nT} = \left(\frac{1}{n} \sum_{i=1}^n Vec\left(\hat{\mathbf{V}}_{i,nT}\right) Vec\left(\hat{\mathbf{V}}_{i,nT}\right)'\right),$$

and

$$\hat{\mathbf{V}}_{i,nT} = \frac{1}{T-2} \sum_{t=2}^{T-1} \left(\Delta\hat{\mathbf{u}}_{it} \Delta\mathbf{z}_{i,t-1}' + \Delta\hat{\mathbf{u}}_{it} \Delta\mathbf{u}_{it}' + \Delta\hat{\mathbf{u}}_{i,t+1} \Delta\mathbf{z}_{it}'\right).$$

[12] Consistency of $\hat{\mathbf{\Sigma}}_{nT}$ in the multivariate case can be established in the same way as in the univariate case in Proposition 1 in appendix.

5.2 Panel VAR(p) models

The BMM procedure can be readily extended to higher order panel AR or panel VARs. Consider the panel VAR model of order p, VAR(p):

$$\mathbf{z}_{it} = \boldsymbol{\alpha}_i + \sum_{\ell=1}^{p} \boldsymbol{\Phi}_\ell \mathbf{z}_{i,t-\ell} + \mathbf{u}_{it}, \tag{59}$$

for $t = -m_i+1, -m_i+2, ..., 1, 2, ..., T$ and $i = 1, 2, ..., n$, with the starting values given by $\mathbf{z}_{i,-m_i-p+1}$, $\mathbf{z}_{i,-m_i-p+2}$, ..., $\mathbf{z}_{i,-m_i}$ for $m_i \geq 0$, and some $p \geq 1$. Suppose that available observations are $\mathbf{z}_{it}$, for $i = 1, 2, ..., n$, and $t = 0, 1, 2, ..., T$; $T \geq p+2$. The number of time periods required is $p+3$. Using the first-differenced form of (59) and self-instrumenting $\Delta\mathbf{z}_{i,t-1}$ we obtain the following quadratic matrix bias-corrected moment conditions

$$E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-1}\right) + E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{u}'_{it}\right) + E\left(\Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}'_{it}\right) = \mathbf{0}. \tag{60}$$

Self-instrumenting the higher order lags we have

$$E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-\ell}\right) = \mathbf{0}, \text{ for } \ell = 2, 3, ..., p. \tag{61}$$

Averaging the above moment conditions over t, and using $\Delta\mathbf{u}_{it} = \Delta\mathbf{z}_{it} - \sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t-\ell}$, we have

$$E\left[\mathbf{M}_{iT}\left(\boldsymbol{\Phi}\right)\right] = \mathbf{0}, \text{ for } i = 1, 2, ..., n, \tag{62}$$

where $\boldsymbol{\Phi}$ is the $k \times pk$ parameter matrix of interest defined by $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, ..., \boldsymbol{\Phi}_p)$, $\mathbf{M}_{iT}\left(\boldsymbol{\Phi}\right) = \left(\mathbf{M}_{iT}^{(1)}\left(\boldsymbol{\Phi}\right), \mathbf{M}_{iT}^{(2)}\left(\boldsymbol{\Phi}\right), ..., \mathbf{M}_{iT}^{(p)}\left(\boldsymbol{\Phi}\right)\right)$,

$$\begin{aligned}\mathbf{M}_{iT}^{(1)}\left(\boldsymbol{\Phi}\right) &= \frac{1}{T-p-1}\sum_{t=p+1}^{T-1}\left(\Delta\mathbf{z}_{it} - \sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t-\ell}\right)\Delta\mathbf{z}'_{i,t-1} \\ &+ \frac{1}{T-p-1}\sum_{t=p+1}^{T-1}\left(\Delta\mathbf{z}_{it} - \sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t-\ell}\right)\left(\Delta\mathbf{z}_{it} - \sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t-\ell}\right)' \\ &+ \frac{1}{T-p-1}\sum_{t=p+1}^{T-1}\left(\Delta\mathbf{z}_{i,t+1} - \sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t+1-\ell}\right)\Delta\mathbf{z}'_{it},\end{aligned} \tag{63}$$

and

$$\mathbf{M}_{iT}^{(\ell)}(\mathbf{\Phi}) = \frac{1}{T-p-1}\sum_{t=p+1}^{T-1}\left(\Delta\mathbf{z}_{it} - \sum_{\ell=1}^{p}\mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell}\right)\Delta\mathbf{z}'_{i,t-\ell}, \text{ for } \ell = 2, 3, ..., p.$$

The BMM estimator of $\mathbf{\Phi}$ can now be computed as

$$\hat{\mathbf{\Phi}}_{nT} = \arg\min_{\mathbf{\Phi}\in\mathbf{\Theta}} \left\|\bar{\mathbf{M}}_{nT}(\mathbf{\Phi})\right\|, \tag{64}$$

where as before $\bar{\mathbf{M}}_{nT}(\mathbf{\Phi}) = n^{-1}\sum_{i=1}^{n}\mathbf{M}_{iT}(\mathbf{\Phi})$, and $\mathbf{\Theta}$ is a compact set of admissible values of $\mathbf{\Phi}$ such that all roots of the characteristic equation $\left|\mathbf{I}_k - \sum_{\ell=1}^{p}\mathbf{\Phi}_\ell x^\ell\right| = 0$ lie outside or on the unit circle. Note that there are k^2p unknown coefficients in $\mathbf{\Phi}$, and the same number, k^2p moment conditions in (63) and (64).

The BMM estimator, $\hat{\mathbf{\Phi}}_{nT}$, has the same asymptotic properties as in the case of the VAR(1) specification and the same lines of proof applies here. The proof can be simplified by using a VAR(1) companion form of (59).

It is also possible to extend the BMM procedure to accommodate unbalanced panels and time effects. Consistent estimation of average error covariances, $\bar{\mathbf{\Omega}}_t$, for $t = 1, 2, ..., T$, is also possible. For details see Sections S.4 and S.5 of the online supplement.

6 Monte Carlo Evidence

We now provide some evidence on the small sample performance of the BMM estimator as compared to a number of key GMM estimators proposed in the literature.

6.1 Data generating process (DGP)

The dependent variable is generated as

$$y_{it} = \alpha_i + \phi y_{i,t-1} + u_{it}, \tag{65}$$

for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m_i + 2, ..., T$. We consider two values for ϕ, namely $\phi = 0.4$, and 0.8. Individual effects are generated as

$$\alpha_i = \alpha + w_i, \tag{66}$$

where $w_i \sim IIDN\left(0, \sigma_w^2\right)$. The processes are initialized as

$$y_{i,-m_i} = \kappa_i \mu_i + \upsilon_i, \tag{67}$$

where $\mu_i = \alpha_i / (1-\phi)$, κ_i is generated as $\kappa_i \sim IIDU(0.5, 1.5)$ so that $E(\kappa_i) = 1$, and $\upsilon_i \sim IIDN(\mu_\upsilon, 1)$. We consider two options for μ_υ, namely $\mu_\upsilon = 0$ and 1. We set $\alpha = \sigma_w^2 = 1$, which ensures that $V(\alpha_i) = 1$.

DGP (65)-(67) implies (also see (4))

$$\Delta y_{i1} = -(1-\phi)\phi^{m_i}\left[(\kappa_i - 1)\mu_i + \upsilon_i\right] + u_{i1} - (1-\phi)\sum_{\ell=0}^{m_i-1}\phi^\ell u_{i,-\ell}. \tag{68}$$

The times at which the processes are generated, namely $-m_i$ before the initial observation y_{i0}, are allowed to differ across i and are generated as $m_i \sim IIDU[1,4]$. The idiosyncratic errors, u_{it}, are generated as non-Gaussian processes with heteroskedastic error variances, namely $u_{it} = (e_{it} - 2)\sigma_{ia}/2$ for $t \leq [T/2]$, and $u_{it} = (e_{it} - 2)\sigma_{ib}/2$ for $t > [T/2]$, with $\sigma_{ia}^2 \sim IIDU(0.25, 0.75)$, $\sigma_{ib}^2 \sim IIDU(1,2)$, and $e_{it} \sim IID\chi^2(2)$, where $[T/2]$ is the integer part of $T/2$. σ_{ia}^2 and σ_{ib}^2 are generated independently of e_{it}. This ensures that the errors have zero means, and are conditionally heteroskedastic, in particular, $V(u_{it}|\sigma_{ia}) = \sigma_{ia}^2$ for $t \leq [T/2]$, and $V(u_{it}|\sigma_{ib}) = \sigma_{ib}^2$ for $t > [T/2]$. We set $T = 3, 5, 10, 20$, and $n = 250, 500, 1,000$, and run each experiment with $2,000$ replications.[13]

Besides the parameter of interest ϕ, the key parameters of the MC design is μ_υ. This parameter affects the relative performance of BMM and IV/GMM estimators. It is easily seen that

$$E\left[\mu_i\left(y_{i,-m} - \mu_i\right)\right] = \frac{\alpha\mu_\upsilon}{1-\phi}.$$

AH, BMM and AB estimators are valid for all values of μ_υ. But, as it is already shown in Subsection 2.1, the validity of the BB estimator requires $\mu_\upsilon = 0$. See condition (11).

Finally, since our theory suggests that the BMM estimator should work even for n and T larger, in addition to small values of $T = 5, 10, 20$, we also consider its performance when T is large. But to save space we provide results for BMM estimator for values of $T = 100, 250, 500$, and $n = 250, 500, 1000$ in the online supplement.

[13] Results for experiments with larger values of n, namely $n = 2000, 5000$, and $10,000$ are provided in the online supplement.

6.2 Estimation methods

6.2.1 BMM estimator

We implement the BMM estimator given by (16) with its variance estimated using (36).

6.2.2 Alternative GMM estimators

As alternatives to the BMM estimator we also included the IV estimator due to Anderson and Hsiao (1981, 1982), the first-difference GMM methods based on the moment conditions considered by Arellano and Bond (1991), and the system GMM methods based on the moment conditions considered by Arellano and Bover (1995) and Blundell and Bond (1998).

Anderson and Hsiao (1981, 1982), hereafter AH estimator makes use of the following moment conditions

$$E\left[\Delta y_{i,t-1}\left(\Delta y_{it}-\phi\Delta y_{i,t-1}\right)\right]=0,\ \text{for } t=3,4,...,T. \tag{69}$$

The resultant estimator is obtained by averaging the above moment conditions over t, which leads to the AH estimator given by equation (8.1) of Anderson and Hsiao (1981). A consistent estimator of the asymptotic variance of the AH estimator is provided in Section S.6 of the online supplement.

The first-difference and system GMM methods are based on a larger set of moment conditions and are thus overidentified. For both of these estimators, we consider two sets of moment conditions – a full set and a subset, with the latter aimed at lowering the number of moment conditions, and thus possibly improving the small sample performance of these estimators at the expense of asymptotic efficiency. Specifically, the first set of first-difference moment conditions proposed by Arellano and Bond (1991) and denoted as "DIF1" consists of

$$\text{DIF1: } E\left[y_{is}\left(\Delta y_{it}-\phi\Delta y_{i,t-1}\right)\right]=0, \text{ for } s=0,1,...,t-2, \text{ and } t=2,3,...,T, \tag{70}$$

which gives $h=3,10,45,190$ moment conditions for $T=3,5,10,20$, respectively. The second set of moment conditions, denoted as "DIF2", is a subset of DIF1 and consists of

$$\text{DIF2: } E\left[y_{i,t-2-\ell}\left(\Delta y_{it}-\phi\Delta y_{i,t-1}\right)\right]=0, \text{ with } \ell=0 \text{ for } t=2, \text{ and } \ell=0,1 \text{ for } t=3,4,...,T, \tag{71}$$

which gives $h=3,7,17,27$ moment conditions for $T=3,5,10,20$, respectively. Hence, for $T=3$, DIF1 and DIF2 are the same, but differ for $T > 3$.

We also consider the system-GMM estimators (Arellano and Bover (1995) and Blundell and Bond (1998)) and augment DIF1 and DIF2 moment conditions with

$$E\left[\Delta y_{i,t-1}\left(y_{it} - \phi y_{i,t-1}\right)\right] = 0, \text{ for } t = 2, 3, ..., T, \tag{72}$$

and denote the estimators based on the augmented sets of moment conditions as "SYS1" and "SYS2", respectively. As discussed in Section 2, additional moment conditions in (72) have been proposed to deal with the mentioned deficiency of the first-difference GMM methods in the case of a highly persistent dependent variable at the expense of stricter requirements on the initialization of dynamic processes. For SYS1, we have $h = 5, 14, 54, 209$ moment conditions for $T = 3, 5, 10, 20$, respectively, while for SYS2 we have $h = 5, 11, 26, 56$ moment conditions for $T = 3, 5, 10, 15$, respectively.

We implement one-step, two-step and continuous updating (CU) versions of DIF and SYS type GMM estimators, based on each of the four sets of moment conditions outlined above.[14]

First-difference and system GMM inference In addition to the conventional standard errors, we also consider Windmeijer (2005)'s standard errors with finite sample corrections for the two-step GMM estimators and Newey and Windmeijer (2009)'s alternative standard errors for the CU-GMM estimators. For the computation of optimal weighting matrix, a centered version is used except for the CU-GMM.

6.3 Monte Carlo findings

Here we focus on reporting the results for the more challenging case of $\phi = 0.8$, and relegate the findings for the experiments with $\phi = 0.4$ to the online supplement.

First we consider the experiment where the deviations of the initial values from the long-run means have zero means ($\mu_v = 0$), labeled as Experiment 1. As noted already, all the estimators considered are valid asymptotically, and any observed differences across them must be due to small samples and the fact that they differ in asymptotic efficiency. Table 1a reports findings for the bias and RMSE (both $\times 100$) of estimating ϕ, and Table 1.b shows the results for size and power

[14] We use the Matlab code provided to us by Hayakawa and Pesaran (2015). See Section 4.1 of Hayakawa and Pesaran (2015) where a more detailed description of these methods is provided.

of the tests. As to be expected, the performance of all estimators improve with n, although there are significant differences in the small sample performances of the different estimators, with the AH estimator doing quite poorly when $T < 10$, which has also been documented in the literature. Similarly, the first-difference GMM methods, which rely on lagged levels as instruments, are not doing that well when T is small, although they clearly beat the AH estimator in terms of the bias and RMSE. The BMM and system-GMM estimators are the best performing. Despite the fact that the system-GMM estimators are asymptotically more efficient than the BMM estimator, for $T = 3$ and all values of n considered the BMM estimator performs better in terms of RMSE. For larger choices of T (or larger choices of n reported in Table S1a of the online supplement), it becomes clear that the BMM estimator is asymptotically not the most efficient, since it does not exploit the additional restrictions, (11), that underlie the system-GMM estimators. In terms of bias, the BMM estimator performs quite well in comparison with the system GMM estimators whose bias seems to vary considerably across estimators and sample sizes.

Size and power of tests based on the different estimators at the 5% nominal level are reported in Table 1b. The BMM estimator achieves good size (close to 5%) for all choices of n and T. The size of the AH estimator is also good, but its power is very low, as to be expected based on the RMSE findings. The tests based on the first difference GMM estimators are in majority of cases oversized, but the size distortions decrease in n and eventually disappear for a sufficiently large $n > 1,000$ (reported in Table S1b in the online supplement for $n = 2,000$, $5,000$ and $10,000$). The size distortions of the system-GMM methods appear to be more serious, and in a few cases (for $T > 5$) the reported rejection rates exceed 50%.[15] These are serious small sample problems, which, as in the case of the first-difference GMM methods, eventually disappear once n becomes sufficiently large. The power of the BMM estimator is quite good compared with the GMM estimators, but the power comparisons are rather complicated because of the size distortions of the first-difference and system GMM estimators. For $T = 3$, and $n = 1000$ the power of the tests based on the BMM estimator is 58.9, about five-fold increase over the first-difference GMM methods, and it is matched only by CU-GMM estimators, which are slightly oversized.

Next we consider the experiment with non-zero values of μ_υ (Experiment 2, $\mu_\upsilon = 1$), whilst keeping all other design parameters unchanged. As noted earlier, not all of the moment conditions

[15] The use of standard errors proposed by Windmeijer (2005) and Newey and Windmeijer (2009) to overcome the size distortion of the system-GMM estimators help but do not fully resolve the problem unless n is sufficiently large.

for the system GMM hold under this scenario, and the system GMM estimators are no longer consistent. This is confirmed by the large biases reported for the system GMM estimators in Table 2.a, and the substantial size distortions reported for these estimators in Table 2.b.[16] The remaining estimators, BMM, AH and the first-difference GMM methods are robust to the choice of μ_v and continue to perform well. In fact, increases in μ_v can result in an improved correlation between the target variable $\Delta y_{i,t-1}$ and its instruments (be it lagged differences, or lagged levels), and we see that the gains in RMSE and bias are quite large for the first-difference GMM estimators, but very minor for the BMM estimator, which is not subject to the weak instrument problem and was previously performing well anyway.[17] As in Experiment 1, the BMM estimator need not be asymptotically the most efficient, but its performance appears very good in small samples. For $T = 3, 5$, its RMSE again outperforms the first-difference GMM estimators by a large margin. The tests based on the BMM estimator continue to perform well in terms of size, whereas the performance of the first-difference GMM estimators is mixed, with severe over-rejections reported for selected larger values of T, with the exception of the 1-step estimator based on the restricted set of moment conditions "DIF2". Such deterioration in inference as T is increased is therefore likely to be due to the proliferation of moment conditions resulting from an increase in T. In terms of the power findings, we again see that the tests based on the BMM estimator are substantially more powerful compared with the first-difference GMM methods for smaller values of $T < 10$, where majority of the first-difference GMM methods do not show very large size distortions. All of the system GMM methods are, not surprisingly grossly oversized and therefore the power comparisons are not meaningful. We conclude from Experiment 2, where only the BMM, AH, and the first-difference GMM estimators are asymptotically valid, that the BMM estimator can significantly outperform its main competitors.

Overall, the MC findings show that the BMM estimator is robust and outperform its 'cousin' AH estimator by a large margin. The system GMM estimators are not robust to $\mu_v \neq 0$, and are thus more restrictive. In the case of Experiment 1 (with $\mu_v = 0$) where all estimators are valid, the BMM estimator is not the most efficient asymptotically, but it performs comparably well for the

[16] The bias and size distortions of the System-GMM estimators do not vanish even for larger values of n. See Tables S2a and S2b in the online supplement.

[17] The improvement in the performance of the first-difference GMM estimators is in line with Hayakawa (2009) and Hayakawaa and Nagata (2016), who investigate the effects of mean-nonstationarity on the first-difference GMM estimators.

choices of n and T considered, and in some instances better than the first-difference and system GMM methods. It is also remarkable that, in contrast to the first-difference and system-GMM estimators, the size of the tests based on the BMM estimator is reliable for all choices of n and T in all experiments considered. Hence, we conclude that the BMM estimator works well, regardless of whether μ_v is zero or not, albeit it could be less efficient for some choices of n and T. However, in practice it is not known whether conditions required for the system GMM estimators regarding the initialization of dynamic processes are satisfied, and violation of these conditions can cause large biases and wrong inference.

The important parameter that affects the performance of the BMM estimator is the magnitude of $\bar{B}_T$ given by (31). The BMM estimator will not perform well in designs with $\bar{B}_T$ close or equal to zero. As highlighted in Remark 4, $\bar{B}_T$ is zero in the leading unit root case with homoskedastic errors. The performance of the BMM estimator when $\phi = 1$ is documented in Table S5 in the online supplement.

Finally, it is worth noting that BMM and AH estimators remain applicable also when T is large, whereas the first-difference and system GMM methods deteriorate and eventually become infeasible as T increases, unless the proliferation of moment condition is somehow controlled. To demonstrate that the BMM and AH estimators remain applicable regardless whether T is small or not in relation with n, we also report selected results for $T = 100, 250, 500$ in an online supplement (Table S6). These experiments confirm that the BMM estimator continues to perform well for large values of T, and also that the power of the tests based on the AH estimator will improve with an increase in T. However, when both n and T are large, alternative estimators developed in the literature that allow for slope heterogeneity and unobserved common factors, both of which are likely to be important in practice, can be applied. Therefore the main appeal of the BMM estimators developed in this paper is, in our view, for panels where the more general large-n, large-T estimators break down due to small time dimension.

7 Conclusion

This paper proposes the idea of self-instrumenting target variables instead of searching for instruments that are uncorrelated with the errors, in cases where the correlation between the target variables and the errors can be derived. In this paper this idea is applied to the estimation of short-T dynamic panel data models, and a simple bias-corrected methods of moments (BMM) estimators are proposed. The BMM estimators are applicable under less restrictive conditions on the initialization of the dynamic processes and the individual effects as compared to the leading first-difference and system-GMM methods advanced in the literature. It is, however, acknowledged that the BMM estimators can be less efficient asymptotically when the stricter requirements of the first-difference and system GMM estimators hold. The robustness of the BMM estimators is likely to be an advantage in practice where it is not possible to know if the stronger requirements of the GMM estimators are met, and thus avoid possible estimation bias and incorrect inference.

The idea of self-instrumenting opens new exciting research avenues. This idea could be considered in other settings, including spatial panel data models. The idea can also be exploited to estimate unknown parameters of a known distributional functional form of slope coefficients in short-T autoregressive or vector autoregressive panels with heterogenous slope coefficients. Last but not least, we have also left the topic of combining the new moment condition proposed in this paper with the existing moment conditions considered in the GMM literature to future research.

Table 1a: Bias and RMSE of alternative estimates of ϕ for Experiment 1

$\phi = 0.8$, $\mu_v = 0$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n						**Bias (x100)**								
3	**250**	1.08	-104.85	-39.67	-38.98	-29.58	-39.67	-38.98	-29.58	3.16	6.69	4.53	3.16	6.69	4.53
3	**500**	0.35	208.31	-27.79	-27.29	-18.20	-27.79	-27.29	-18.20	0.93	4.39	1.36	0.93	4.39	1.36
3	**1000**	0.26	208.87	-14.77	-13.02	-4.60	-14.77	-13.02	-4.60	0.32	2.47	0.48	0.32	2.47	0.48
5	**250**	1.07	24.14	-23.55	-21.18	-2.56	-18.98	-16.71	0.86	4.81	3.67	0.93	5.16	3.42	0.46
5	**500**	0.48	14.20	-13.64	-10.66	2.05	-9.84	-8.21	2.23	2.63	1.73	-0.02	2.82	1.51	-0.30
5	**1000**	0.22	5.85	-7.64	-5.87	0.94	-5.20	-4.34	0.67	1.35	0.86	0.04	1.46	0.73	-0.12
10	**250**	0.79	7.87	-10.99	-9.04	2.26	-10.03	-8.86	2.27	4.90	3.40	0.04	6.13	3.52	-0.91
10	**500**	0.32	2.97	-6.53	-4.92	1.11	-5.79	-4.95	0.17	3.05	1.52	-0.07	3.75	1.43	-0.53
10	**1000**	0.12	1.29	-3.49	-2.45	0.71	-2.77	-2.32	0.13	1.81	0.58	-0.05	2.21	0.46	-0.32
20	**250**	0.42	3.03	-4.37	-3.52	1.04	-3.21	-2.91	0.23	3.50	2.80	0.29	6.09	3.54	-1.35
20	**500**	0.15	1.89	-2.40	-1.69	0.75	-1.65	-1.46	0.08	2.14	1.15	0.00	3.77	1.37	-0.83
20	**1000**	0.05	0.54	-1.29	-0.75	0.47	-0.81	-0.76	0.06	1.19	0.39	-0.03	2.14	0.37	-0.44
							RMSE(x100)								
3	**250**	10.71	1517.32	81.92	89.23	103.34	81.92	89.23	103.34	13.81	14.25	18.05	13.81	14.25	18.05
3	**500**	6.69	6911.35	63.47	69.56	81.76	63.47	69.56	81.76	10.89	9.94	9.55	10.89	9.94	9.55
3	**1000**	4.75	6607.87	43.84	44.76	50.00	43.84	44.76	50.00	7.02	6.00	5.27	7.02	6.00	5.27
5	**250**	8.29	499.48	34.84	36.39	50.06	34.95	35.59	48.25	8.22	7.44	7.74	8.41	7.41	7.87
5	**500**	5.62	134.32	23.40	22.57	27.36	22.81	22.16	27.70	6.04	4.58	3.83	6.11	4.46	3.80
5	**1000**	3.86	40.34	15.88	15.04	16.02	15.65	14.91	15.74	4.36	2.88	2.47	4.37	2.81	2.46
10	**250**	6.11	48.62	14.11	13.70	14.00	17.08	17.26	21.53	6.35	5.04	3.17	7.26	5.29	3.65
10	**500**	4.22	27.66	9.33	8.76	8.02	11.36	11.02	10.38	4.52	2.89	1.94	4.99	3.10	2.31
10	**1000**	2.78	18.01	6.05	5.54	5.30	7.28	6.84	6.21	3.07	1.61	1.30	3.29	1.80	1.63
20	**250**	3.85	24.42	5.38	6.19	10.62	6.42	6.64	5.51	4.30	3.76	4.75	6.64	4.51	2.96
20	**500**	2.59	17.48	3.28	3.25	3.44	4.09	3.99	3.52	2.84	1.81	1.28	4.28	2.31	1.86
20	**1000**	1.81	11.72	2.09	1.94	1.96	2.78	2.66	2.46	1.86	0.92	0.76	2.65	1.31	1.19

Notes: The DGP is given by $y_{it} = (1-\phi)\mu_i + \phi y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m_i + 2, ..., T$, with $y_{i,-m_i} = \kappa_i \mu_i + \upsilon_i$, where $\kappa_i \sim IIDU(0.5, 1.5)$ and $\upsilon_i \sim IIDN(\mu_\upsilon, 1)$ measure the extent to which starting values deviate from the long-run values $\mu_i = (\alpha + w_i)/(1-\phi)$, and $w_i \sim IIDN(0, \sigma_w^2)$. We set $\alpha = 1$, and σ_w^2 is set to ensure $V(\alpha_i) = 1$. Individual effects are generated to be cross-sectionally heteroskedastic and non-normal, $u_{it} = (e_{it} - 2)\sigma_{ia}/2$ for $t \leq [T/2]$, and $u_{it} = (e_{it} - 2)\sigma_{ib}/2$ for $t > [T/2]$, with $\sigma_{ia}^2 \sim IIDU(0.25, 0.75)$, $\sigma_{ib}^2 \sim IIDU(1, 2)$, $e_{it} \sim IID\chi^2(2)$, and $[T/2]$ is the integer part of $T/2$. The BMM estimator is given by (16). Anderson and Hsiao (AH) IV estimator is given by (S.33). Moment conditions employed in the first-difference GMM methods (Arellano and Bond) are "DIF1" and "DIF2", given by (70) and (71), respectively. Moment conditions employed in the system-GMM methods (Blundell and Bond) are "SYS1" given by (70) and (72), and "SYS2" given by (71) and (72). We implement one-step (1Step), two-step (2Step) and continuous updating (CU) GMM estimators, based on the each set of the moment conditions. Subsection 6.2 provides the full description of individual estimation methods.

Table 1b: Size and Power of tests for ϕ in the case of Experiments 1

$$\phi = 0.8,\ \mu_v = 0$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n			**Size (5% level, ×100**, $H_0: \phi = 0.8$)																			
3	**250**	5.4	6.5	7.7	13.7	11.2	15.7	19.2	7.7	13.7	11.2	15.7	19.2	13.9	23.9	12.6	23.1	18.9	13.9	23.9	12.6	23.1	18.9
3	**500**	5.0	6.4	7.7	12.3	10.1	13.1	14.1	7.7	12.3	10.1	13.1	14.1	11.4	16.7	9.1	13.0	10.0	11.4	16.7	9.1	13.0	10.0
3	**1000**	5.3	4.5	5.9	7.0	5.3	6.9	6.3	5.9	7.0	5.3	6.9	6.3	7.9	13.0	7.5	9.2	7.3	7.9	13.0	7.5	9.2	7.3
5	**250**	4.9	6.3	15.5	22.2	11.7	19.5	15.7	9.6	15.8	9.4	14.0	13.3	22.5	37.0	12.6	25.3	13.8	24.2	32.8	13.2	22.7	13.2
5	**500**	4.7	6.8	10.6	13.0	8.1	12.9	9.4	7.0	9.4	5.9	9.6	7.6	15.8	24.0	7.5	12.6	8.1	16.4	20.7	7.9	12.4	7.4
5	**1000**	4.4	4.3	7.7	8.1	6.6	8.1	6.3	5.2	6.3	5.3	5.9	5.4	10.5	14.5	6.9	10.0	6.5	11.0	12.4	6.6	8.5	6.4
10	**250**	3.9	5.1	21.6	36.4	11.0	32.9	29.2	6.7	16.6	6.8	14.2	12.8	33.5	60.7	11.7	41.7	27.0	44.2	50.8	19.2	31.4	14.8
10	**500**	5.4	5.4	14.5	20.0	10.9	17.5	14.7	6.2	10.8	7.0	8.6	8.0	21.7	42.1	8.0	23.7	15.4	28.0	32.0	10.4	17.8	8.8
10	**1000**	4.3	4.8	10.3	11.8	7.6	10.7	9.2	5.0	7.0	6.1	6.8	6.9	13.8	23.1	6.6	14.2	9.4	17.5	16.2	6.6	13.2	8.3
20	**250**	3.7	4.9	28.6	76.7	0.0	85.1	91.5	5.8	16.6	6.5	12.6	16.2	37.5	91.2	0.8	91.0	95.8	69.5	71.2	27.3	48.5	33.4
20	**500**	3.7	5.9	16.8	40.8	5.4	45.1	51.0	5.2	9.5	5.2	8.5	9.7	23.8	64.5	0.9	48.4	53.8	48.9	39.6	12.1	29.2	19.0
20	**1000**	4.7	4.4	11.1	20.4	6.6	21.3	22.8	5.7	7.6	5.9	6.6	7.7	15.2	35.2	2.4	23.9	24.5	30.9	20.7	6.1	16.1	11.9
				Power (5% level, ×100, $H_1: \phi = 0.9$)																			
3	**250**	29.4	8.3	11.0	11.0	14.0	18.2	21.2	11.0	16.7	14.0	18.2	21.2	1.1	23.4	14.2	34.3	34.4	1.1	23.4	14.2	34.3	34.4
3	**500**	41.0	8.4	11.9	11.9	12.2	14.7	16.1	11.9	15.3	12.2	14.7	16.1	0.8	30.8	22.3	41.6	43.5	0.8	30.8	22.3	41.6	43.5
3	**1000**	58.9	6.3	12.0	12.0	9.5	9.6	8.8	12.0	12.7	9.5	9.6	8.8	11.1	49.2	44.2	59.5	60.2	11.1	49.2	44.2	59.5	60.2
5	**250**	38.8	9.0	28.9	28.9	20.5	22.5	19.1	18.5	23.2	15.4	17.0	16.8	4.4	54.9	27.3	73.4	63.9	4.1	53.1	29.2	72.9	63.9
5	**500**	52.3	10.3	25.2	25.2	17.7	15.9	12.6	16.7	18.5	15.2	12.3	11.2	18.2	75.9	57.0	86.5	83.6	16.9	75.4	59.1	86.7	83.6
5	**1000**	75.1	9.2	25.3	25.3	20.9	14.7	12.6	19.4	18.8	17.3	13.3	12.3	58.6	94.3	89.2	97.2	97.1	57.4	94.7	90.1	97.4	96.8
10	**250**	50.1	10.0	70.3	70.3	47.0	40.6	34.3	28.9	39.4	24.7	19.4	18.5	23.2	89.6	38.4	99.1	97.2	14.0	79.5	45.4	97.6	93.0
10	**500**	70.0	11.2	71.4	71.4	57.7	40.8	34.6	38.1	41.4	36.0	22.5	24.1	58.8	99.3	90.0	100.0	100.0	50.0	97.8	88.4	99.9	99.6
10	**1000**	89.7	12.8	78.0	78.0	71.8	56.6	50.4	49.8	53.7	51.6	37.7	38.9	93.6	100.0	100.0	100.0	100.0	92.2	100.0	100.0	100.0	100.0
20	**250**	72.1	9.9	99.8	99.8	3.3	89.4	93.1	75.1	83.2	71.9	66.2	72.1	80.4	99.9	2.1	99.8	99.5	35.7	95.1	66.3	100.0	99.8
20	**500**	92.8	12.5	100.0	100.0	98.8	98.7	98.7	92.0	95.3	92.7	88.7	90.8	99.3	100.0	99.1	100.0	99.8	89.2	100.0	99.3	100.0	100.0
20	**1000**	99.6	17.8	100.0	100.0	100.0	100.0	100.0	99.5	99.9	99.8	99.5	99.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See notes to Table 1a. Two-step Arrelano and Bond's first-difference GMM and Blundell and Bond's system GMM estimators with the suffix "w" use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators with the suffix "nw" use Newey and Windmeijer (2009)'s standard errors.

Table 2a: Bias and RMSE of alternative estimates of ϕ for Experiment 2

$$\phi = 0.8,\ \mu_v = 1$$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**250**	1.11	156.20	-13.06	-10.19	-3.47	-13.06	-10.19	-3.47	10.06	13.41	13.19	10.06	13.41	13.19
3	**500**	0.33	38.60	-7.33	-5.01	-0.62	-7.33	-5.01	-0.62	9.06	11.67	9.23	9.06	11.67	9.23
3	**1000**	0.26	-232.04	-3.46	-2.19	0.21	-3.46	-2.19	0.21	8.75	11.00	6.87	8.75	11.00	6.87
5	**250**	1.07	90.95	-11.63	-7.23	0.64	-8.45	-5.04	0.84	9.11	9.78	6.46	9.50	9.41	6.15
5	**500**	0.45	18.44	-5.91	-3.03	0.98	-4.10	-2.16	0.76	7.78	8.69	3.11	8.00	8.12	2.92
5	**1000**	0.23	7.73	-3.10	-1.59	0.49	-2.09	-1.12	0.34	7.07	8.03	1.05	7.19	7.39	0.99
10	**250**	0.77	9.14	-7.34	-4.01	2.04	-4.36	-2.25	1.05	7.23	6.62	0.23	8.45	7.37	-0.25
10	**500**	0.29	3.28	-4.16	-2.03	1.00	-2.43	-1.30	0.30	5.82	4.86	-0.19	6.58	5.48	-0.74
10	**1000**	0.11	1.41	-2.10	-0.92	0.60	-1.11	-0.61	0.20	4.89	3.76	-0.22	5.37	4.35	-0.57
20	**250**	0.40	3.15	-3.83	-2.55	1.34	-1.90	-1.23	0.35	4.43	3.82	0.35	6.93	5.34	-1.48
20	**500**	0.14	1.95	-2.08	-1.12	0.88	-0.98	-0.62	0.17	3.25	2.33	-0.01	4.94	3.40	-0.98
20	**1000**	0.05	0.55	-1.10	-0.45	0.51	-0.46	-0.31	0.11	2.36	1.50	-0.08	3.49	2.35	-0.62
		RMSE(x100)													
3	**250**	10.49	6511.29	39.34	41.85	49.92	39.34	41.85	49.92	17.37	19.57	27.22	17.37	19.57	27.22
3	**500**	6.52	2563.47	25.22	25.57	28.23	25.22	25.57	28.23	13.89	15.02	18.37	13.89	15.02	18.37
3	**1000**	4.66	6152.14	16.80	16.04	15.70	16.80	16.04	15.70	11.31	12.91	13.37	11.31	12.91	13.37
5	**250**	8.12	2834.50	20.56	18.76	18.75	19.58	18.02	19.80	11.46	12.34	14.75	11.76	12.25	14.89
5	**500**	5.41	366.99	13.08	11.56	11.58	12.52	11.30	11.20	9.80	10.74	9.83	9.95	10.30	10.14
5	**1000**	3.76	49.65	8.55	7.76	7.77	8.34	7.74	7.72	8.35	9.45	5.88	8.44	8.85	6.00
10	**250**	5.98	60.13	10.15	8.34	8.53	9.32	7.93	7.60	8.21	7.84	4.15	9.24	8.76	5.79
10	**500**	4.06	29.05	6.51	5.41	5.22	6.11	5.31	5.02	6.71	5.82	2.02	7.34	6.57	2.47
10	**1000**	2.70	18.82	4.15	3.53	3.52	4.04	3.63	3.56	5.58	4.43	1.35	6.00	5.12	1.73
20	**250**	3.78	24.93	4.78	5.25	8.69	4.38	4.11	3.84	5.10	4.53	5.23	7.39	6.10	3.09
20	**500**	2.55	17.81	2.91	2.68	2.94	2.89	2.66	2.55	3.75	2.77	1.34	5.31	3.97	1.95
20	**1000**	1.78	11.93	1.84	1.60	1.69	1.99	1.82	1.77	2.78	1.80	0.79	3.82	2.78	1.28

Notes: See notes to Tables 1a.

Table 2b: Size and Power of tests for ϕ in the case of Experiments 2

$$\phi = 0.8,\ \mu_v = 1$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n	**Size (5% level, ×100, $H_0:\phi=0.8$)**																					
3	**250**	5.1	6.5	4.4	5.7	3.8	6.0	6.2	4.4	5.7	3.8	6.0	6.2	36.4	53.7	39.6	54.1	49.7	36.4	53.7	39.6	54.1	49.7
3	**500**	4.6	5.7	5.0	4.9	3.7	4.5	3.9	5.0	4.9	3.7	4.5	3.9	46.1	64.4	47.2	53.3	47.1	46.1	64.4	47.2	53.3	47.1
3	**1000**	5.2	4.8	5.6	4.4	4.1	4.4	4.3	5.6	4.4	4.1	4.4	4.3	53.1	73.9	51.6	49.4	42.4	53.1	73.9	51.6	49.4	42.4
5	**250**	4.5	6.1	7.8	11.6	6.0	9.7	9.5	5.1	8.2	5.3	8.0	7.4	52.7	71.6	40.0	51.1	38.4	53.7	67.5	42.8	46.9	36.1
5	**500**	4.3	6.8	6.7	7.5	5.4	7.6	7.1	5.1	5.8	5.0	5.9	6.2	55.2	72.3	41.0	36.5	25.5	56.2	68.4	46.8	35.4	24.6
5	**1000**	4.5	4.5	5.5	5.8	4.8	6.1	6.2	5.0	5.1	5.0	5.6	5.5	62.1	80.1	50.7	27.1	16.2	62.9	75.9	56.1	25.9	15.6
10	**250**	3.8	5.2	15.6	22.6	6.9	24.9	26.3	4.8	8.8	5.4	9.1	10.9	60.2	85.6	24.4	49.8	32.5	69.6	80.8	46.3	43.6	22.2
10	**500**	5.3	5.5	11.1	13.6	7.2	13.9	13.9	5.5	7.2	5.6	7.5	8.1	60.5	82.8	36.4	33.1	18.4	68.4	77.7	48.6	27.6	13.8
10	**1000**	4.2	4.7	7.6	8.2	6.2	9.3	9.2	4.9	6.3	5.9	6.7	6.8	63.5	82.0	51.2	25.5	13.1	69.6	79.4	56.5	24.6	12.9
20	**250**	3.6	4.9	26.3	74.4	0.0	83.5	91.5	5.3	13.0	5.8	12.0	16.5	56.4	96.5	2.2	93.1	95.5	83.0	87.4	48.7	55.9	37.8
20	**500**	4.0	5.9	15.9	37.1	4.1	43.8	54.4	5.3	8.6	5.1	8.6	9.8	51.7	87.6	5.4	57.6	56.7	78.2	80.1	43.8	39.7	23.8
20	**1000**	4.7	4.3	10.1	17.2	5.6	20.7	24.0	5.7	6.7	5.9	6.3	7.4	49.8	80.4	16.8	36.3	28.0	72.0	74.1	46.1	29.3	16.9
		Power (5% level, ×100, $H_1:\phi=0.9$)																					
3	**250**	29.0	7.9	9.7	9.7	7.0	7.8	8.0	9.7	9.6	7.0	7.8	8.0	6.4	29.9	16.8	49.3	48.1	6.4	29.9	16.8	49.3	48.1
3	**500**	41.3	7.3	12.6	12.6	9.0	7.7	7.1	12.6	10.9	9.0	7.7	7.1	7.5	34.7	23.3	61.0	56.3	7.5	34.7	23.3	61.0	56.3
3	**1000**	60.2	6.4	15.9	15.9	12.6	10.2	10.5	15.9	13.9	12.6	10.2	10.5	9.0	41.3	25.6	78.4	66.7	9.0	41.3	25.6	78.4	66.7
5	**250**	39.2	9.1	26.8	26.8	17.4	15.4	14.5	18.0	16.9	13.9	12.3	12.7	10.5	54.0	23.2	87.0	75.4	10.5	47.7	21.1	85.0	73.3
5	**500**	54.2	10.2	29.6	29.6	21.6	15.9	15.2	22.9	20.3	19.0	14.9	14.3	11.8	58.4	28.8	94.6	88.6	11.8	53.2	28.5	93.6	86.3
5	**1000**	76.6	8.9	37.5	37.5	33.2	25.2	25.9	32.9	30.8	31.7	24.3	25.9	19.0	61.3	28.2	98.8	96.8	18.5	58.5	33.9	98.6	95.6
10	**250**	51.3	9.8	75.3	75.3	50.5	43.4	45.0	43.6	45.6	37.9	29.9	36.7	13.3	77.0	15.0	99.6	97.9	9.4	63.0	21.8	99.0	95.3
10	**500**	72.1	11.1	84.3	84.3	71.6	59.7	60.4	64.2	65.9	63.6	55.1	57.8	34.4	89.2	47.6	100.0	100.0	26.3	76.5	43.0	100.0	99.8
10	**1000**	91.0	12.6	94.1	94.1	91.7	85.4	84.3	84.5	86.9	86.7	82.4	82.2	67.6	98.1	82.0	100.0	100.0	61.5	91.5	72.2	100.0	100.0
20	**250**	73.0	9.9	99.9	99.9	3.3	90.7	94.0	91.0	93.8	88.2	87.6	90.8	72.6	99.6	3.1	99.6	99.3	29.8	85.7	40.9	100.0	99.9
20	**500**	93.2	12.3	100.0	100.0	99.8	99.6	99.6	99.4	99.7	99.4	99.0	99.2	98.0	100.0	94.5	100.0	99.9	83.5	99.3	90.3	100.0	100.0
20	**1000**	99.6	17.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	100.0

Notes: See notes to Tables 1a and 1b.

A Appendix

This appendix is organized as follows. Section A.1 derives $\bar{B}_3$ given by (34) and (35), as well as a number of results used in Section 4. Section A.2 provides lemmas for the univariate case. Section A.3 provides lemmas for the multivariate case. Additional propositions and proofs are given in Section A.4.

A.1 Derivation of $\bar{B}_3$

Recall that

$$\bar{B}_T = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\left(B_{iT}\right), \tag{A.1}$$

where

$$B_{iT} = \frac{1}{T-2} \left(\sum_{t=2}^{T-1} \Delta y_{i,t-1}^2 + \Delta y_{it}^2 + 2\Delta u_{it} \Delta y_{i,t-1} \right),$$

and $B_{i3} = \Delta y_{i1}^2 + \Delta y_{i2}^2 + 2\Delta u_{i2}\Delta y_{i1}$. But $E\left(\Delta u_{i2}\Delta y_{i1}\right) = -\sigma_{i1}^2$, and

$$\begin{aligned} E\left(\Delta y_{i2}^2\right) &= E\left(\phi^2 \Delta y_{i1}^2 + \Delta u_{i2}^2 + 2\phi \Delta u_{i2} \Delta y_{i1}\right) \\ &= \phi^2 E\left(\Delta y_{i1}^2\right) + \left(\sigma_{i2}^2 + \sigma_{i1}^2\right) - 2\phi\sigma_{i1}^2 \\ &= \phi^2 E\left(\Delta y_{i1}^2\right) + \left(1-2\phi\right)\sigma_{i1}^2 + \sigma_{i2}^2, \end{aligned} \tag{A.2}$$

Hence

$$\begin{aligned} E\left(B_{i3}\right) &= E\left(\Delta y_{i1}^2\right) + \phi^2 E\left(\Delta y_{i1}^2\right) + \left(1-2\phi\right)\sigma_{i1}^2 + \sigma_{i2}^2 - 2\sigma_{i1}^2 \\ &= \left(1+\phi^2\right) E\left(\Delta y_{i1}^2\right) + \left(\sigma_{i2}^2 - \sigma_{i1}^2\right) - 2\phi\sigma_{i1}^2. \end{aligned} \tag{A.3}$$

We derive $E\left(\bar{B}_3\right)$ in terms of $\bar{\sigma}_t^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{it}^2$, for $t=1,2$, and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(y_{i0}-\mu_i\right)^2$, and $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right]$. Note that $\Delta y_{i1} = u_{i1} - (1-\phi)\left(y_{i0}-\mu_i\right)$. Hence

$$E\left(\Delta y_{i1}^2\right) = \sigma_{i1}^2 + (1-\phi)^2 E\left(y_{i0}-\mu_i\right)^2 - 2(1-\phi) E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right].$$

Using this result in (A.3), we have

$$\begin{aligned} E\left(B_{i3}\right) &= \left(\sigma_{i2}^2 - \sigma_{i1}^2\right) + (1-\phi)^2 \sigma_{i1}^2 \\ &\quad + \left(1+\phi^2\right)\left\{(1-\phi)^2 E\left(y_{i0}-\mu_i\right)^2 - 2(1-\phi)E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right]\right\}, \end{aligned}$$

which in view of (A.1), yields

$$\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2 + (1-\phi)^2 \bar{\sigma}_1^2 + \left(1+\phi^2\right)(1-\phi)\psi_0, \tag{A.4}$$

where

$$\psi_0 = (1-\phi) \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left(y_{i0}-\mu_i\right)^2 - 2 \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right],$$

as required.

A.2 Lemmas for the univariate case

Lemma A.1 *Suppose y_{it}, for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m + 2, ..., 1, 2, ..., T$, are generated by (1) with starting values $y_{i,-m}$. Let Assumptions 1-3 hold. Consider*

$$\bar{Q}_{nT} = \frac{1}{n}\sum_{i=1}^{n} Q_{iT}, \text{ and } \bar{B}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{iT} + Q_{iT}^{+} + 2H_{iT}\right),$$

where $Q_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta y_{i,t-1}^2$, $Q_{iT}^{+} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta y_{it}^2$, and $H_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta u_{it}\Delta y_{i,t-1}$. Suppose that T is fixed. Then, we have

$$\bar{Q}_{nT} = E\left(\bar{Q}_{nT}\right) + O_p\left(n^{-1/2}\right), \tag{A.5}$$

$$\bar{B}_{nT} = E\left(\bar{B}_{nT}\right) + O_p\left(n^{-1/2}\right). \tag{A.6}$$

Proof. Under Assumptions 1-3, the fourth moments of u_{it} and b_i are bounded, and hence, using Loève's inequality,[18] for each i the fourth moment of Δy_{it} :

$$\Delta y_{it} = \phi^{t-1}\left[b_i + u_{i1} - (1-\phi)\sum_{\ell=0}^{m_i-1}\phi^{\ell}u_{i,-\ell}\right] + \sum_{\ell=0}^{t-2}\phi^{\ell}\Delta u_{i,t-\ell},$$

is also bounded, for all values of $|\phi| \leq 1$ and $m_i \geq 0$. Since T is fixed, it follows that the second moment of $Q_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta y_{i,t-1}^2$ must be bounded, and hence there must exist K such that $E\left[Q_{iT} - E\left(Q_{iT}\right)\right]^2 < K$. Consider next the cross-sectional average of $Q_{iT} - E\left(Q_{iT}\right)$. We have $E\left[Q_{iT} - E\left(Q_{iT}\right)\right] = 0$ by construction, and also $Q_{iT} - E\left(Q_{iT}\right)$ is independently distributed across i, since, under Assumptions 1-3, Δy_{it} is independently distributed across i. Hence,

$$Var\left\{n^{-1}\sum_{i=1}^{n}\left[Q_{iT} - E\left(Q_{iT}\right)\right]\right\} \leq n^{-2}\sum_{i=1}^{n}E\left[Q_{iT} - E\left(Q_{iT}\right)\right]^2 < \frac{K}{n},$$

and therefore $n^{-1}\sum_{i=1}^{n}Q_{iT} - n^{-1}\sum_{i=1}^{n}E\left(Q_{iT}\right) = O_p\left(n^{-1/2}\right)$. This completes the proof of (A.5).

Result (A.6) is established similarly. Note that

$$\bar{B}_{nT} = \frac{1}{n}\sum_{i=1}^{n}Q_{iT} + \frac{1}{n}\sum_{i=1}^{n}Q_{iT}^{+} + 2\frac{1}{n}\sum_{i=1}^{n}H_{iT} = \bar{Q}_{nT} + \bar{Q}_{nT}^{+} + 2\bar{H}_{nT}.$$

The order of $\bar{Q}_{nT} - E\left(\bar{Q}_{nT}\right)$ is given by (A.5). Using the same arguments as in the proof of (A.5), we have

$$\bar{Q}_{nT}^{+} - E\left(\bar{Q}_{nT}^{+}\right) = O_p\left(n^{-1/2}\right), \text{ and } \bar{H}_{nT} - E\left(\bar{H}_{nT}\right) = O_p\left(n^{-1/2}\right).$$

Hence, $\bar{B}_{nT} - E\left(\bar{B}_{nT}\right) = \bar{Q}_{nT} - E\left(\bar{Q}_{nT}\right) + \bar{Q}_{nT}^{+} - E\left(\bar{Q}_{nT}^{+}\right) + 2\left[\bar{H}_{nT} - E\left(\bar{H}_{nT}\right)\right] = O_p\left(n^{-1/2}\right)$, and result (A.6) follows. This completes the proof. ■

Lemma A.2 *Suppose y_{it}, for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T$, are generated by (1) with starting values $y_{i,-m_i}$. Let Assumptions 1-3 hold. Consider*

$$\bar{V}_{nT} = \frac{1}{n}\sum_{i=1}^{n}V_{iT},$$

[18] See equation (9.62) of Davidson (1994).

where $V_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right)$. *Suppose that* T *is fixed. Then, we have*

$$\bar{V}_{nT} = O_p\left(n^{-1/2}\right). \tag{A.7}$$

If, in addition, $S_T = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(V_{iT}^2\right)$*, and* T *is fixed as* $n \to \infty$*, then*

$$\sqrt{n}\bar{V}_{nT} \to_d N\left(0, S_T\right). \tag{A.8}$$

Proof. Under Assumptions 2 and 3, V_{iT} is independently distributed of V_{jT} for all $i \neq j$, $i, j = 1, 2, ..., n$. In addition, (using (13))

$$E\left(V_{iT}\right) = \frac{1}{T-2}\sum_{t=2}^{T-1} E\left(\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right) = 0. \tag{A.9}$$

Also, by Assumptions 2 and 3, $\sup_{i,t} E\left|u_{it}\right|^{4+\epsilon} < K$, and $\sup_i E\left|b_i\right|^{4+\epsilon} < K$, for some $\epsilon > 0$, and hence, using Loève's inequality,[19] we have $\sup_{i,t} E\left|\Delta y_{it}\right|^{4+\epsilon} < K$. Using Loève's inequality again, we have

$$E\left|\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right|^{2+\epsilon/2} \leq K\left(E\left|\Delta u_{it}\Delta y_{i,t-1}\right|^{2+\epsilon/2} + E\left|\Delta u_{it}^2\right|^{2+\epsilon/2} + E\left|\Delta u_{i,t+1}\Delta y_{it}\right|^{2+\epsilon/2}\right).$$

But $\sup_{it} E\left|\Delta u_{it}^2\right|^{2+\epsilon/2} = \sup_{it} E\left|\Delta u_{it}\right|^{4+\epsilon} < K$, as well as $\sup_{i,t} E\left|\Delta u_{it}\Delta y_{i,t-1}\right|^{2+\epsilon/2} < K$, and $\sup_{i,t} E\left|\Delta u_{i,t+1}\Delta y_{it}\right|^{2+\epsilon/2} < K$. Hence, $\sup_{it} E\left|\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right|^{2+\epsilon/2} < K$, and using Loève's inequality again, we have

$$\sup_i E\left(\left|V_{iT}\right|^{2+\epsilon/2}\right) < K. \tag{A.10}$$

It follows also that $\sup_i E\left(V_{iT}^2\right) < K$, and given that V_{iT} is independently distributed over i, we have

$$E\left(\bar{V}_{nT}^2\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(V_{iT}V_{jT}\right) = n^{-2}\sum_{i=1}^{n} E\left(V_{iT}^2\right) < \frac{K}{n},$$

and result (A.7) follows. To establish (A.8), we note that (A.10) holds, and therefore the Lyapunov condition holds (see Theorem 23.12 of Davidson, 1994). Hence, noting also that $n^{-1}\sum_{i=1}^{n} E\left(V_{iT}^2\right) \to S_T$ by assumption, we obtain $\sqrt{n}\bar{V}_{nT} \to_d N\left(0, S_T\right)$, as required. ■

A.3 Lemmas for multivariate case

Lemmas A.3 and A.4 below are direct extensions of Lemmas A.1 and A.2, respectively, to the multivariate case.

Lemma A.3 *Suppose* $\mathbf{z}_{it}$*, for* $i = 1, 2, ..., n$*, and* $t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T$*, are generated by (46) with starting values* $\mathbf{z}_{i,-m_i}$*. Let Assumptions 4-6 hold. Consider*

$$\bar{\mathbf{Q}}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Q}_{iT}, \text{ and } \bar{\mathbf{B}}_{nT} = n^{-1}\sum_{i=1}^{n}\bar{\mathbf{B}}_{iT},$$

where $\mathbf{Q}_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta\mathbf{z}_{i,t-1}\Delta\mathbf{z}_{i,t-1}'$, $\bar{\mathbf{B}}_{iT} = \left(\mathbf{I}_k \otimes \mathbf{H}_{iT}\right)\mathbf{R} + \left(\mathbf{H}_{iT} + \mathbf{Q}_{iT} + \mathbf{Q}_{iT}^{+}\right) \otimes \mathbf{I}_k$, $\mathbf{H}_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta\mathbf{u}_{it}\Delta\mathbf{z}_{i,t-1}'$, $\mathbf{Q}_{iT}^{+} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta\mathbf{z}_{it}\Delta\mathbf{z}_{it}'$, *and* $\mathbf{R}$ *is the unique* $k^2 \times k^2$ *re-*

[19] See equation (9.62) of Davidson (1994).

ordering matrix defined by $Vec(\mathbf{X}) = \mathbf{R}Vec(\mathbf{X}')$ *for any* $k \times k$ *matrix* $\mathbf{X}$. *Suppose that* T *is fixed. Then, we have*

$$\bar{\mathbf{Q}}_{nT} = E\left(\bar{\mathbf{Q}}_{nT}\right) + O_p\left(n^{-1/2}\right), \tag{A.11}$$

$$\bar{\mathbf{B}}_{nT} = E\left(\bar{\mathbf{B}}_{nT}\right) + O_p\left(n^{-1/2}\right). \tag{A.12}$$

Proof. Lemma A.3 can be established using the same arguments as in the proof of Lemma A.1. ■

Lemma A.4 *Suppose* $\mathbf{z}_{it}$, *for* $i = 1, 2, ..., n$, *and* $t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T$, *are generated by (46) with starting values* $\mathbf{z}_{i,-m_i}$. *Let Assumptions 4-6 hold. Consider*

$$\bar{\mathbf{V}}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_{iT},$$

where $\mathbf{V}_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-1} + \Delta\mathbf{u}_{it}\Delta\mathbf{u}'_{it} + \Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}'_{it}\right)$. *Suppose that* T *is fixed. Then, we have*

$$\bar{\mathbf{V}}_{nT} = O_p\left(n^{-1/2}\right). \tag{A.13}$$

If, in addition, $\mathbf{S}_T = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[Vec(\mathbf{V}_{iT})\, Vec(\mathbf{V}_{iT})'\right]$, *and* T *is fixed as* $n \to \infty$, *then*

$$\sqrt{n}Vec\left(\bar{\mathbf{V}}_{nT}\right) \to_d N\left(\mathbf{0}, \mathbf{S}_T\right). \tag{A.14}$$

Proof. Lemma A.4 can be established using the same arguments as in the proof of Lemma A.2. ■

A.4 Propositions and Proofs

Theorems 1 and 2 are established in the main text. This section presents propositions for the consistency of $\hat{\Sigma}_{nT}$.

Proposition 1 *Suppose conditions of Theorem 1 hold, and consider* $\hat{\Sigma}_{nT}$ *defined by (36), namely*

$$\hat{\Sigma}_{nT} = \hat{\bar{B}}_{nT}^{-2}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{V}_{i,nT}^2\right),$$

where $\hat{\bar{B}}_{nT} = n^{-1}\sum_{i=1}^{n}\left(Q_{iT} + Q_{iT}^{+} + 2\hat{H}_{i,nT}\right)$, $\hat{H}_{i,nT} = (T-2)^{-1}\sum_{t=2}^{T-1}\Delta\hat{u}_{it}\Delta y_{i,t-1}$, $\Delta\hat{u}_{it} = \Delta y_{it} - \hat{\phi}_{nT}\Delta y_{i,t-1}$,

$$\hat{V}_{i,nT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta\hat{u}_{it}\Delta y_{i,t-1} + \Delta\hat{u}_{it}^2 + \Delta\hat{u}_{i,t+1}\Delta y_{it}\right),$$

and $\hat{\phi}_{nT}$ *is the* $\sqrt{n}$*-consistent BMM estimator given by (16). Let* T *be fixed as* $n \to \infty$. *Then,*

$$\hat{\Sigma}_{nT} \to_p \Sigma_T, \tag{A.15}$$

where Σ_T *is defined in (32)*

Proof. Using Theorem 1, we have $\hat{\phi}_{nT} = \phi_0 + O_p\left(n^{-1/2}\right)$, and therefore $\Delta\hat{u}_{it} = \Delta y_{it} - \hat{\phi}_{nT}\Delta y_{i,t-1}$ is consistent, namely $\Delta\hat{u}_{it} - \Delta u_{it} = \Delta y_{it} - \left(\hat{\phi}_{nT} - \phi_0\right)\Delta y_{i,t-1} = O_p\left(n^{-1/2}\right)$. This implies $\hat{H}_{i,nT}$ is consistent, which in turn implies $\hat{\bar{B}}_{nT} - \bar{B}_{nT} \rightarrow_p 0$. But, using result (A.6) of Lemma A.1, we have $\bar{B}_{nT} \rightarrow_p E\left(\bar{B}_{nT}\right)$, and $E\left(\bar{B}_{nT}\right) \rightarrow B_T$. Therefore $\hat{\bar{B}}_{nT} \rightarrow_p \bar{B}_T$. Since $\bar{B}_T > 0$ by assumption, it follows that

$$\hat{\bar{B}}_{nT}^{-2} \rightarrow_p \bar{B}_T^{-2}. \tag{A.16}$$

Next consider $n^{-1}\sum_{i=1}^{n}\hat{V}_{i,nT}^2$, and note that

$$\hat{V}_{i,nT}^2 = \left[\left(\hat{V}_{i,nT} - V_{iT}\right) + V_{iT}\right]^2 = \left(\hat{V}_{i,nT} - V_{iT}\right)^2 + 2\left(\hat{V}_{i,nT} - V_{iT}\right)V_{iT} + V_{iT}^2,$$

where $V_{iT} = (T-2)^{-1}\sum_{t=2}^{T-1}\left(\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right)$. Using $\Delta\hat{u}_{n,it} - \Delta u_{n,it} = O_p\left(n^{-1/2}\right)$, we have $\hat{V}_{i,nT} - V_{iT} = O_p\left(n^{-1/2}\right)$. Noting also that $V_{iT} = O_p(1)$, we then have

$$n^{-1}\sum_{i=1}^{n}\left(\hat{V}_{i,nT} - V_{iT}\right)^2 \rightarrow_p 0, \text{ and } n^{-1}\sum_{i=1}^{n}\left(\hat{V}_{i,nT} - V_{iT}\right)V_{iT} \rightarrow_p 0. \tag{A.17}$$

Finally, to obtain the limiting property of $n^{-1}\sum_{i=1}^{n}V_{iT}^2$, note that by assumption V_{iT} is independently distributed over i. Also, as established in (A.10), we have $\sup_i E\left|V_{iT}\right|^{2+\epsilon/2} < K$ for some $\epsilon > 0$. It follows that $n^{-1}\sum_{i=1}^{n}\left[V_{iT}^2 - E\left(V_{iT}^2\right)\right] \rightarrow_p 0$, and therefore (noting that $n^{-1}\sum_{i=1}^{n}E\left(V_{iT}^2\right) \rightarrow S_T$ by assumption) we have

$$n^{-1}\sum_{i=1}^{n}V_{iT}^2 \rightarrow_p S_T. \tag{A.18}$$

Result (A.15) now follows from (A.16), (A.17), and (A.18). ■

References

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

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

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

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies 58*, 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51.

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

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*, 795–837.

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

Bruno, G. S. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters 87*, 361–366.

Bun, M. (2003). Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews 22*, 29–58.

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

Bun, M. J. and M. A. Carree (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters 92*, 220–227.

Bun, M. J. and F. Kleibergen (2013). Identification and inference in moments based analysis of linear dynamic panel data models. UvA Econometrics Discussion Paper No. 2013/07.

Bun, M. J. G. and J. Kiviet (2003). On the diminishing returns of higher order terms in asymptotic expansions of bias. *Economic Letters 19*, 145–152.

Choi, C., N. C. Mark, and D. Sul (2010). Bias reduction in dynamic panel data models by common recursive mean adjustment. *Oxford Bulletin of Economics and Statistics 72*, 567–599.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

Chudik, A., M. H. Pesaran, and J.-C. Yang (2016). Half-panel jackknife fixed effects estimation of panels with weakly exogenous regressors. Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute Working Paper No. 281.

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

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*(3), 991–1030.

Everaert, G. and L. Ponzi (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control 31*, 1160–1184.

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

Hahn, J. and H. Moon (2006). Reducing bias of MLE in a dynamic panel model. *Econometric Theory 22*, 499–512.

Han, C. and P. C. B. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory 26*, 119–151.

Han, C., P. C. B. Phillips, and D. Sul (2014). X-Differencing and dynamic panel model estimation. *Econometric Theory 30*, 201–251.

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

Hayakawa, K. (2012). GMM estimation of short dynamic panel data models with interactive fixed effects. *Journal of the Japan Statistical Society 42*, 109–123.

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

Hayakawaa, K. and S. Nagata (2016). On the behaviour of the GMM estimator in persistent dynamic panel data models with unrestricted initial conditions. *Computational Statistics and Data Analysis 100*, 265–303.

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*, 107–150.

Juodis, A. (2013). A note on bias-corrected estimation in dynamic panel data models. *Economics Letters 118*, 435–438.

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

Kiviet, J. F. (1999). Expectations of expansions for estimators in a dynamic panel data model; some results for weakly-exogenous regressors. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panel Data and Limited Dependent Variables*. Cambridge University Press, Cambridge.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*, 687–719.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlations. *Journal of Econometrics 161*, 182–202.

Phillips, P. C. and D. Sul (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics 137*, 162–188.

Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal 6*, 217–259.

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

S Online Supplement to "A Bias-Corrected Method of Moments Approach to Estimation of Dynamic Short-T Panels" by A. Chudik and M. H. Pesaran

This supplement provides additional theoretical results and further Monte Carlo (MC) findings. Section S.1 considers the implications of relaxing the assumptions that errors and initial values are cross-sectionally independent. Section S.2 derives conditional model for y_{it} when $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$ is given by a panel VAR model. Section S.3 derives conditions under which the autoregressive parameter of interest, ϕ, is identified for the set of alternative GMM moment conditions considered in the paper. To simplify the derivations we focus on the case where $T = 3$, and the order of the underlying AR process is one. Section S.4 extends the BMM procedure to unbalanced panels with time effects. Section S.5 discusses consistent estimation of $\bar{\mathbf{\Omega}}_t$, for $t = 1, 2, ..., T$. Section S.6 derives consistent estimator of asymptotic variance of Anderson and Hsiao (AH) estimator. Section S.7 reports MC results for $\phi = 0.4$, as well as for large values of n, namely $n = 2,000, 5,000, 10,000$, and $T = 3, 5, 10, 20$. It also considers properties of BMM and AH estimators when both n and T are large, and when $\phi = 1$. Section S.8 presents rejection frequency plots for selected MC experiments.

S.1 Relaxing cross-sectional independence of errors and initial values

Assumption 2 requires errors to be cross-sectionally independent. This assumption can be relaxed as follows.

ASSUMPTION S1 *(Cross-sectionally correlated errors) For each $i = 1, 2, ..., n$, the process $\{u_{it}, t = -m_i + 1, -m_i + 2, ..., 1, 2, ..., T\}$ is distributed with mean 0, $E\left(u_{it}^2\right) = \sigma_{it}^2$, and there exist positive constants c and K such that $0 < c < \sigma_{it}^2 < K$, for all i, t, and $\sup_{i,t} E\left|u_{it}\right|^{4+\epsilon} < K$ for some $\epsilon > 0$. For each t, u_{it} is independently distributed over i. For each i, u_{it} is serially uncorrelated over t. In addition, there exist constants $0 \leq \delta_\rho < 1$ and $0 \leq \delta_\varkappa < 1$ such that the following conditions hold:*

$$\sup_{i,t} \sum_{j=1}^{n} \left|E\left(u_{it}u_{jt}\right)\right| = O\left(n^{\delta_\rho}\right), \tag{S.1}$$

and

$$\sup_{i,t} \sum_{j=1}^{n} \left|E\left(\tilde{u}_{it}^2 \tilde{u}_{jt}^2\right)\right| = O\left(n^{\delta_\varkappa}\right), \tag{S.2}$$

where $\tilde{u}_{it}^2 = u_{it}^2 - \sigma_{it}^2$.

Cross-sectional dependence of initial values could also be allowed, as postulated in the following assumption, which replaces Assumption 3.

ASSUMPTION S2 *(Initialization and individual effects in the cross-sectionally correlated case) Let* $b_i \equiv -\phi^{m_i}\left[(1-\phi)\, y_{i,-m_i} - \alpha_i\right]$. *It is assumed that* $\sup_i E\left|b_i\right|^{4+\epsilon} < K$, *for some* $\epsilon > 0$. *In addition, the following conditions hold:*

$$E\left(\Delta u_{it} b_i\right) = 0, \text{ for } i = 1, 2, ..., n, \text{ and } t = 2, 3, ..., T, \tag{S.3}$$

and there exist constants $0 \leq \delta_\varphi < 1$ *and* $0 \leq \delta_b < 1$ *such that*

$$\sup_{i,t} \sum_{j=1}^{n} \left|E\left(u_{it} u_{jt} b_i b_j\right)\right| = O\left(n^{\delta_\varphi}\right), \tag{S.4}$$

and

$$\sup_{i} \sum_{j=1}^{n} E\left(\tilde{b}_i^2 \tilde{b}_j^2\right) = O\left(n^{\delta_b}\right), \tag{S.5}$$

where $\tilde{b}_i^2 = b_i^2 - \varsigma_i^2$ *and* $\varsigma_i^2 = E\left(b_i^2\right)$.

Remark 10 *Assumption S1 allows* u_{it} *to be weakly cross-sectionally correlated such that conditions (S.1) and (S.2) of Assumption S1 hold. These conditions rule out the presence of strong unobserved common factors in errors (strong in a sense that the cross-section arithmetic average of Euclidean norm of loadings is bounded away from zero as* $n \to \infty$*),*[S1] *but it allows for more general processes than commonly used spatial processes in the literature. For example, condition (S.1) allows for the largest eigenvalues of the* $n \times n$ *covariance matrices of error vectors* $\mathbf{u}_t = \left(u_{1t}, u_{2t}, ..., u_{nt}\right)'$, *denoted as* $\mathbf{\Omega}_t = E\left(\mathbf{u}_t \mathbf{u}_t'\right)$, *to diverge as* $n \to \infty$ *but at a rate slower than* n, *whereas commonly used spatial processes in the literature typically assume that these eigenvalues are all bounded. For further discussion, see Section 2 of Pesaran and Tosetti (2011).*

Remark 11 *Assumption S1 is sufficient for*

$$n^{-1} \sum_{i=1}^{n} u_{it}^2 \to_p \bar{\sigma}_t^2, \tag{S.6}$$

as $n \to \infty$, *at any point in time* $t = -m_{\min} + 1, -m_{\min} + 2, ..., 1, 2, ..., T$, *as well as*

$$n^{-1} \sum_{i=1}^{n} u_{it} u_{it'} \to_p 0, \tag{S.7}$$

as $n \to \infty$, *for any* $t \neq t'$, $t, t' = -m_{\min} + 1, -m_{\min} + 2, ..., 1, 2, ..., T$, *where* $m_{\min} = \min\left\{m_1, m_2, ..., m_n\right\}$, *and as before* $\bar{\sigma}_t^2 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma_{it}^2$. *This is established in Lemma S1 below. Conditions (S.6) and (S.7) are required for the consistency of the BMM estimator.*

[S1] See Chudik, Pesaran, and Tosetti (2011) for definitions and discussions of the concepts of strong factors, and weak and strong cross-sectional dependence.

Lemma S1 *Under Assumption S1, we have*

$$n^{-1}\sum_{i=1}^{n}u_{it}^{2}=\bar{\sigma}_{nt}^{2}+O_{p}\left(n^{(\delta_{\varkappa}-1)/2}\right), \tag{S.8}$$

for $t=-m_{\min}+1,-m_{\min}+2,...,1,2,...,T,$ *where* $m_{\min}=\min\{m_1,m_2,...,m_n\}$, $\bar{\sigma}_{nt}^2=n^{-1}\sum_{i=1}^{n}\sigma_{it}^2$, *and*

$$n^{-1}\sum_{i=1}^{n}u_{it}u_{it'}=O_{p}\left(n^{(\delta_{\rho}-1)/2}\right), \tag{S.9}$$

for $t\neq t'$, $t,t'=-m_{\min}+1,-m_{\min}+2,...,1,2,...,T.$

Proof. Note that

$$n^{-1}\sum_{i=1}^{n}u_{it}^{2}=n^{-1}\sum_{i=1}^{n}\tilde{u}_{it}^{2}+\bar{\sigma}_{nt}^{2}, \tag{S.10}$$

where $\bar{\sigma}_{nt}^2=n^{-1}\sum_{i=1}^{n}\sigma_{it}^2$, $\tilde{u}_{it}^2=u_{it}^2-\sigma_i^2$, and $E\left(\tilde{u}_{it}^2\right)=E\left(u_{it}^2\right)-\sigma_{it}^2=0$ by construction. Taking variance of the first term on the right side of (S.10), and using condition (S.2) of Assumption S1, we have

$$\begin{aligned} Var\left(n^{-1}\sum_{i=1}^{n}\tilde{u}_{it}^{2}\right) &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\tilde{u}_{it}^{2}\tilde{u}_{jt}^{2}\right), \\ &\leq n^{-1}\sup_{i}\sum_{j=1}^{n}\left|E\left(\tilde{u}_{it}^{2}\tilde{u}_{jt}^{2}\right)\right|, \\ &= O\left(n^{\delta_{\varkappa}-1}\right), \end{aligned}$$

for $t=-m+1,-m+2,...,1,2,...,T$. Hence

$$n^{-1}\sum_{i=1}^{n}\tilde{u}_{it}^{2}=O_{p}\left(n^{(\delta_{\varkappa}-1)/2}\right),$$

for $t=-m+1,-m+2,...,1,2,...,T$, and result (S.8) follows.

Consider next $n^{-1}\sum_{i=1}^{n}u_{it}u_{it'}$ for any $t\neq t'$, $t,t'=-m_{\min}+1,-m_{\min}+2,...,1,2,...,T$. We have $E\left(u_{it}u_{it'}\right)=0$ for $t\neq t'$, and

$$\begin{aligned} Var\left(n^{-1}\sum_{i=1}^{n}u_{it}u_{it'}\right) &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(u_{it}u_{it'}u_{jt}u_{jt'}\right), \\ &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right), \end{aligned}$$

where $E\left(u_{it}u_{it'}u_{jt}u_{jt'}\right)=E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right)$ follows from the independence of u_{it} and $u_{it'}$ for $t\neq t'$. Using condition (S.1) of Assumption S1, and the boundedness of variances σ_{it}^2, we obtain

$$
\begin{aligned}
n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right) &\leq Kn^{-1}\sup_i\sum_{j=1}^{n}\left|E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right)\right|, \\
&= O\left(n^{\delta_\rho-1}\right), \qquad \text{(S.11)}
\end{aligned}
$$

for $t\neq t'$, $t,t'=-m_{\min}+1,-m_{\min}+2,...,1,2,...,T$, where by Cauchy-Schwarz inequality

$$
\begin{aligned}
\sum_{j=1}^{n}\left|E\left(u_{it}u_{jt}\right)E\left(u_{it'}u_{jt'}\right)\right| &\leq \left(\sum_{j=1}^{n}\left[E\left(u_{it}u_{jt}\right)\right]^2\right)^{1/2}\left(\sum_{j=1}^{n}\left[E\left(u_{it'}u_{jt'}\right)\right]^2\right)^{1/2} \\
&\leq K\left(\sum_{j=1}^{n}\rho_{ijt}^2\right)^{1/2}\left(\sum_{j=1}^{n}\rho_{ijt'}^2\right)^{1/2} \\
&= O\left(n^{\delta_\rho}\right),
\end{aligned}
$$

$\rho_{ijt}=E\left(u_{it}u_{jt}\right)/\left(\sigma_{it}\sigma_{jt}\right)$ is the correlation coefficient of u_{it} and u_{jt}, $\sup_{i,t}\sigma_{it}^2<K$ by Assumption S1, $\left|\rho_{ijt}\right|\leq 1$ by definition, and therefore $\sup_{it}\sum_{j=1}^{n}\rho_{ijt}^2\leq\sup_{it}\sum_{j=1}^{n}\left|\rho_{ijt}\right|$, but (due to bounded error variances) $\sup_{i,t}\sum_{j=1}^{n}\left|\rho_{ijt}\right|=O\left(n^{\delta_\rho}\right)$ is implied by condition (S.1) of Assumption S1. Result (S.9) now follows from (S.11), as required. ■

Assumptions S1-S2 can be used to replace Assumptions 2-3, respectively. It can be established that the presence of cross-sectional correlation has no consequence for the consistency of the BMM estimator, so long as $\delta=\max\left\{\delta_\rho,\delta_\varkappa,\delta_\varphi,\delta_b\right\}<1$. The inference on ϕ, however, is no longer valid in the presence of cross-sectional error dependence.

S.2 Derivation of conditional model for y_{it} when $\mathbf{z}_{it}=\left(y_{it},\mathbf{x}_{it}'\right)'$ is given by a panel VAR model

Suppose $\mathbf{z}_{it}=\left(y_{it},\mathbf{x}_{it}'\right)'$ is given by a panel VAR(1) model given by equation (46) in the paper, which we reproduce below for convenience,

$$
\mathbf{z}_{it}=\boldsymbol{\alpha}_i+\boldsymbol{\Phi}\mathbf{z}_{i,t-1}+\mathbf{u}_{it}, \qquad \text{(S.12)}
$$

for $t=-m_i+1,-m_i+2,...,1,2,...,T$, and $i=1,2,...,n$, with the starting values given by $\mathbf{z}_{i,-m}$ for $m\geq 0$, where $\boldsymbol{\alpha}_i$ is a $k\times 1$ vector of individual effects, $\boldsymbol{\Phi}$ is a $k\times k$ matrix of slope coefficients, $\mathbf{u}_{it}=\left(u_{i1t},u_{i2t},...,u_{ikt}\right)'$ is a $k\times 1$ vector of idiosyncratic errors, k is finite and does not depend on n. Individual equations for y_{it}

and $\mathbf{x}_{it}$ in (S.12) are

$$y_{it} = \alpha_{iy} + \phi_{11} y_{i,t-1} + \boldsymbol{\phi}'_{yx}\mathbf{x}_{i,t-1} + u_{y,it}, \tag{S.13}$$

$$\mathbf{x}_{it} = \boldsymbol{\alpha}_{ix} + \boldsymbol{\phi}_{xy} y_{i,t-1} + \boldsymbol{\Phi}_{xx}\mathbf{x}_{i,t-1} + \mathbf{u}_{x,it}, \tag{S.14}$$

where $\boldsymbol{\alpha}_i = (\alpha_{iy}, \boldsymbol{\alpha}'_{ix})'$, $\mathbf{u}_{it} = \left(u_{y,it}, \mathbf{u}'_{x,it}\right)'$, and $\boldsymbol{\Phi}$ is partitioned as:

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \boldsymbol{\phi}'_{yx} \\ \boldsymbol{\phi}_{xy} & \boldsymbol{\Phi}_{xx} \end{pmatrix}.$$

Let

$$E\left(\mathbf{u}_{it}\mathbf{u}'_{it}\right) = \boldsymbol{\Omega}_{it} = \begin{pmatrix} \omega_{yy,it} & \boldsymbol{\omega}'_{xy,it} \\ \boldsymbol{\omega}_{xy,it} & \boldsymbol{\Omega}_{xx,it} \end{pmatrix},$$

for all i and t. Using linear projection of $u_{y,it}$ on $\mathbf{u}_{x,it}$, we have

$$u_{y,it} = \boldsymbol{\theta}'_{it}\mathbf{u}_{x,it} + \eta_{it}, \tag{S.15}$$

where $\boldsymbol{\theta}_{it} = \boldsymbol{\Omega}^{-1}_{xx,it}\boldsymbol{\omega}_{xy,it}$, and $cov\left(\eta_{it}, \mathbf{u}_{x,it}\right) = \mathbf{0}$. Then using (S.15) and (S.14) in (S.13), we have

$$\begin{aligned} y_{it} &= \alpha_{iy} + \phi_{11} y_{i,t-1} + \boldsymbol{\phi}'_{yx}\mathbf{x}_{i,t-1} + \boldsymbol{\theta}'_{it}\left(\mathbf{x}_{it} - \boldsymbol{\alpha}_{ix} - \boldsymbol{\phi}_{xy} y_{i,t-1} - \boldsymbol{\Phi}_{xx}\mathbf{x}_{i,t-1}\right) + \eta_{it}, \\ &= \left(\alpha_{iy} - \boldsymbol{\theta}'_{it}\boldsymbol{\alpha}_{ix}\right) + \left(\phi_{11} - \boldsymbol{\theta}'_{it}\boldsymbol{\phi}_{xy}\right) y_{i,t-1} + \left(\boldsymbol{\phi}'_{yx} - \boldsymbol{\theta}'_{it}\boldsymbol{\Phi}_{xx}\right)\mathbf{x}_{i,t-1} + \eta_{it}, \end{aligned} \tag{S.16}$$

where $cov\left(\eta_{it}, \mathbf{x}_{is}\right) = \mathbf{0}$ for all i, t and s, and recall that η_{it} is serially uncorrelated. It is clear that the conditional model (46) will have homogeneous slopes only if $\boldsymbol{\theta}_{it} = \boldsymbol{\theta}$ for all i and t.

S.3 Identification of ϕ under alternative GMM conditions when $T = 3$

We consider three sets of alternative GMM conditions advanced in the literature for identification of ϕ, given by equations (43)-(45) in the paper, and reproduced below for convenience. To simplify the analysis we set $T = 3$ and note that the IV estimator proposed by Anderson and Hsiao (1981, 1982) for $T = 3$ can be written as

$$\text{AH:} \quad E\left[\Delta y_{i1}\left(\Delta y_{i3} - \phi \Delta y_{i2}\right)\right] = 0. \tag{S.17}$$

The moment conditions proposed by Arellano and Bond (1991) can be written as:

$$\text{AB:} \quad E\left[y_{i0}\left(\Delta y_{i2} - \phi\Delta y_{i1}\right)\right] = 0,\; E\left[y_{i0}\left(\Delta y_{i3} - \phi\Delta y_{i2}\right)\right] = 0, \text{ and } E\left[y_{i1}\left(\Delta y_{i3} - \phi\Delta y_{i2}\right)\right] = 0. \tag{S.18}$$

Finally, we consider the moment conditions of Arellano and Bover (1995) and Blundell and Bond (1998):

$$\text{BB:} \quad E\left[\Delta y_{i1}\left(y_{i2}-\phi y_{i1}\right)\right]=0,\ E\left[\Delta y_{i1}\left(y_{i3}-\phi y_{i2}\right)\right]=0, \text{ and } E\left[\Delta y_{i2}\left(y_{i3}-\phi y_{i2}\right)\right]=0. \tag{S.19}$$

S.3.1 Identification of ϕ under AH

In view of (S.17), consistent estimation of ϕ by the AH estimator requires that $E\left(\Delta y_{i2}\Delta y_{i1}\right)\neq 0$. But

$$E\left(\Delta y_{i2}\Delta y_{i1}\right)=\phi E\left(\Delta y_{i1}^{2}\right)+E\left(\Delta y_{i1}\Delta u_{i2}\right)=\phi E\left(\Delta y_{i1}^{2}\right)-\sigma_{i1}^{2},$$

and $\Delta y_{i1}=u_{i1}-(1-\phi)\left(y_{i0}-\mu_{i}\right)$. Hence,

$$E\left(\Delta y_{i1}^{2}\right)=\sigma_{i1}^{2}+(1-\phi)^{2}E\left(y_{i0}-\mu_{i}\right)^{2}-2(1-\phi)E\left[u_{i1}\left(y_{i0}-\mu_{i}\right)\right], \tag{S.20}$$

and

$$\begin{aligned} E\left(\Delta y_{i2}\Delta y_{i1}\right) &= \phi\left\{\sigma_{i1}^{2}+(1-\phi)^{2}E\left(y_{i0}-\mu_{i}\right)^{2}-2(1-\phi)E\left[u_{i1}\left(y_{i0}-\mu_{i}\right)\right]\right\}-\sigma_{i1}^{2} \\ &= -(1-\phi)\sigma_{i1}^{2}+\phi(1-\phi)^{2}E\left(y_{i0}-\mu_{i}\right)^{2}-2\phi(1-\phi)E\left[u_{i1}\left(y_{i0}-\mu_{i}\right)\right]. \end{aligned}$$

Therefore, $E\left(\Delta y_{i2}\Delta y_{i1}\right)=0$ if $\phi=1$, irrespective of whether $\sigma_{i1}^{2}=\sigma_{i2}^{2}$. Otherwise, $E\left(\Delta y_{i2}\Delta y_{i1}\right)\neq 0$, in general.

S.3.2 Identification of ϕ under AB

In view of (S.18), for identification of AB estimator it is necessary that

$$\Delta_{AB}=\omega_{1}E\left(y_{i1}\Delta y_{i2}\right)+\omega_{2}E\left(y_{i0}\Delta y_{i2}\right)+\omega_{3}E\left(y_{i0}\Delta y_{i1}\right)\neq 0,$$

for some constants ω_1,ω_2, and ω_3. To derive Δ_{AB}, in addition to Assumptions 2-3, following the literature we also assume that $E\left(u_{it}y_{i0}\right)=0$, for $t=1,2$, and $E\left(u_{i1}\mu_{i}\right)=0$. Consider the three terms in the Δ_{AB}, separately. We have

$$E\left(y_{i1}\Delta y_{i2}\right)=E\left(\Delta y_{i1}\Delta y_{i2}\right)+E\left(y_{i0}\Delta y_{i2}\right),$$

where (noting that $E\left(y_{i0}\Delta u_{i2}\right)=0$ by assumption)

$$E\left(y_{i0}\Delta y_{i2}\right)=E\left[y_{i0}\left(\phi\Delta y_{i1}+\Delta u_{i2}\right)\right]=\phi E\left(y_{i0}\Delta y_{i1}\right). \tag{S.21}$$

Using $\Delta y_{i1} = u_{i1} - (1-\phi)(y_{i0} - \mu_i)$, and noting that $E(y_{i0}u_{i1}) = 0$ by assumption,

$$E(y_{i0}\Delta y_{i1}) = E\{y_{i0}[u_{i1} - (1-\phi)(y_{i0} - \mu_i)]\} = -(1-\phi)E[y_{i0}(y_{i0} - \mu_i)]. \tag{S.22}$$

In addition,

$$E(\Delta y_{i1}\Delta y_{i2}) = E[\Delta y_{i1}(\phi\Delta y_{i1} + \Delta u_{i2})] = \phi E\left(\Delta y_{i1}^2\right) - \sigma_{i1}^2.$$

Hence,

$$\begin{aligned}\Delta_{AB} &= \omega_1[E(\Delta y_{i1}\Delta y_{i2}) + E(y_{i0}\Delta y_{i2})] + \omega_2 E(y_{i0}\Delta y_{i2}) + \omega_3 E(y_{i0}\Delta y_{i1}) \\ &= \omega_1\left[\phi E\left(\Delta y_{i1}^2\right) - \sigma_{i1}^2\right] + (\omega_1 + \omega_2)E(y_{i0}\Delta y_{i2}) + \omega_3 E(y_{i0}\Delta y_{i1})\end{aligned}$$

and using (S.21),

$$\Delta_{AB} = \omega_1\left[\phi E\left(\Delta y_{i1}^2\right) - \sigma_{i1}^2\right] + [(\omega_1 + \omega_2)\phi + \omega_3]E(y_{i0}\Delta y_{i1}).$$

Using (S.22), we have

$$\Delta_{AB} = \omega_1\left[\phi E\left(\Delta y_{i1}^2\right) - \sigma_{i1}^2\right] - [(\omega_1 + \omega_2)\phi + \omega_3]\{(1-\phi)E[y_{i0}(y_{i0} - \mu_i)]\}.$$

But (using (S.20) and noting that $E(u_{i1}y_{i0}) = 0$ and $E(u_{i1}\mu_i) = 0$),

$$\begin{aligned}\phi E\left(\Delta y_{i1}^2\right) - \sigma_{i1}^2 &= \phi\left\{\sigma_{i1}^2 + (1-\phi)^2 E(y_{i0} - \mu_i)^2\right\} - \sigma_{i1}^2 \\ &= -(1-\phi)\sigma_{i1}^2 + \phi(1-\phi)^2 E(y_{i0} - \mu_i)^2.\end{aligned}$$

Hence, overall

$$\begin{aligned}\Delta_{AB} &= -[(\omega_1 + \omega_2)\phi + \omega_3](1-\phi)E[y_{i0}(y_{i0} - \mu_i)] \\ &\quad + \omega_1\left\{-(1-\phi)\sigma_{i1}^2 + \phi(1-\phi)^2 E(y_{i0} - \mu_i)^2\right\}.\end{aligned}$$

It now follows that $\Delta_{AB} = 0$, if $\phi = 1$, no matter what weights are chosen, and irrespective of whether $\sigma_{i1}^2 = \sigma_{i2}^2$. If $|\phi| < 1$, $\Delta_{AB} \neq 0$ for a suitable choice of $\{\omega_1, \omega_2, \omega_3\}$.

S.3.3 Identification of ϕ under BB

In view of (S.19) for identification of BB estimator it is necessary that

$$\Delta_{BB} = \omega_1 E(\Delta y_{i1}y_{i1}) + \omega_2 E(\Delta y_{i2}y_{i2}) + \omega_3 E(\Delta y_{i1}y_{i2}) \neq 0, \tag{S.23}$$

for some constants ω_1, ω_2 and ω_3. To derive Δ_{BB}, in addition to Assumptions 2-3, following the literature we also assume that $E(u_{it}y_{i0}) = 0$, and $E[\mu_i(y_{i0} - \mu_i)] = 0$ for $t = 1, 2$, and $E(u_{i1}\mu_i) = 0$. Consider the individual terms in (S.23). We have

$$E(\Delta y_{i1} y_{i1}) = E\left(\Delta y_{i1}^2\right) + E(\Delta y_{i1} y_{i0}),$$

and using $\Delta y_{i1} = u_{i1} - (1-\phi)(y_{i0} - \mu_i)$, and assuming $E(u_{i1}y_{i0}) = 0$, we obtain

$$\begin{aligned} E(y_{i1}\Delta y_{i1}) &= E\left(\Delta y_{i1}^2\right) + E\{[u_{i1} - (1-\phi)(y_{i0} - \mu_i)]\, y_{i0}\} \\ &= E\left(\Delta y_{i1}^2\right) - (1-\phi) E[y_{i0}(y_{i0} - \mu_i)] \end{aligned} \tag{S.24}$$

Also

$$E(\Delta y_{i2} y_{i2}) = E\left(\Delta y_{i2}^2\right) + E(\Delta y_{i2} y_{i1}) = E\left(\Delta y_{i2}^2\right) + \phi E(\Delta y_{i1} y_{i1}) + E(\Delta u_{i2} y_{i1}),$$

where (using condition (6) of Assumption 3, and $E(u_{it}y_{i0}) = 0$, for $t = 1, 2$), and

$$\begin{aligned} E(\Delta u_{i2} y_{i1}) &= E[\Delta u_{i2}(\Delta y_{i1} + y_{i0})] \\ &= E(\Delta u_{i2} y_{i0}) + E[\Delta u_{i2}(u_{i1} - (1-\phi)(y_{i0} - \mu_i))] \\ &= -\sigma_{i1}^2. \end{aligned}$$

Hence,

$$E(\Delta y_{i2} y_{i2}) = E\left(\Delta y_{i2}^2\right) + \phi E(\Delta y_{i1} y_{i1}) - \sigma_{i1}^2. \tag{S.25}$$

Furthermore,

$$\begin{aligned} E(\Delta y_{i1} y_{i2}) &= E[\Delta y_{i1}(\phi y_{i1} + u_{i2})] = \phi E(\Delta y_{i1} y_{i1}) + E(\Delta y_{i1} u_{i2}) \\ &= \phi E(\Delta y_{i1} y_{i1}) + E([u_{i1} - (1-\phi)(y_{i0} - \mu_i)]\, u_{i2}) \\ &= \phi E(\Delta y_{i1} y_{i1}). \end{aligned} \tag{S.26}$$

Using (S.24), (S.25) and (S.26) in (S.23), and setting $\omega_2 + \omega_3 = \tilde{\omega}_3$, we have

$$\Delta_{BB} = (\omega_1 + \phi\tilde{\omega}_3)\left\{E\left(\Delta y_{i1}^2\right) - (1-\phi) E[y_{i0}(y_{i0} - \mu_i)]\right\} + \omega_2\left[E\left(\Delta y_{i2}^2\right) - \sigma_{i1}^2\right]. \tag{S.27}$$

But (see (A.2))

$$E\left(\Delta y_{i2}^2\right) - \sigma_{i1}^2 = \phi^2 E\left(\Delta y_{i1}^2\right) - 2\phi\sigma_{i1}^2 + \sigma_{i2}^2, \tag{S.28}$$

and (see (S.20) and recall that by assumption $E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right]=E\left(u_{i1}y_{i0}\right)-E\left(u_{i1}\mu_i\right)=0$),

$$E\left(\Delta y_{i1}^2\right)=\sigma_{i1}^2+(1-\phi)^2E\left(y_{i0}-\mu_i\right)^2. \tag{S.29}$$

Using the above results in (S.27) we now have

$$\begin{aligned}\Delta_{BB} &= \left(\omega_1+\phi\tilde{\omega}_3\right)E\left(\Delta y_{i1}^2\right)-(1-\phi)\left(\omega_1+\phi\tilde{\omega}_3\right)E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right] \\ &\quad +\omega_2\left[\phi^2E\left(\Delta y_{i1}^2\right)-2\phi\sigma_{i1}^2+\sigma_{i2}^2\right]. \\ &= \left(\omega_1+\phi\tilde{\omega}_3+\omega_2\phi^2\right)E\left(\Delta y_{i1}^2\right)+\omega_2\left(\sigma_{i2}^2-2\phi\sigma_{i1}^2\right)-(1-\phi)\left(\omega_1+\phi\tilde{\omega}_3\right)E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right],\end{aligned}$$

and upon using (S.29), and after some algebra we have

$$\begin{aligned}\Delta_{BB} &= \left[\omega_1-\phi(1-\phi)\omega_2+\phi\omega_3\right]\sigma_{i1}^2+\omega_2\sigma_{i2}^2-(1-\phi)\left[\omega_1+\phi\left(\omega_2+\omega_3\right)\right]E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right] \\ &\quad \left[\omega_1+\phi\left(\omega_2+\omega_3\right)+\omega_2\phi^2\right](1-\phi)^2E\left(y_{i0}-\mu_i\right)^2,\end{aligned}$$

which can be written equivalently as

$$\begin{aligned}\Delta_{BB} &= \omega_1\left[\sigma_{i1}^2-(1-\phi)E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right]+(1-\phi)^2E\left(y_{i0}-\mu_i\right)^2\right] \\ &\quad +\omega_2\left[\sigma_{i2}^2-\phi(1-\phi)\sigma_{i1}^2-\phi(1-\phi)E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right]+\phi(1+\phi)(1-\phi)^2E\left(y_{i0}-\mu_i\right)^2\right] \\ &\quad +\omega_3\phi\left[\sigma_{i1}^2-(1-\phi)E\left[y_{i0}\left(y_{i0}-\mu_i\right)\right]+(1-\phi)^2E\left(y_{i0}-\mu_i\right)^2\right]\end{aligned}$$

When $\phi=1$, then it is clear that $\Delta_{BB}=\left(\omega_1+\omega_3\right)\sigma_{i1}^2+\omega_2\sigma_{i2}^2$, therefore, in general, $\Delta_{BB}\neq 0$ for all values of $|\phi|<1$.

S.4 Extension to unbalanced panels with fixed and time effects

Extending the panel VAR(p) model, (59), to include time effects we have:

$$\mathbf{z}_{it}=\boldsymbol{\alpha}_i+\boldsymbol{\delta}_t+\sum_{\ell=1}^{p}\boldsymbol{\Phi}_\ell\mathbf{z}_{i,t-\ell}+\mathbf{u}_{it}, \tag{S.30}$$

for $i=1,2,...,n$, and $t=-m_i+1,-m_i+2...,1,2,...,T$, with the starting values given by $\mathbf{z}_{i,-m_i-p+1}$, $\mathbf{z}_{i,-m_i-p+2}$, ..., $\mathbf{z}_{i,-m_i}$ for $m_i\geq 0$, and some $p\geq 1$. Suppose that available observations are $\mathbf{z}_{it}$ for $i=1,2,...,n$, and possibly only some $t\in\{0,1,2,...T\}$. Hence the panel of available observations is potentially unbalanced.[S2] To deal with unbalanced panels, additional notations are required. Let $\mathcal{T}_i\subseteq$

[S2] When panel is unbalanced, it is assumed that the identity of missing observations is purely random (independent of model parameters and errors).

$\{p+1, p+2, ..., T-1\}$ denote an indexed set for which observations $\Delta \mathbf{z}_{i,t+1}, \Delta \mathbf{z}_{it}, \Delta \mathbf{z}_{i,t-1}, ..., \Delta \mathbf{z}_{i,t-p}$ are all available (for a given i). Assume that $\mathcal{T}_i$ is non-empty for all $i = 1, 2, ..., n$, and let $\mathcal{T} \equiv \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_n\}$. In addition, for a given t, denote the index set of available observations on $\Delta \mathbf{z}_{it}$ as $\mathcal{N}_t \subseteq \{1, 2, ..., n\}$, and the cardinality of $\mathcal{N}_t$ and $\mathcal{T}_i$, by $\#\mathcal{N}_t$ and $\#\mathcal{T}_i$, respectively. Further, suppose that $\inf_t \mathcal{N}_t \to \infty$, and T is fixed as $n \to \infty$.

Define demeaned first-differences

$$\Delta \tilde{\mathbf{z}}_{it} = \Delta \mathbf{z}_{it} - \Delta \bar{\mathbf{z}}_t, \text{ where } \Delta \bar{\mathbf{z}}_t = (\#\mathcal{N}_t)^{-1} \sum_{i \in \mathcal{N}_t} \Delta \mathbf{z}_{it}.$$

The BMM estimator of $\mathbf{\Phi}$ can be computed as

$$\hat{\mathbf{\Phi}}_{nT} = \arg \min_{\mathbf{\Phi} \in \mathbf{\Theta}} \left\| \bar{\mathbf{M}}_{nT} (\mathbf{\Phi}) \right\|, \tag{S.31}$$

where $\bar{\mathbf{M}}_{n\mathcal{T}} (\mathbf{\Phi}) = n^{-1} \sum_{i=1}^{n} \mathbf{M}_{i\mathcal{T}_i} (\mathbf{\Phi})$, $\mathbf{M}_{i\mathcal{T}_i} (\mathbf{\Phi}) = \left[\mathbf{M}_{i\mathcal{T}_i}^{(1)} (\mathbf{\Phi}), \mathbf{M}_{i\mathcal{T}_i}^{(2)} (\mathbf{\Phi}), ..., \mathbf{M}_{i\mathcal{T}_i}^{(p)} (\mathbf{\Phi}) \right]$, and the individual elements $\mathbf{M}_{i\mathcal{T}_i} (\mathbf{\Phi})$ are given by

$$\begin{aligned} \mathbf{M}_{i\mathcal{T}_i}^{(1)} (\mathbf{\Phi}) \quad = \quad & \frac{1}{\#\mathcal{T}_i} \sum_{t \in \mathcal{T}_i} \left(\Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell} \right) \Delta \mathbf{z}'_{i,t-1} \\ & + \frac{1}{\#\mathcal{T}_i} \sum_{t \in \mathcal{T}_i} \left(\Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell} \right) \left(\Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell} \right)' \\ & + \frac{1}{\#\mathcal{T}_i} \sum_{t \in \mathcal{T}_i} \left(\Delta \mathbf{z}_{i,t+1} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t+1-\ell} \right) \Delta \mathbf{z}'_{it}, \end{aligned} \tag{S.32}$$

and

$$\mathbf{M}_{i\mathcal{T}_i}^{(\ell)} (\mathbf{\Phi}) = \frac{1}{\#\mathcal{T}_i} \sum_{t \in \mathcal{T}_i} \left(\Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell} \right) \Delta \mathbf{z}'_{i,t-\ell}, \text{ for } \ell = 2, 3, ..., p.$$

S.5 Consistent estimation of $\bar{\mathbf{\Omega}}_t$, for $t = 1, 2, ..., T$

Consider the panel VAR(p) data model given by (59) and suppose that $\Delta \mathbf{u}_{it} = \Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{z}_{i,t-\ell}$ is consistently estimated by $\widehat{\Delta \mathbf{u}_{it}} = \Delta \mathbf{z}_{it} - \sum_{\ell=1}^{p} \hat{\mathbf{\Phi}}_\ell \mathbf{z}_{i,t-\ell}$, where $\hat{\mathbf{\Phi}}_\ell$ represents a consistent estimator of $\mathbf{\Phi}_\ell$. Consider the following average error covariance matrices

$$\bar{\mathbf{\Omega}}_t = \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \mathbf{\Omega}_{it},$$

where $\mathbf{\Omega}_{it} = E(\mathbf{u}_{it} \mathbf{u}'_{it})$. It is then easily established that $\bar{\mathbf{\Omega}}_t$, for $t = 1, 2, ..., T$, can be consistently estimated by

$$\hat{\bar{\mathbf{\Omega}}}_{n,t} = -\frac{1}{n} \sum_{i=1}^{n} \widehat{\Delta \mathbf{u}_{i,t+1}} \widehat{\Delta \mathbf{u}_{i,t}}, \text{ for } t = 2, 3, ..., T,$$

and

$$\widehat{\widehat{\mathbf{\Omega}}}_{n,1} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\Delta\mathbf{u}_{i,2}}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\widehat{\Delta\mathbf{u}_{i,3}}\widehat{\Delta\mathbf{u}_{i,2}}.$$

S.6 Consistent estimation of the asymptotic variance of the Anderson and Hsiao estimator

Denote the AH estimator given by equation (8.1) of Anderson and Hsiao (1981) as

$$\hat{\phi}_{nT}^{AH} = \frac{\sum_{i=1}^{n}\sum_{t=3}^{T}\Delta y_{it}\Delta y_{i,t-2}}{\sum_{i=1}^{n}\sum_{t=3}^{T}\Delta y_{i,t-1}\Delta y_{i,t-2}}. \tag{S.33}$$

A consistent estimator of the asymptotic variance of $\hat{\phi}_{nT}^{AH}$ is given by

$$\hat{\Sigma}_{nT}^{AH} = \left(\widehat{\widehat{B}}_{nT}^{AH}\right)^{-2}\frac{1}{n}\sum_{i=1}^{n}\left(\hat{V}_{i,nT}^{AH}\right)^2, \tag{S.34}$$

where

$$\widehat{\widehat{B}}_{nT}^{AH} = \frac{1}{n(T-2)}\sum_{i=1}^{n}\sum_{t=3}^{T}\Delta y_{i,t-1}\Delta y_{i,t-2},\ \hat{V}_{i,nT}^{AH} = \frac{1}{T-2}\sum_{t=3}^{T}\Delta y_{i,t-2}\Delta\hat{u}_{it}^{AH}, \tag{S.35}$$

and $\Delta\hat{u}_{it}^{AH} = \Delta y_{it} - \hat{\phi}_{nT}^{AH}\Delta y_{i,t-1}$ ($\Delta\hat{u}_{it}^{AH}$ depends on n and T, but subscripts n and T are omitted to simplify the notations).

S.7 MC findings

This section presents additional MC findings, not reported in the paper. List of the experiments, based on the choices of parameters $\phi = 0.4, 0.8$, and $\mu_v = 0, 1$ is provided in the following Table.

Table: List of Monte Carlo experiments based on $\phi = 0.4, 0.8$ **, and** $\mu_v = 0.1$.

	Parameters				Parameters		
Exp.	ϕ	μ_v	Tables	Exp.	ϕ	μ_v	Tables
1	0.8	0	1.a-b in the paper for $n \leq 1000$	3	0.4	0	S3a-b for $n \leq 1000$
			S1a-b for $n > 1000$				S3c-d for $n > 1000$
2	0.8	1	2.a-b in the paper for $n \leq 1000$	4	0.4	1	S4a-b for $n \leq 1000$
			S2a-b for $n > 1000$				S4c-d for $n > 1000$

Notes: ϕ is the autoregressive parameter of interest. μ_v governs the mean of the deviations of the initial values from μ_i, respectively. Detailed description of the design is provided in Subsection 6.1.

In addition to experiments listed in the Table above, we report MC findings for the performance of the BMM estimators in experiment with $\phi = 1$ in Table S5. Results for the BMM estimator and large values of $T = 100, 250, 500$ in experiment with $\phi = 0.8$ are presented in Table S6.

Table S1a: Bias and RMSE of alternative estimates of ϕ for Experiment 1 for large $n > 1000$

$\phi = 0.8$, $\mu_v = 0$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n								**Bias (x100)**						
3	**2000**	0.05	-49.53	-5.95	-4.24	1.66	-5.95	-4.24	1.66	-0.08	1.26	0.07	-0.08	1.26	0.07
3	**5000**	0.05	-101.89	-2.84	-1.94	0.11	-2.84	-1.94	0.11	0.08	0.53	0.08	0.08	0.53	0.08
3	**10000**	0.07	-62.82	-1.07	-0.63	0.36	-1.07	-0.63	0.36	0.02	0.28	0.05	0.02	0.28	0.05
5	**2000**	0.11	3.02	-4.15	-3.06	0.32	-2.64	-2.18	0.26	0.73	0.43	0.06	0.79	0.36	-0.02
5	**5000**	0.00	1.25	-1.57	-1.15	0.25	-0.97	-0.81	0.18	0.25	0.10	-0.05	0.27	0.07	-0.09
5	**10000**	0.00	-0.12	-0.93	-0.68	0.01	-0.60	-0.48	0.01	0.11	0.07	0.00	0.12	0.06	-0.02
10	**2000**	0.10	1.11	-1.83	-1.18	0.43	-1.35	-1.08	0.13	0.94	0.19	-0.02	1.16	0.11	-0.16
10	**5000**	0.03	0.40	-0.73	-0.46	0.20	-0.45	-0.37	0.10	0.42	0.04	-0.02	0.51	0.00	-0.08
10	**10000**	0.00	0.04	-0.45	-0.30	0.03	-0.35	-0.28	-0.04	0.17	0.00	-0.02	0.22	-0.02	-0.05
20	**2000**	0.01	0.42	-0.67	-0.37	0.24	-0.39	-0.36	0.05	0.64	0.11	-0.03	1.16	0.02	-0.27
20	**5000**	0.02	0.20	-0.28	-0.16	0.08	-0.19	-0.16	0.00	0.26	0.02	-0.01	0.48	-0.04	-0.11
20	**10000**	-0.01	0.10	-0.16	-0.09	0.03	-0.12	-0.11	-0.03	0.12	0.00	-0.01	0.23	-0.05	-0.07
									RMSE(x100)						
3	**2000**	3.30	1960.45	25.88	25.87	26.35	25.88	25.87	26.35	5.07	3.78	3.12	5.07	3.78	3.12
3	**5000**	2.01	5529.03	14.59	14.20	13.87	14.59	14.20	13.87	3.02	2.01	1.89	3.02	2.01	1.89
3	**10000**	1.41	1870.41	10.45	10.11	10.08	10.45	10.11	10.08	2.10	1.39	1.34	2.10	1.39	1.34
5	**2000**	2.55	25.73	10.99	10.33	10.20	10.79	10.24	10.20	3.13	1.93	1.77	3.14	1.92	1.78
5	**5000**	1.59	15.16	6.70	6.35	6.35	6.69	6.36	6.38	1.99	1.12	1.09	1.99	1.13	1.10
5	**10000**	1.11	10.26	4.72	4.46	4.43	4.70	4.45	4.44	1.39	0.78	0.76	1.39	0.78	0.76
10	**2000**	1.87	12.91	3.96	3.62	3.55	4.89	4.61	4.41	2.13	1.02	0.90	2.22	1.20	1.12
10	**5000**	1.18	7.88	2.49	2.31	2.30	3.01	2.84	2.80	1.32	0.57	0.54	1.34	0.71	0.70
10	**10000**	0.86	5.63	1.68	1.56	1.54	2.07	1.93	1.90	0.92	0.40	0.39	0.92	0.50	0.49
20	**2000**	1.27	8.34	1.36	1.26	1.27	1.90	1.83	1.76	1.22	0.53	0.49	1.63	0.82	0.81
20	**5000**	0.79	5.17	0.80	0.75	0.74	1.19	1.14	1.12	0.70	0.30	0.29	0.86	0.50	0.49
20	**10000**	0.56	3.60	0.57	0.54	0.53	0.83	0.80	0.80	0.48	0.20	0.20	0.56	0.34	0.34

Notes: The DGP is given by $y_{it} = (1-\phi)\mu_i + \phi y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m_i + 2, ..., T$, with $y_{i,-m_i} = \kappa_i \mu_i + \upsilon_i$, where $\kappa_i \sim IIDU(0.5, 1.5)$ and $\upsilon_i \sim IIDN(\mu_v, 1)$ measure the extent to which starting values deviate from the long-run values $\mu_i = (\alpha + w_i)/(1-\phi)$, and $w_i \sim IIDN(0, \sigma_w^2)$. We set $\alpha = 1$, and σ_w^2 is set to ensure $V(\alpha_i) = 1$. Individual effects are generated to be cross-sectionally heteroskedastic and non-normal, $u_{it} = (e_{it} - 2)\sigma_{ia}/2$ for $t \leq [T/2]$, and $u_{it} = (e_{it} - 2)\sigma_{ib}/2$ for $t > [T/2]$, with $\sigma_{ia}^2 \sim IIDU(0.25, 0.75)$, $\sigma_{ib}^2 \sim IIDU(1, 2)$, $e_{it} \sim IID\chi^2(2)$, and $[T/2]$ is the integer part of $T/2$. The BMM estimator is given by (16). Anderson and Hsiao (AH) IV estimator is given by (S.33). Moment conditions employed in the first-difference GMM methods (Arellano and Bond) are "DIF1" and "DIF2", given by (70) and (71), respectively. Moment conditions employed in the system-GMM methods (Blundell and Bond) are "SYS1" given by (70) and (72), and "SYS2" given by (71) and (72). We implement one-step (1Step), two-step (2Step) and continuous updating (CU) GMM estimators, based on the each set of the moment conditions. Subsection 6.2 provides the full description of individual estimation methods

Table S1b: Size and Power of tests for ϕ in the case of Experiments 1 for large $n > 1000$

$\phi = 0.8$, $\mu_v = 0$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n	**Size (5% level, ×100**, $H_0: \phi = 0.8$)																					
3	**2000**	5.8	5.4	3.9	4.4	3.2	3.3	2.8	3.9	4.4	3.2	3.3	2.8	7.3	9.1	6.0	7.1	6.2	7.3	9.1	6.0	7.1	6.2
3	**5000**	4.9	5.5	4.8	4.0	3.7	3.2	3.1	4.8	4.0	3.7	3.2	3.1	4.8	5.5	4.9	4.7	4.3	4.8	5.5	4.9	4.7	4.3
3	**10000**	4.8	5.6	4.9	5.3	5.0	5.1	5.1	4.9	5.3	5.0	5.1	5.1	5.3	5.5	5.3	5.1	4.7	5.3	5.5	5.3	5.1	4.7
5	**2000**	4.7	4.5	6.7	6.7	5.8	6.0	5.2	5.4	5.5	4.8	5.0	5.0	7.8	10.6	5.9	7.9	6.6	7.8	9.7	6.5	7.3	6.5
5	**5000**	4.7	3.9	6.2	5.4	5.5	5.6	5.3	4.9	4.8	5.0	5.0	4.9	6.4	6.5	5.3	6.4	5.1	6.4	6.4	5.3	6.3	5.2
5	**10000**	4.2	5.2	5.4	5.4	5.1	4.7	4.6	5.2	5.2	5.0	4.6	4.8	5.2	5.9	5.2	5.3	4.8	5.1	5.9	5.1	5.2	4.7
10	**2000**	4.3	5.6	7.6	7.6	6.2	7.7	6.9	4.8	6.0	5.6	5.9	5.4	10.0	14.8	6.9	9.6	7.5	11.8	11.3	6.4	9.1	7.0
10	**5000**	4.7	4.8	6.5	7.2	6.4	6.6	6.1	5.7	5.8	5.6	5.8	6.1	6.9	7.5	5.4	6.1	5.5	7.9	7.5	5.7	7.2	6.3
10	**10000**	6.1	5.4	5.9	5.2	4.8	4.9	4.6	5.0	4.8	4.7	4.8	4.5	6.4	6.0	4.7	5.5	5.0	6.9	6.5	5.8	6.1	5.6
20	**2000**	5.4	4.6	8.7	11.7	6.4	13.1	13.4	5.3	6.1	5.5	6.3	7.0	10.1	17.2	3.9	13.0	12.2	19.6	11.2	4.8	10.5	8.5
20	**5000**	4.2	4.9	5.4	7.0	5.0	7.4	7.5	5.4	5.7	5.3	5.2	5.6	6.5	8.6	4.7	7.4	6.5	9.2	7.5	5.3	6.9	6.0
20	**10000**	5.1	4.5	6.7	6.6	5.7	6.8	6.8	4.8	5.7	5.5	5.5	5.8	5.6	6.7	4.9	6.2	5.8	7.2	6.3	5.3	6.2	5.7
		Power (5% level, ×100, $H_1: \phi = 0.9$)																					
3	**2000**	83.5	7.4	10.8	10.8	7.6	6.4	4.9	10.8	10.0	7.6	6.4	4.9	57.2	77.9	76.9	85.0	87.4	57.2	77.9	76.9	85.0	87.4
3	**5000**	99.2	7.4	15.6	15.6	13.2	10.6	10.0	15.6	13.9	13.2	10.6	10.0	98.8	99.1	99.5	99.8	99.9	98.8	99.1	99.5	99.8	99.9
3	**10000**	100.0	7.9	21.1	21.1	20.3	18.1	18.0	21.1	20.6	20.3	18.1	18.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	**2000**	93.6	10.7	29.5	29.5	26.2	18.1	16.5	22.8	22.7	22.5	16.1	16.2	94.1	99.7	99.6	99.9	99.7	93.6	99.8	99.6	100.0	99.9
5	**5000**	99.9	14.2	43.3	43.3	43.6	35.6	35.6	40.1	40.8	40.9	35.7	36.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	**10000**	100.0	20.3	67.0	67.0	68.3	62.7	62.1	63.6	66.8	66.5	62.4	61.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	**2000**	98.8	15.3	92.2	92.2	91.5	82.8	80.2	72.3	75.6	75.3	66.4	66.4	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	99.9
10	**5000**	100.0	26.8	99.6	99.6	99.6	99.3	99.2	95.2	97.5	97.4	96.2	96.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	**10000**	100.0	45.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**2000**	100.0	26.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**5000**	100.0	48.6	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**10000**	100.0	76.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9

Notes: See notes to Table 1a. Two-step Arrelano and Bond's first-difference GMM and Blundell and Bond's system GMM estimators with the suffix "w" use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators with the suffix "nw" use Newey and Windmeijer (2009)'s standard errors.

Table S2a: Bias and RMSE of alternative estimates of ϕ for Experiment 2 for large $n > 1000$

$$\phi = 0.8,\ \mu_v = 1$$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**2000**	0.06	10.59	-1.03	-0.35	0.82	-1.03	-0.35	0.82	8.79	10.79	5.81	8.79	10.79	5.81
3	**5000**	0.04	41.32	-0.73	-0.47	-0.02	-0.73	-0.47	-0.02	8.74	10.60	3.76	8.74	10.60	3.76
3	**10000**	0.07	-174.06	-0.20	-0.07	0.15	-0.20	-0.07	0.15	8.91	10.82	2.54	8.91	10.82	2.54
5	**2000**	0.11	3.76	-1.56	-0.75	0.29	-1.01	-0.51	0.22	6.64	7.31	-0.09	6.71	6.70	-0.08
5	**5000**	0.00	1.51	-0.62	-0.31	0.11	-0.40	-0.22	0.07	6.46	6.87	-0.35	6.49	6.27	-0.30
5	**10000**	0.00	-0.11	-0.34	-0.17	0.04	-0.22	-0.11	0.03	6.37	6.72	-0.31	6.39	6.14	-0.24
10	**2000**	0.09	1.18	-1.07	-0.40	0.36	-0.54	-0.27	0.13	4.35	3.13	-0.20	4.68	3.75	-0.40
10	**5000**	0.03	0.43	-0.42	-0.15	0.16	-0.18	-0.09	0.07	3.94	2.72	-0.20	4.17	3.36	-0.32
10	**10000**	0.00	0.04	-0.27	-0.12	0.04	-0.16	-0.09	-0.01	3.76	2.56	-0.20	3.95	3.21	-0.28
20	**2000**	0.01	0.43	-0.56	-0.20	0.27	-0.22	-0.14	0.07	1.89	1.11	-0.08	2.67	1.88	-0.45
20	**5000**	0.02	0.20	-0.24	-0.09	0.10	-0.11	-0.07	0.02	1.55	0.87	-0.08	2.09	1.64	-0.29
20	**10000**	0.00	0.10	-0.14	-0.05	0.04	-0.08	-0.05	-0.01	1.42	0.79	-0.08	1.87	1.56	-0.24
		RMSE(x100)													
3	**2000**	3.24	7032.74	11.13	10.73	10.73	11.13	10.73	10.73	10.21	11.98	11.08	10.21	11.98	11.08
3	**5000**	1.98	3060.85	6.76	6.56	6.54	6.76	6.56	6.54	9.28	11.21	8.22	9.28	11.21	8.22
3	**10000**	1.39	15547.02	4.95	4.79	4.79	4.95	4.79	4.79	9.16	11.14	6.36	9.16	11.14	6.36
5	**2000**	2.48	29.10	5.80	5.34	5.32	5.69	5.32	5.30	7.35	8.18	2.48	7.41	7.57	2.58
5	**5000**	1.55	16.84	3.65	3.46	3.46	3.62	3.45	3.46	6.76	7.27	1.21	6.79	6.66	1.23
5	**10000**	1.08	11.29	2.53	2.39	2.39	2.51	2.38	2.39	6.54	6.96	0.86	6.55	6.38	0.86
10	**2000**	1.82	13.45	2.70	2.39	2.41	2.76	2.56	2.54	4.76	3.52	0.95	5.06	4.19	1.20
10	**5000**	1.15	8.20	1.70	1.54	1.55	1.74	1.63	1.63	4.15	2.89	0.59	4.36	3.55	0.77
10	**10000**	0.84	5.86	1.15	1.05	1.04	1.19	1.11	1.10	3.88	2.65	0.45	4.06	3.32	0.58
20	**2000**	1.25	8.48	1.21	1.08	1.12	1.38	1.29	1.27	2.19	1.30	0.51	2.92	2.14	0.89
20	**5000**	0.78	5.26	0.72	0.65	0.65	0.87	0.81	0.81	1.70	0.96	0.31	2.22	1.75	0.57
20	**10000**	0.56	3.66	0.51	0.47	0.47	0.61	0.58	0.58	1.50	0.84	0.22	1.95	1.62	0.41

Notes: See notes to Tables S1a.

Table S2b: Size and Power of tests for ϕ in the case of Experiments 2 for large $n > 1000$

$$\phi = 0.8,\ \mu_v = 1$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n	**Size (5% level, ×100, $H_0: \phi = 0.8$)**																					
3	**2000**	5.2	3.8	4.8	5.1	4.8	4.9	4.7	4.8	5.1	4.8	4.9	4.7	67.8	86.8	67.2	47.3	38.5	67.8	86.8	67.2	47.3	38.5
3	**5000**	5.0	4.5	4.7	4.1	3.9	3.7	3.9	4.7	4.1	3.9	3.7	3.9	88.9	97.0	91.4	38.9	24.8	88.9	97.0	91.4	38.9	24.8
3	**10000**	4.7	4.2	5.5	5.5	5.5	5.3	5.5	5.5	5.5	5.5	5.3	5.5	98.8	100.0	99.7	35.3	17.8	98.8	100.0	99.7	35.3	17.8
5	**2000**	5.0	4.8	5.1	5.3	4.9	5.0	5.1	4.6	4.6	4.6	4.5	4.9	74.0	89.6	71.6	19.6	10.6	74.6	86.5	72.5	19.7	10.6
5	**5000**	4.7	4.0	5.4	5.6	5.7	5.7	5.6	4.6	5.5	5.5	5.5	5.5	94.3	99.4	95.9	19.3	11.4	94.4	98.6	95.3	18.2	11.1
5	**10000**	4.2	5.2	4.6	4.8	4.7	5.1	5.0	4.6	4.8	4.8	4.9	4.8	99.3	100.0	100.0	19.3	12.4	99.3	100.0	99.8	18.5	11.6
10	**2000**	4.3	5.5	6.6	6.3	5.3	7.4	7.7	4.7	5.3	5.0	5.2	5.5	74.8	89.9	68.0	22.4	11.2	80.5	89.1	72.5	21.1	11.7
10	**5000**	5.0	4.9	6.2	5.7	5.4	6.2	6.1	5.7	5.3	5.4	5.5	5.7	92.8	98.3	92.2	20.2	10.0	95.0	98.4	94.1	21.0	12.3
10	**10000**	6.1	5.5	5.3	5.1	4.9	4.9	4.9	5.0	4.6	4.6	4.6	4.6	98.9	99.9	99.0	22.9	12.4	99.3	99.9	99.5	23.0	13.6
20	**2000**	5.5	4.7	8.4	11.0	6.0	12.8	14.0	4.9	5.6	5.0	6.2	6.6	53.9	78.9	33.7	25.5	16.8	75.0	78.0	56.0	23.9	14.5
20	**5000**	4.4	4.9	5.5	6.7	5.0	7.6	7.7	5.8	4.9	4.9	5.1	5.1	71.7	88.8	62.4	20.2	10.1	88.9	93.3	83.6	23.5	14.2
20	**10000**	5.2	4.6	6.6	6.5	5.8	6.9	7.0	5.0	4.9	4.8	5.3	5.4	88.4	97.5	86.5	19.8	9.8	97.0	99.2	97.6	23.6	14.7
		Power (5% level, ×100, $H_1: \phi = 0.9$)																					
3	**2000**	84.3	5.7	17.9	17.9	15.7	13.3	13.7	17.9	16.0	15.7	13.3	13.7	12.4	45.1	26.7	92.0	76.5	12.4	45.1	26.7	92.0	76.5
3	**5000**	99.6	5.9	36.5	36.5	36.0	33.5	33.4	36.5	36.0	36.0	33.5	33.4	16.6	48.9	24.3	98.9	89.6	16.6	48.9	24.3	98.9	89.6
3	**10000**	100.0	6.1	57.0	57.0	56.4	55.1	55.0	57.0	56.4	56.4	55.1	55.0	20.3	52.1	22.8	99.7	96.4	20.3	52.1	22.8	99.7	96.4
5	**2000**	94.6	10.3	56.5	56.5	53.6	45.6	45.6	51.1	50.9	51.3	45.6	45.9	34.9	64.4	30.4	100.0	99.7	33.7	65.6	41.8	100.0	99.7
5	**5000**	100.0	13.0	85.2	85.2	85.8	82.9	82.9	82.9	85.0	85.0	82.7	82.5	68.7	81.8	47.5	100.0	100.0	67.8	85.2	64.4	100.0	100.0
5	**10000**	100.0	18.5	98.6	98.6	99.1	98.9	98.7	98.3	99.1	99.0	98.9	98.7	90.3	89.3	66.9	100.0	100.0	90.0	94.4	83.0	100.0	100.0
10	**2000**	99.1	14.6	99.4	99.4	99.5	98.7	98.5	98.2	98.8	98.7	98.2	98.0	94.9	100.0	98.4	100.0	100.0	94.1	99.1	95.2	100.0	100.0
10	**5000**	100.0	25.6	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	**10000**	100.0	42.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**2000**	100.0	25.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**5000**	100.0	47.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**10000**	100.0	75.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See notes to Tables S1a and S1b.

Table S3a: Bias and RMSE of alternative estimates of ϕ for Experiment 3

$$\phi = 0.4,\ \mu_v = 0$$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**250**	0.38	16.08	-5.51	-4.46	2.71	-5.51	-4.46	2.71	-0.16	1.87	0.33	-0.16	1.87	0.33
3	**500**	0.06	4.26	-2.74	-2.10	1.79	-2.74	-2.10	1.79	-0.18	0.97	0.01	-0.18	0.97	0.01
3	**1000**	-0.06	2.22	-1.49	-1.17	0.76	-1.49	-1.17	0.76	-0.03	0.50	-0.01	-0.03	0.50	-0.01
5	**250**	0.02	0.43	-3.25	-2.22	0.66	-2.75	-2.25	0.45	1.33	0.59	-0.13	1.61	0.35	-0.56
5	**500**	0.26	1.00	-1.34	-0.81	0.67	-0.83	-0.63	0.78	0.83	0.36	0.07	1.00	0.21	-0.18
5	**1000**	0.04	0.29	-0.69	-0.37	0.39	-0.52	-0.40	0.31	0.38	0.15	0.01	0.46	0.06	-0.12
10	**250**	-0.08	0.53	-2.09	-1.20	0.39	-1.53	-0.91	0.48	1.25	0.57	0.05	2.13	0.38	-0.62
10	**500**	-0.06	-0.05	-0.99	-0.55	0.21	-0.86	-0.56	0.15	0.73	0.17	-0.05	1.15	0.04	-0.36
10	**1000**	-0.02	0.12	-0.46	-0.18	0.21	-0.36	-0.20	0.16	0.40	0.02	-0.06	0.62	-0.07	-0.21
20	**250**	0.11	0.84	-1.22	-0.64	0.53	-0.59	-0.26	0.46	1.06	0.84	0.31	2.05	0.39	-0.65
20	**500**	-0.01	0.12	-0.62	-0.34	0.16	-0.28	-0.08	0.29	0.54	0.17	-0.06	1.04	0.03	-0.36
20	**1000**	-0.09	-0.11	-0.37	-0.16	0.10	-0.19	-0.12	0.07	0.20	0.03	-0.02	0.46	-0.12	-0.27
		RMSE(x100)													
3	**250**	8.23	432.92	29.82	29.03	34.40	29.82	29.03	34.40	10.19	7.58	7.45	10.19	7.58	7.45
3	**500**	5.60	121.58	20.98	20.64	22.07	20.98	20.64	22.07	7.19	5.18	5.05	7.19	5.18	5.05
3	**1000**	3.77	25.10	14.56	14.35	14.77	14.56	14.35	14.77	4.98	3.61	3.56	4.98	3.61	3.56
5	**250**	6.11	16.54	10.72	10.47	11.37	12.20	11.95	12.57	6.53	4.75	4.75	6.70	4.78	4.76
5	**500**	4.27	11.91	7.29	7.07	7.41	8.38	8.10	8.43	4.75	3.28	3.23	4.87	3.32	3.25
5	**1000**	3.10	8.13	5.42	5.13	5.24	6.14	5.88	5.98	3.43	2.27	2.25	3.47	2.32	2.29
10	**250**	4.20	10.97	4.67	4.54	4.97	5.67	5.39	5.47	3.94	2.96	3.12	4.52	3.32	3.37
10	**500**	3.05	8.04	3.23	3.06	3.19	4.00	3.77	3.79	2.93	2.00	2.00	3.19	2.27	2.25
10	**1000**	2.11	5.61	2.26	2.11	2.17	2.81	2.59	2.60	2.02	1.27	1.28	2.19	1.51	1.52
20	**250**	2.81	7.65	2.46	3.33	5.56	2.84	2.79	2.92	2.45	2.88	4.99	3.28	2.34	2.52
20	**500**	1.97	5.28	1.69	1.78	2.00	2.01	1.88	1.93	1.65	1.37	1.55	2.08	1.57	1.61
20	**1000**	1.42	3.79	1.14	1.12	1.19	1.38	1.30	1.31	1.14	0.85	0.88	1.38	1.07	1.09

Notes: See notes to Tables S1a.

Table S3b: Size and Power of tests for ϕ in the case of Experiment 3

$$\phi = 0.4,\ \mu_v = 0$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
T	n	BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
										Size (5% level, ×100, $H_0: \phi = 0.4$)													
3	**250**	5.4	5.8	5.1	6.0	4.8	6.6	6.0	5.1	6.0	4.8	6.6	6.0	4.5	8.1	6.7	9.1	8.7	4.5	8.1	6.7	9.1	8.7
3	**500**	5.1	4.9	4.4	5.1	4.9	5.7	5.0	4.4	5.1	4.9	5.7	5.0	5.1	6.3	6.0	6.5	6.4	5.1	6.3	6.0	6.5	6.4
3	**1000**	4.3	4.2	4.9	5.9	5.4	5.7	5.3	4.9	5.9	5.4	5.7	5.3	4.8	6.0	5.3	5.6	5.3	4.8	6.0	5.3	5.6	5.3
5	**250**	5.1	5.9	6.2	9.8	6.2	11.2	8.2	6.0	9.3	7.0	9.8	8.6	6.3	14.7	6.6	14.7	11.1	6.4	12.4	6.1	12.7	9.0
5	**500**	4.2	4.5	5.1	7.2	5.9	9.0	6.9	5.3	6.5	5.5	7.5	6.5	6.6	9.9	6.2	9.2	7.9	6.8	9.3	6.1	8.1	7.5
5	**1000**	5.1	4.9	5.2	5.8	4.9	6.6	5.5	5.9	5.8	5.0	6.0	5.4	5.3	6.9	5.5	6.9	6.5	5.3	7.2	6.1	7.0	6.8
10	**250**	5.7	5.4	8.5	19.2	6.1	24.6	21.1	5.9	10.3	5.6	11.6	10.4	6.8	27.1	2.9	30.9	27.8	9.5	18.0	5.8	18.6	15.2
10	**500**	5.6	5.9	7.6	11.9	5.0	13.8	11.6	5.8	7.3	5.1	8.3	7.4	7.0	17.6	6.1	18.2	16.1	7.8	11.0	5.3	11.1	9.3
10	**1000**	4.7	5.4	6.2	8.5	5.6	9.2	8.6	5.1	5.3	4.2	5.7	5.1	5.5	9.2	4.8	9.7	8.8	6.1	6.9	5.2	7.5	6.6
20	**250**	5.1	5.7	8.7	70.9	0.1	81.4	92.5	5.8	12.7	5.2	15.5	17.6	7.6	80.9	0.1	88.0	94.6	13.9	25.0	3.6	27.7	28.0
20	**500**	4.5	5.1	7.7	33.3	2.8	39.9	49.6	5.9	8.4	5.1	9.6	10.0	6.2	38.4	0.4	43.5	54.3	9.7	14.2	5.2	15.8	14.9
20	**1000**	5.3	5.0	5.9	15.9	4.9	18.7	21.2	4.9	6.9	4.7	6.7	6.8	5.5	18.4	3.1	20.2	23.6	6.7	9.3	5.2	9.8	9.7
										Power (5% level, ×100, $H_1: \phi = 0.5$)													
3	**250**	33.8	9.5	10.4	10.4	9.7	9.2	8.8	10.4	10.0	9.7	9.2	8.8	9.4	28.4	25.1	36.1	36.2	9.4	28.4	25.1	36.1	36.2
3	**500**	51.2	9.1	11.8	11.8	10.8	8.9	8.6	11.8	11.3	10.8	8.9	8.6	24.2	47.6	45.9	53.1	54.4	24.2	47.6	45.9	53.1	54.4
3	**1000**	74.5	10.7	14.4	14.4	14.2	12.6	12.4	14.4	14.2	14.2	12.6	12.4	49.4	73.8	73.5	77.6	78.5	49.4	73.8	73.5	77.6	78.5
5	**250**	42.8	14.4	28.6	28.6	25.5	26.6	21.3	21.5	24.6	21.1	20.8	18.8	29.2	68.5	54.3	73.2	68.4	26.7	68.5	55.8	73.8	69.3
5	**500**	62.7	15.9	35.5	35.5	33.2	32.4	28.0	26.8	29.2	26.8	24.7	22.7	50.1	87.9	82.8	89.6	87.9	48.2	87.8	83.8	90.1	88.5
5	**1000**	88.2	25.6	53.9	53.9	55.2	51.5	48.9	42.4	46.4	45.4	41.4	40.7	80.9	99.2	98.9	99.4	99.1	78.9	99.1	98.9	99.2	99.1
10	**250**	68.0	17.6	81.4	81.4	70.7	78.2	74.7	56.8	65.3	55.0	54.8	52.2	66.7	97.9	82.8	98.0	97.7	52.0	93.3	82.2	96.5	94.9
10	**500**	90.3	29.2	94.4	94.4	94.0	93.9	92.6	81.1	85.9	82.2	79.6	78.5	90.3	100.0	99.8	100.0	99.9	84.8	99.7	99.0	99.9	99.7
10	**1000**	99.7	44.6	99.7	99.7	99.7	99.6	99.6	96.5	98.4	98.3	97.9	97.6	99.8	100.0	100.0	100.0	100.0	99.6	100.0	100.0	100.0	100.0
20	**250**	93.4	25.7	99.9	99.9	7.5	96.7	98.8	96.8	98.6	96.3	97.4	97.5	97.5	100.0	0.6	98.9	99.0	87.0	99.9	97.0	100.0	100.0
20	**500**	99.8	47.7	100.0	100.0	100.0	100.0	100.0	99.8	99.9	99.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0
20	**1000**	100.0	76.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9

Notes: See notes to Tables S1a and S1b.

Table S3c: Bias and RMSE of alternative estimates of ϕ for Experiment 3 for large $n > 1000$

$$\phi = 0.4,\ \mu_v = 0$$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**2000**	0.08	1.66	-0.39	-0.33	0.69	-0.39	-0.33	0.69	0.06	0.36	0.09	0.06	0.36	0.09
3	**5000**	0.00	0.54	-0.36	-0.31	0.10	-0.36	-0.31	0.10	-0.04	0.11	0.00	-0.04	0.11	0.00
3	**10000**	0.00	0.14	-0.15	-0.09	0.10	-0.15	-0.09	0.10	-0.04	0.06	0.00	-0.04	0.06	0.00
5	**2000**	0.03	0.29	-0.39	-0.22	0.16	-0.26	-0.20	0.17	0.19	0.05	-0.02	0.24	-0.01	-0.09
5	**5000**	-0.03	-0.06	-0.28	-0.18	-0.02	-0.28	-0.24	-0.09	0.02	0.02	-0.01	0.04	-0.01	-0.04
5	**10000**	0.00	0.05	-0.11	-0.07	0.00	-0.07	-0.05	0.02	0.04	0.02	0.01	0.05	0.01	-0.01
10	**2000**	0.06	0.17	-0.16	0.01	0.20	-0.07	0.01	0.19	0.27	0.04	0.01	0.38	0.01	-0.05
10	**5000**	0.00	0.03	-0.08	-0.01	0.06	-0.06	-0.02	0.05	0.10	0.02	0.01	0.14	-0.01	-0.03
10	**10000**	0.00	0.00	-0.05	-0.02	0.02	-0.02	-0.01	0.03	0.04	0.00	0.00	0.06	-0.01	-0.02
20	**2000**	0.00	0.01	-0.15	-0.06	0.06	-0.07	-0.05	0.04	0.14	0.01	-0.01	0.27	-0.07	-0.14
20	**5000**	0.00	0.02	-0.05	-0.01	0.04	-0.01	0.01	0.04	0.06	0.01	0.00	0.12	-0.02	-0.05
20	**10000**	0.00	-0.01	-0.03	-0.01	0.02	0.00	0.00	0.02	0.03	0.00	0.00	0.06	-0.01	-0.02
		RMSE(x100)													
3	**2000**	2.71	17.18	10.52	10.16	10.34	10.52	10.16	10.34	3.73	2.58	2.54	3.73	2.58	2.54
3	**5000**	1.73	10.18	6.54	6.37	6.39	6.54	6.37	6.39	2.28	1.64	1.63	2.28	1.64	1.63
3	**10000**	1.22	7.14	4.68	4.49	4.50	4.68	4.49	4.50	1.59	1.15	1.14	1.59	1.15	1.14
5	**2000**	2.18	5.83	3.70	3.50	3.53	4.28	4.05	4.07	2.42	1.60	1.59	2.47	1.62	1.61
5	**5000**	1.38	3.66	2.47	2.33	2.33	2.75	2.59	2.59	1.54	1.01	1.00	1.55	1.03	1.03
5	**10000**	0.95	2.55	1.69	1.61	1.61	1.92	1.83	1.83	1.09	0.70	0.70	1.10	0.71	0.71
10	**2000**	1.49	3.82	1.57	1.50	1.54	1.95	1.85	1.86	1.45	0.95	0.95	1.54	1.11	1.11
10	**5000**	0.93	2.48	0.97	0.91	0.91	1.22	1.16	1.16	0.88	0.55	0.55	0.93	0.66	0.66
10	**10000**	0.67	1.76	0.71	0.66	0.66	0.88	0.83	0.83	0.63	0.40	0.40	0.67	0.48	0.48
20	**2000**	0.99	2.69	0.82	0.80	0.81	1.00	0.94	0.95	0.82	0.59	0.59	0.95	0.75	0.76
20	**5000**	0.64	1.71	0.50	0.46	0.46	0.63	0.57	0.58	0.51	0.35	0.35	0.59	0.45	0.45
20	**10000**	0.45	1.18	0.36	0.33	0.33	0.44	0.41	0.41	0.36	0.24	0.24	0.40	0.32	0.32

Notes: See notes to Tables S1a.

Table S3d: Size and Power of tests for ϕ in the case of Experiments 3 for large $n > 1000$

$$\phi = 0.4,\ \mu_v = 0$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
T	n	BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
		Size (5% level, ×100, $H_0: \phi = 0.4$)																					
3	**2000**	5.1	4.9	5.3	5.5	5.3	5.7	5.5	5.3	5.5	5.3	5.7	5.5	6.4	5.5	5.4	5.3	5.3	6.4	5.5	5.4	5.3	5.3
3	**5000**	5.6	5.2	4.6	5.6	5.6	5.8	5.7	4.6	5.6	5.6	5.8	5.7	4.8	5.4	5.3	5.5	5.7	4.8	5.4	5.3	5.5	5.7
3	**10000**	5.3	4.7	4.6	4.8	4.8	4.7	5.0	4.6	4.8	4.8	4.7	5.0	4.4	5.1	5.1	5.1	5.2	4.4	5.1	5.1	5.1	5.2
5	**2000**	5.8	5.3	4.7	4.7	4.2	5.0	4.7	4.5	5.1	4.8	5.3	5.2	4.9	7.3	6.3	6.7	6.3	5.1	6.6	6.0	6.3	6.1
5	**5000**	5.6	5.2	6.2	5.7	5.3	5.5	5.5	5.6	5.7	5.2	5.7	5.5	5.4	5.6	5.5	5.6	5.4	6.0	5.7	5.3	5.1	5.2
5	**10000**	4.9	4.6	5.2	4.9	4.9	5.1	4.9	5.4	5.1	5.0	5.1	5.0	4.8	5.4	5.2	5.2	5.1	4.6	5.1	4.9	5.0	4.9
10	**2000**	5.4	4.2	5.4	7.4	6.1	7.7	7.4	4.6	5.5	5.1	5.9	5.7	5.5	9.0	7.0	9.0	8.2	5.9	7.3	6.1	7.2	6.6
10	**5000**	4.7	5.5	4.6	5.1	4.7	5.6	5.6	4.8	5.1	4.9	5.3	5.5	4.6	5.2	4.7	5.3	5.1	5.6	4.7	4.3	4.7	4.5
10	**10000**	5.6	5.8	4.6	5.7	5.4	5.6	5.6	5.5	5.6	5.5	5.7	5.8	5.3	6.2	5.6	6.2	6.1	5.4	5.5	5.2	5.2	5.1
20	**2000**	4.9	5.5	5.7	11.9	6.4	12.4	13.6	4.9	6.0	5.7	6.1	6.1	5.7	12.1	5.1	12.6	13.7	5.9	8.0	5.8	7.7	7.3
20	**5000**	5.0	5.3	5.1	6.5	4.5	6.6	6.7	4.8	5.1	4.8	5.2	5.2	4.7	7.0	5.0	7.0	7.0	5.7	4.7	4.1	5.3	5.2
20	**10000**	4.9	4.8	5.0	6.1	5.4	6.5	6.4	5.0	5.2	5.2	5.2	5.1	5.2	5.9	4.8	5.9	6.0	4.8	5.0	4.6	4.8	4.9
		Power (5% level, ×100, $H_1: \phi = 0.5$)																					
3	**2000**	93.0	13.6	20.1	20.1	19.7	17.1	16.7	20.1	19.5	19.7	17.1	16.7	80.2	94.8	95.0	95.6	95.6	80.2	94.8	95.0	95.6	95.6
3	**5000**	100.0	20.9	37.2	37.2	38.4	36.3	36.6	37.2	38.6	38.4	36.3	36.6	99.6	100.0	100.0	100.0	100.0	99.6	100.0	100.0	100.0	100.0
3	**10000**	100.0	31.7	59.1	59.1	62.2	61.3	60.2	59.1	62.7	62.2	61.3	60.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	**2000**	99.4	41.6	79.3	79.3	81.6	78.9	78.0	67.1	70.6	69.9	67.1	66.7	98.8	100.0	100.0	100.0	100.0	98.5	100.0	100.0	100.0	100.0
5	**5000**	100.0	78.0	98.5	98.5	99.0	98.9	98.8	95.3	96.8	96.7	96.3	96.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	**10000**	100.0	96.8	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	**2000**	100.0	70.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
10	**5000**	100.0	97.6	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9
10	**10000**	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
20	**2000**	100.0	95.1	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**5000**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
20	**10000**	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0

Notes: See notes to Tables S1a and S1b.

Table S4a: Bias and RMSE of alternative estimates of ϕ for Experiment 4

$\phi = 0.4$, $\mu_v = 1$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**250**	0.42	16.28	-4.65	-2.97	1.61	-4.65	-2.97	1.61	-4.53	1.71	-1.68	-4.53	1.71	-1.68
3	**500**	0.07	8.87	-2.36	-1.30	1.34	-2.36	-1.30	1.34	-4.82	0.69	-1.96	-4.82	0.69	-1.96
3	**1000**	-0.06	2.32	-1.27	-0.72	0.56	-1.27	-0.72	0.56	-4.69	0.03	-1.97	-4.69	0.03	-1.97
5	**250**	0.02	0.41	-3.04	-1.84	0.63	-2.55	-1.75	0.47	-0.54	0.55	-1.13	-0.34	0.32	-1.47
5	**500**	0.27	1.00	-1.23	-0.63	0.68	-0.79	-0.43	0.73	-1.30	0.30	-0.89	-1.22	0.19	-1.08
5	**1000**	0.04	0.28	-0.65	-0.30	0.36	-0.48	-0.28	0.31	-1.86	0.08	-0.93	-1.89	0.05	-1.00
10	**250**	-0.07	0.53	-2.03	-1.14	0.39	-1.42	-0.71	0.55	0.55	0.62	-0.23	1.30	0.44	-1.11
10	**500**	-0.06	-0.05	-0.98	-0.53	0.20	-0.80	-0.46	0.18	0.00	0.27	-0.31	0.28	0.16	-0.84
10	**1000**	-0.02	0.12	-0.45	-0.18	0.19	-0.34	-0.15	0.17	-0.38	0.13	-0.32	-0.32	0.04	-0.71
20	**250**	0.11	0.84	-1.21	-0.70	0.43	-0.58	-0.24	0.46	0.82	0.82	0.14	1.74	0.50	-0.84
20	**500**	-0.01	0.11	-0.61	-0.35	0.14	-0.27	-0.06	0.30	0.27	0.23	-0.13	0.70	0.14	-0.57
20	**1000**	-0.09	-0.11	-0.37	-0.17	0.09	-0.18	-0.11	0.07	-0.07	0.11	-0.09	0.10	-0.01	-0.49
		RMSE(x100)													
3	**250**	8.07	1864.43	23.34	22.78	25.10	23.34	22.78	25.10	12.19	7.92	7.34	12.19	7.92	7.34
3	**500**	5.52	181.46	16.76	16.32	16.53	16.76	16.32	16.53	9.65	5.51	5.18	9.65	5.51	5.18
3	**1000**	3.74	26.76	11.66	11.41	11.53	11.66	11.41	11.53	7.39	3.78	3.89	7.39	3.78	3.89
5	**250**	6.06	16.67	10.04	9.70	10.43	11.09	10.67	11.11	6.63	4.88	4.80	6.73	4.87	4.90
5	**500**	4.25	12.01	6.80	6.56	6.86	7.55	7.24	7.49	5.04	3.40	3.31	5.12	3.42	3.40
5	**1000**	3.08	8.18	5.06	4.72	4.81	5.55	5.22	5.29	4.01	2.35	2.38	4.05	2.39	2.45
10	**250**	4.19	11.01	4.57	4.42	4.81	5.38	5.04	5.13	3.84	3.02	3.15	4.23	3.41	3.50
10	**500**	3.04	8.05	3.15	2.98	3.10	3.79	3.54	3.57	2.88	2.07	2.05	3.02	2.33	2.38
10	**1000**	2.10	5.63	2.21	2.05	2.10	2.66	2.43	2.44	2.04	1.32	1.34	2.13	1.56	1.67
20	**250**	2.80	7.66	2.45	3.32	5.43	2.79	2.73	2.86	2.36	2.87	5.08	3.09	2.40	2.60
20	**500**	1.97	5.28	1.68	1.77	1.99	1.97	1.84	1.89	1.59	1.38	1.58	1.92	1.60	1.67
20	**1000**	1.41	3.79	1.13	1.12	1.18	1.36	1.27	1.28	1.14	0.87	0.90	1.30	1.10	1.17

Notes: See notes to Tables S1a.

Table S4b: Size and Power of tests for ϕ in the case of Experiments 4

$\phi = 0.4$, $\mu_v = 1$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n	**Size (5% level, ×100, $H_0 : \phi = 0.4$)**																					
3	**250**	5.4	6.0	4.3	4.9	3.6	5.8	5.4	4.3	4.9	3.6	5.8	5.4	2.8	11.1	7.5	10.4	11.7	2.8	11.1	7.5	10.4	11.7
3	**500**	4.7	5.0	4.7	5.4	4.3	5.2	4.6	4.7	5.4	4.3	5.2	4.6	4.7	8.8	6.5	9.4	10.9	4.7	8.8	6.5	9.4	10.9
3	**1000**	4.4	4.3	4.8	5.7	5.2	5.6	5.4	4.8	5.7	5.2	5.6	5.4	8.1	7.8	6.7	10.7	12.1	8.1	7.8	6.7	10.7	12.1
5	**250**	4.8	5.9	6.6	9.4	6.1	10.5	8.5	5.9	8.7	6.9	9.5	8.2	5.6	16.9	6.3	17.8	14.4	5.9	15.6	6.4	16.1	13.0
5	**500**	4.4	4.5	5.2	7.3	5.5	9.0	7.5	5.1	6.4	5.3	7.8	6.8	5.9	12.9	6.8	12.7	10.7	5.9	11.5	6.3	12.0	10.7
5	**1000**	4.9	4.8	5.2	6.1	5.0	6.6	5.9	5.7	5.9	5.2	6.1	5.4	9.0	10.9	6.0	11.2	10.5	8.9	10.4	6.4	11.0	10.7
10	**250**	5.8	5.4	8.2	18.7	6.1	23.6	21.6	5.8	10.3	5.6	11.1	10.2	6.4	31.2	2.1	33.9	30.0	7.0	20.7	6.2	22.1	18.9
10	**500**	5.8	5.8	7.0	12.2	5.2	13.3	11.9	5.9	7.0	4.9	7.7	7.5	6.4	21.5	5.6	21.2	18.4	6.2	13.9	6.0	16.4	13.3
10	**1000**	4.7	5.4	6.1	8.1	5.0	8.8	8.1	5.0	5.3	4.4	5.5	5.2	6.7	12.5	4.4	12.6	10.6	5.9	9.8	5.5	12.3	11.5
20	**250**	5.2	5.5	8.6	71.1	0.0	81.8	92.8	5.6	12.5	5.2	15.4	18.1	6.6	80.8	0.0	88.8	95.7	11.5	26.8	3.9	30.0	31.5
20	**500**	4.4	5.1	7.5	32.7	2.7	40.1	49.8	5.8	8.6	4.8	9.1	10.2	5.3	40.2	0.3	45.5	57.0	7.0	15.5	4.9	18.4	18.5
20	**1000**	5.2	5.1	5.9	15.4	4.9	18.0	20.9	4.8	6.6	4.7	6.7	7.1	5.8	22.2	2.6	22.8	25.6	5.0	11.5	5.5	13.5	13.2
		Power (5% level, ×100, $H_1 : \phi = 0.5$)																					
3	**250**	34.0	9.8	10.5	10.5	8.5	8.3	7.9	10.5	10.0	8.5	8.3	7.9	15.7	33.8	28.2	46.8	51.2	15.7	33.8	28.2	46.8	51.2
3	**500**	51.9	9.4	13.2	13.2	11.4	9.6	9.1	13.2	11.9	11.4	9.6	9.1	41.1	52.5	49.2	68.6	74.5	41.1	52.5	49.2	68.6	74.5
3	**1000**	75.0	10.7	17.7	17.7	16.2	14.8	14.3	17.7	16.6	16.2	14.8	14.3	77.5	79.5	78.0	91.5	94.1	77.5	79.5	78.0	91.5	94.1
5	**250**	43.1	14.5	30.7	30.7	26.7	27.2	22.1	23.3	25.9	22.8	22.5	20.4	37.0	71.3	53.3	81.8	78.7	35.5	70.2	55.5	81.4	79.0
5	**500**	62.9	15.6	38.7	38.7	36.1	35.4	30.2	30.8	33.7	31.1	29.0	26.9	64.9	90.1	81.8	95.7	95.1	63.1	89.2	83.4	95.1	94.5
5	**1000**	88.7	25.3	58.5	58.5	60.4	57.7	54.5	49.0	53.6	52.0	48.9	47.8	92.3	99.4	98.5	99.8	99.7	92.2	99.1	98.0	99.9	99.8
10	**250**	68.0	17.6	83.0	83.0	73.2	79.6	77.1	60.9	69.1	58.5	58.6	56.4	71.5	97.8	79.3	98.6	98.1	60.2	93.6	80.0	97.6	96.9
10	**500**	90.2	29.0	95.1	95.1	94.6	94.9	94.0	85.3	88.3	85.9	83.3	82.9	93.8	100.0	99.2	100.0	100.0	90.0	99.7	98.5	100.0	100.0
10	**1000**	99.7	44.3	99.7	99.7	99.9	99.8	99.6	98.2	99.1	99.0	98.9	98.5	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	99.9
20	**250**	93.5	25.7	99.9	99.9	7.7	97.2	98.9	97.1	98.7	97.1	97.7	97.9	98.0	100.0	0.5	98.5	98.9	89.4	99.8	95.8	100.0	100.0
20	**500**	99.8	47.8	100.0	100.0	100.0	100.0	100.0	99.8	99.9	99.9	99.9	99.8	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.9	100.0	100.0
20	**1000**	100.0	75.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9

Notes: See notes to Tables S1a and S1b.

Table S4c: Bias and RMSE of alternative estimates of ϕ for Experiment 4 for large $n > 1000$

$$\phi = 0.4,\ \mu_v = 1$$

				Arellano and Bond						Blundell and Bond					
				"DIF1"			"DIF2"			"SYS1"			"SYS2"		
		BMM	AH	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU	1Step	2Step	CU
T	n	**Bias (x100)**													
3	**2000**	0.09	1.73	-0.38	-0.21	0.46	-0.38	-0.21	0.46	-4.62	-0.11	-1.82	-4.62	-0.11	-1.82
3	**5000**	0.00	1.07	0.00	0.07	0.33	0.00	0.07	0.33	-4.78	-0.40	-1.89	-4.78	-0.40	-1.89
3	**10000**	0.01	0.49	-0.09	-0.07	0.05	-0.09	-0.07	0.05	-4.76	-0.46	-1.91	-4.76	-0.46	-1.91
5	**2000**	0.01	0.27	-0.38	-0.19	0.15	-0.27	-0.16	0.14	-2.11	-0.02	-0.98	-2.17	0.00	-0.99
5	**5000**	0.05	0.06	-0.12	-0.10	0.04	-0.12	-0.11	0.01	-2.18	-0.03	-0.95	-2.27	0.02	-0.90
5	**10000**	0.02	-0.01	-0.05	-0.01	0.06	-0.07	-0.04	0.02	-2.27	-0.03	-0.93	-2.38	0.04	-0.88
10	**2000**	-0.02	0.10	-0.28	-0.12	0.06	-0.17	-0.08	0.09	-0.64	0.10	-0.29	-0.67	0.01	-0.64
10	**5000**	-0.02	0.03	-0.13	-0.07	0.00	-0.13	-0.08	-0.01	-0.77	0.06	-0.31	-0.89	0.03	-0.58
10	**10000**	0.00	0.04	-0.06	-0.04	0.00	-0.04	-0.03	0.00	-0.80	0.08	-0.29	-0.93	0.07	-0.54
20	**2000**	0.01	0.07	-0.12	-0.02	0.10	-0.05	-0.01	0.08	-0.11	0.11	-0.06	-0.08	0.08	-0.32
20	**5000**	0.00	-0.02	-0.07	-0.03	0.02	-0.03	-0.02	0.02	-0.23	0.08	-0.08	-0.26	0.09	-0.27
20	**10000**	0.00	0.00	-0.02	0.00	0.03	-0.01	0.00	0.02	-0.25	0.09	-0.07	-0.31	0.10	-0.25
		RMSE(x100)													
3	**2000**	2.66	17.82	8.33	7.98	8.04	8.33	7.98	8.04	6.35	2.70	3.02	6.35	2.70	3.02
3	**5000**	1.73	11.00	5.12	4.86	4.88	5.12	4.86	4.88	5.47	1.77	2.46	5.47	1.77	2.46
3	**10000**	1.21	7.53	3.68	3.54	3.55	3.68	3.54	3.55	5.12	1.30	2.20	5.12	1.30	2.20
5	**2000**	2.16	5.83	3.44	3.22	3.24	3.84	3.59	3.61	3.30	1.67	1.84	3.37	1.68	1.86
5	**5000**	1.33	3.60	2.23	2.07	2.08	2.42	2.28	2.28	2.71	1.03	1.36	2.80	1.04	1.34
5	**10000**	0.96	2.65	1.59	1.49	1.50	1.73	1.63	1.63	2.55	0.76	1.17	2.65	0.76	1.13
10	**2000**	1.49	3.93	1.54	1.43	1.44	1.86	1.72	1.73	1.56	0.94	0.96	1.63	1.11	1.24
10	**5000**	0.96	2.49	0.97	0.88	0.89	1.18	1.10	1.10	1.19	0.59	0.66	1.30	0.70	0.89
10	**10000**	0.67	1.79	0.69	0.64	0.64	0.83	0.78	0.78	1.02	0.42	0.49	1.15	0.50	0.72
20	**2000**	0.99	2.69	0.79	0.75	0.77	0.96	0.90	0.90	0.81	0.59	0.59	0.91	0.74	0.78
20	**5000**	0.65	1.67	0.51	0.48	0.48	0.64	0.60	0.60	0.57	0.38	0.38	0.65	0.49	0.55
20	**10000**	0.45	1.17	0.35	0.32	0.33	0.44	0.40	0.41	0.44	0.26	0.26	0.51	0.34	0.41

Notes: See notes to Tables S1a.

Table S4d: Size and Power of tests for ϕ in the case of Experiments 4 for large $n > 1000$

$$\phi = 0.4,\ \mu_v = 1$$

				Arellano and Bond										Blundell and Bond									
				"DIF1"					"DIF2"					"SYS1"					"SYS2"				
		BMM	AH	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw	1Step	2Step	2Step_w	CU	CU_nw
T	n	**Size (5% level, ×100, $H_0 : \phi = 0.4$)**																					
3	**2000**	5.0	4.5	5.4	5.4	5.2	5.5	5.0	5.4	5.4	5.2	5.5	5.0	17.8	8.1	7.4	13.1	15.9	17.8	8.1	7.4	13.1	15.9
3	**5000**	6.4	5.2	4.2	4.7	4.7	4.5	4.4	4.2	4.7	4.7	4.5	4.4	44.0	9.2	8.1	24.8	28.7	44.0	9.2	8.1	24.8	28.7
3	**10000**	5.0	5.1	4.8	5.3	5.5	5.6	5.6	4.8	5.3	5.5	5.6	5.6	73.3	10.5	9.0	41.6	47.0	73.3	10.5	9.0	41.6	47.0
5	**2000**	5.8	4.9	4.3	4.3	4.0	4.7	4.4	4.3	5.1	4.5	5.1	4.7	14.2	10.6	7.3	13.6	12.8	14.5	9.6	6.3	12.6	12.2
5	**5000**	4.7	4.4	5.1	4.7	4.6	4.7	4.9	4.3	4.7	4.8	4.7	5.0	28.1	8.7	6.2	20.2	19.9	28.9	8.7	6.1	18.4	18.0
5	**10000**	5.3	5.4	5.1	4.9	5.0	4.9	4.8	5.3	5.1	5.0	5.2	4.9	51.6	9.6	6.7	32.1	32.2	54.3	8.6	6.3	28.8	28.8
10	**2000**	5.8	5.2	5.9	6.9	5.3	6.9	6.5	4.8	5.0	4.5	5.3	5.3	8.3	10.5	5.3	11.5	10.0	8.2	9.3	5.1	13.1	12.3
10	**5000**	5.7	5.8	5.2	4.9	4.2	4.9	4.5	5.3	5.1	4.6	5.1	5.0	15.0	9.5	5.9	13.0	11.5	16.7	8.2	5.5	17.9	17.1
10	**10000**	5.0	5.7	5.5	5.8	5.4	5.5	5.5	4.9	5.1	5.0	5.1	5.1	24.8	8.7	5.5	14.7	12.7	29.2	8.4	5.5	25.0	24.3
20	**2000**	4.4	4.7	5.3	10.0	4.9	11.7	12.4	4.4	5.4	4.5	5.6	5.6	5.3	14.1	4.9	13.8	13.9	4.4	7.8	4.7	9.5	9.5
20	**5000**	6.2	4.8	5.6	7.1	5.4	7.0	7.2	5.6	6.7	6.5	6.6	6.5	8.4	11.5	5.9	10.9	10.3	8.0	8.7	6.5	13.0	13.0
20	**10000**	5.2	4.6	4.9	5.7	4.7	5.5	5.5	4.7	4.5	4.5	4.5	4.6	11.1	9.4	5.8	8.4	8.0	12.3	7.6	6.0	13.2	13.5
		Power (5% level, ×100, $H_1 : \phi = 0.5$)																					
3	**2000**	93.5	12.9	26.6	26.6	25.9	24.1	23.1	26.6	26.5	25.9	24.1	23.1	97.1	96.6	96.1	99.7	99.9	97.1	96.6	96.1	99.7	99.9
3	**5000**	100.0	18.7	50.0	50.0	51.7	49.9	48.8	50.0	52.1	51.7	49.9	48.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
3	**10000**	100.0	28.5	78.0	78.0	81.3	80.4	80.1	78.0	81.1	81.3	80.4	80.1	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9
5	**2000**	99.5	41.4	84.5	84.5	86.6	84.8	84.0	75.3	79.7	78.8	77.2	76.6	99.7	100.0	100.0	100.0	99.9	99.7	100.0	100.0	100.0	100.0
5	**5000**	100.0	77.1	99.4	99.4	99.7	99.6	99.5	98.2	99.2	99.2	99.1	99.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	**10000**	100.0	96.1	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	**2000**	100.0	70.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
10	**5000**	100.0	97.3	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9
10	**10000**	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**2000**	100.0	95.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**5000**	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
20	**10000**	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: See notes to Tables S1a and S1b.

Table S5: MC findings for the performance of the BMM estimator $\hat{\phi}_{nT}$ in the unit root experiment (Bias, RMSE, Size and Power, $\times 100$)

	Bias ($\times 100$)			**RMSE** ($\times 100$)			**Size (5% level,** $\times 100$)			**Power (5% level,** $\times 100$)		
(n, T)	**5**	**10**	**20**	**5**	**10**	**20**	**5**	**10**	**20**	**5**	**10**	**20**
250	1.37	-0.65	-1.13	9.86	5.82	4.08	6.95	9.85	12.20	31.00	40.85	43.80
500	1.41	0.03	-0.46	7.95	4.88	3.31	5.10	8.60	10.80	38.15	44.70	46.35
1000	1.11	0.63	-0.17	6.07	4.18	2.73	4.70	6.90	9.90	51.15	49.60	51.90
5000	0.10	0.47	0.25	2.43	2.68	1.88	4.65	5.95	6.50	93.45	78.85	67.50
20000	0.00	0.15	0.22	1.13	1.28	1.28	3.90	4.70	5.75	100.00	98.80	86.70
200000	-0.01	0.03	0.02	0.36	0.36	0.38	4.25	4.15	4.40	100.00	100.00	100.00

Notes: The DGP is given by $y_{it} = \phi y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., n$, and $t = -m_i + 1, -m_i + 2, ..., T$, with $\phi_0 = 1$, and $y_{i,-m_i} = v_i$, where $v_i \sim IIDN\,(0, 1)$. Idiosyncratic errors are generated to be cross-sectionally heteroskedastic and non-normal, $u_{it} = (e_{it} - 2)\,\sigma_{ia}/2$ for $t \leq [T/2]$, and $u_{it} = (e_{it} - 2)\,\sigma_{ib}/2$ for $t > [T/2]$, with $\sigma^2_{ia} \sim IIDU\,(0.25, 0.75)$, $\sigma^2_{ib} \sim IIDU\,(1, 2)$, $e_{it} \sim IID\chi^2\,(2)$, and $[T/2]$ is the integer part of $T/2$. $B_T = 0.33, 0.12$, and 0.05, for $T = 5, 10$, and 20, respectively.

Table S6: MC Findings for BMM and AH estimators in experiments 4 ($\phi = 0.8$) and 8 ($\phi = 0.4$) with larger values of T

$\mu_{\upsilon} = 1$

		Experiment 8		Experiment 4		Experiment 8		Experiment 4	
		$\phi = 0.4$		$\phi = 0.8$		$\phi = 0.4$		$\phi = 0.8$	
		BMM	AH	BMM	AH	BMM	AH	BMM	AH
T	n	**Bias (x100)**				**Size (5% level, ×100)**			
100	**250**	0.00	0.04	-0.01	0.36	6.1	6.1	5.2	5.8
100	**500**	-0.01	0.00	0.01	0.43	4.5	4.7	5.5	5.2
100	**1000**	-0.01	0.02	0.00	0.20	4.8	5.0	4.6	4.6
250	**250**	-0.01	0.05	-0.01	0.00	4.9	5.2	4.5	5.4
250	**500**	0.00	0.00	-0.01	-0.10	4.9	4.6	5.9	5.4
250	**1000**	-0.01	-0.02	0.00	0.02	5.5	5.2	5.5	5.4
500	**250**	-0.01	-0.01	-0.02	-0.12	4.9	5.6	4.6	5.2
500	**500**	0.00	-0.01	-0.01	-0.04	5.1	5.2	4.7	4.3
500	**1000**	-0.02	-0.03	0.01	0.03	4.9	4.8	4.7	5.2
		RMSE(x100)				**Power (5% level, ×100)**			
100	**250**	1.24	3.35	1.43	10.52	100.0	84.9	100.0	20.8
100	**500**	0.86	2.27	1.01	7.37	100.0	98.8	100.0	30.0
100	**1000**	0.61	1.62	0.70	5.19	100.0	100.0	100.0	49.6
250	**250**	0.77	2.09	0.86	6.35	100.0	99.8	100.0	35.3
250	**500**	0.53	1.40	0.62	4.55	100.0	100.0	100.0	61.4
250	**1000**	0.39	1.02	0.45	3.28	100.0	100.0	100.0	85.9
500	**250**	0.54	1.48	0.61	4.52	100.0	100.0	100.0	61.4
500	**500**	0.38	1.03	0.43	3.18	100.0	100.0	100.0	87.0
500	**1000**	0.27	0.71	0.31	2.29	100.0	100.0	100.0	98.8

Notes: See notes to Tables S1a and S1b.

S.8 Rejection frequency figures

This sections presents rejection frequency plots for the BMM, and GMM estimators in the case of Experiment 1, and the sample size combination $T = 10$ and $n = 1000$. Figures S1-S2 compare the rejection frequencies based on the BMM estimator with the first-difference GMM estimators, using the DIF1 and DIF2 moment conditions, respectively, and Figures S3-S4 compare the rejection frequencies based on the BMM estimator with the system GMM estimators, using SYS1 and SYS2 moment conditions, respectively.

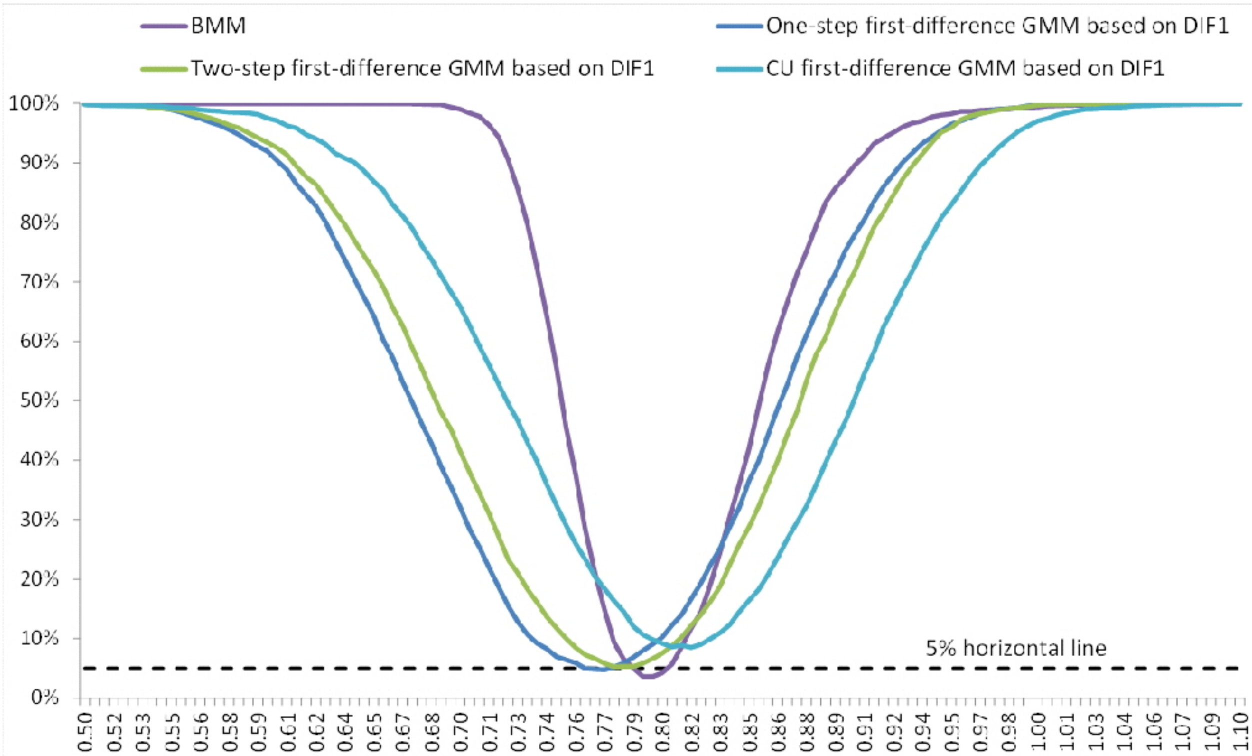

Figure S1: Rejection frequency of the tests based on the BMM and the first-difference GMM estimators based on DIF1 moment conditions in Experiment 1 ($\phi = 0.8$, $\mu_v = 0$), for sample size $n = 1000$ and $T = 10$.[S3]

[S3] Two-step GMM estimators use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators use Newey and Windmeijer (2009)'s standard errors.

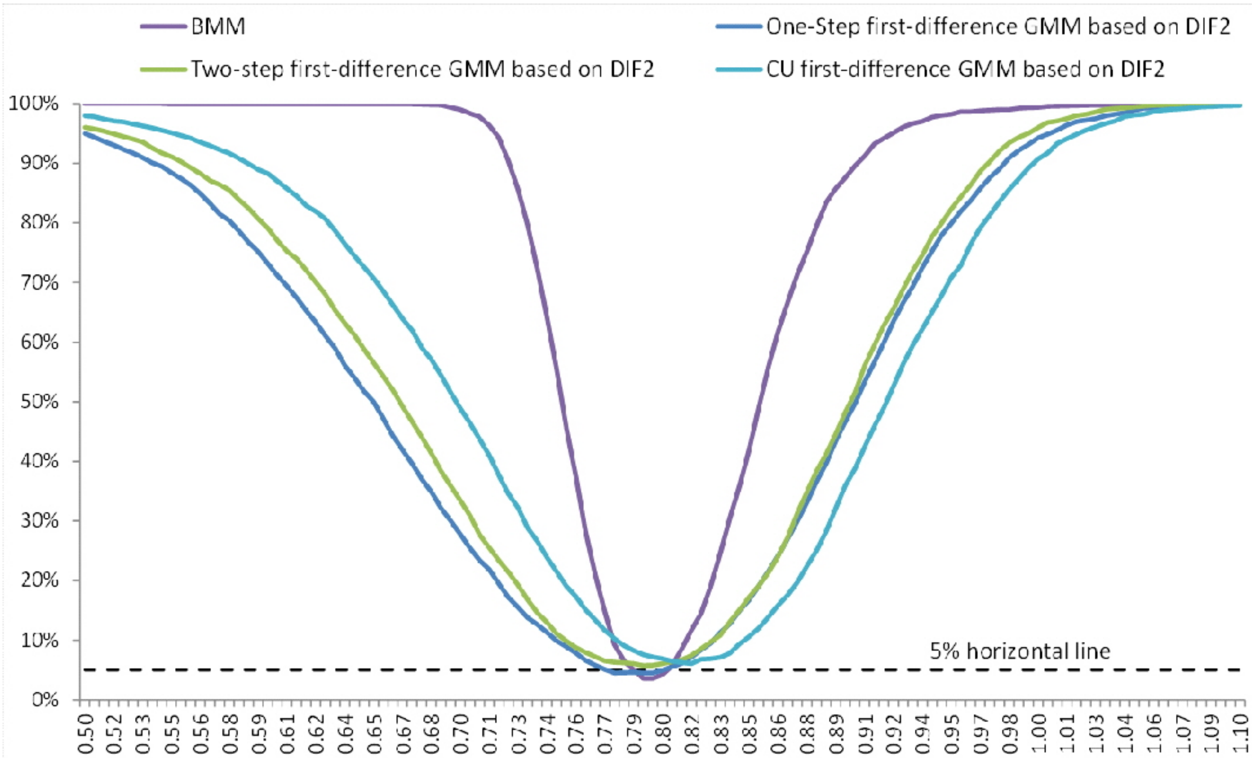

Figure S2: Rejection frequency of the tests based on the BMM and the first-difference GMM estimators based on DIF2 moment conditions in Experiment 1 ($\phi = 0.8$, $\mu_v = 0$), for sample size $n = 1000$ and $T = 10$.[S4]

[S4]Two-step GMM estimators use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators use Newey and Windmeijer (2009)'s standard errors.

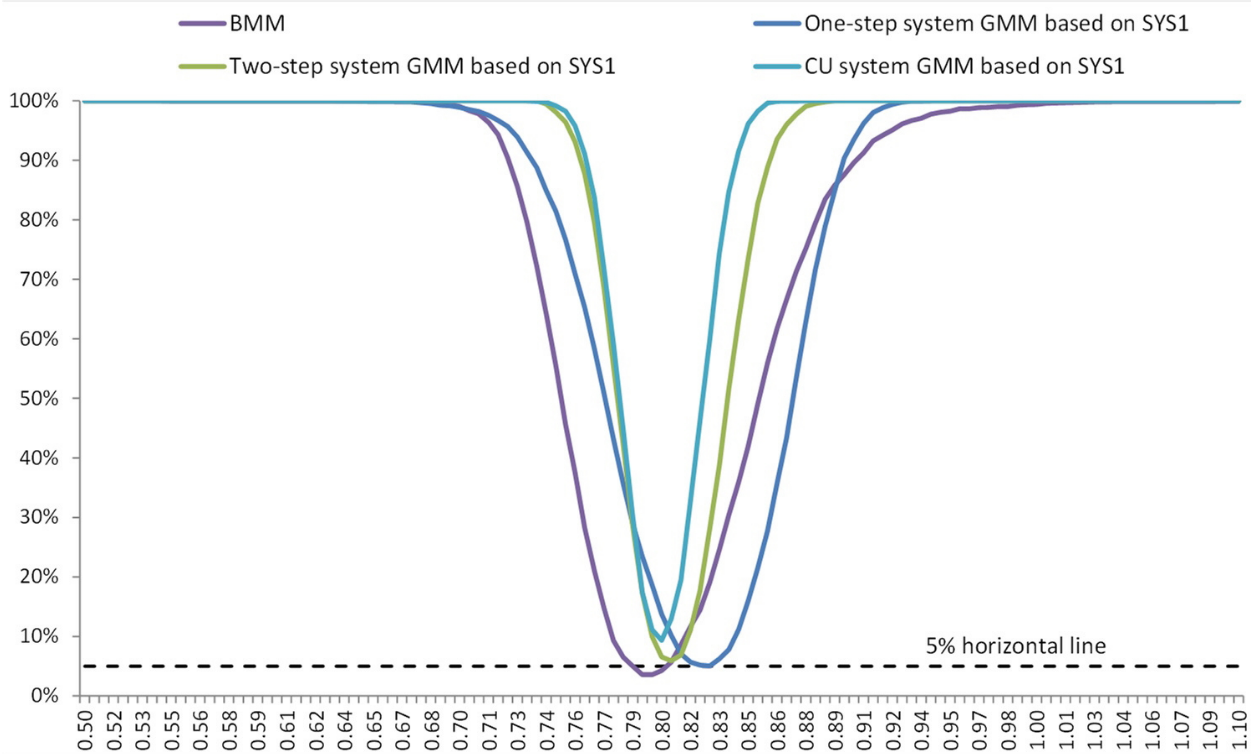

Figure S3: Rejection frequency of the tests based on the BMM and the system GMM estimators based on SYS1 moment conditions in Experiment 1 ($\phi = 0.8$, $\mu_\upsilon = 0$), for sample size $n = 1000$ and $T = 10$.[S5]

[S5] Two-step GMM estimators use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators use Newey and Windmeijer (2009)'s standard errors.

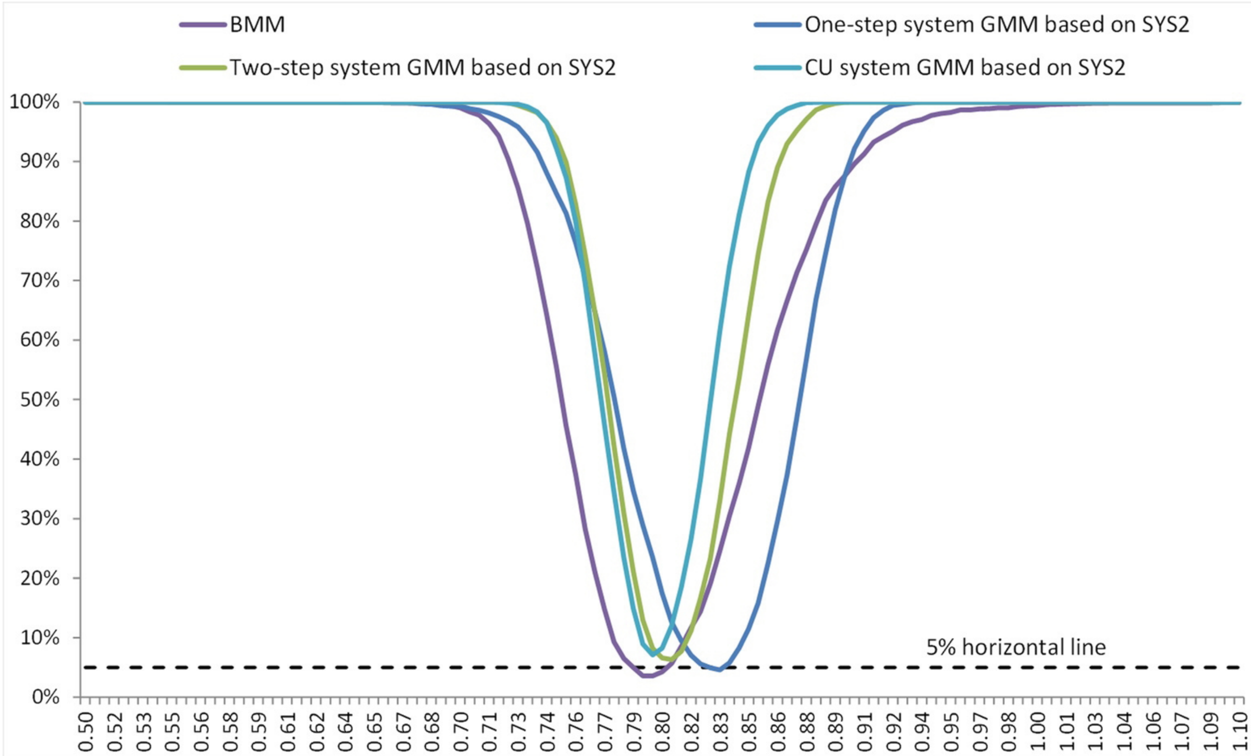

Figure S4: Rejection frequency of the tests based on the BMM and the system GMM estimators based on SYS2 moment conditions in Experiment 1 ($\phi = 0.8$, $\mu_{\upsilon} = 0$), for sample size $n = 1000$ and $T = 10$.[S6]

[S6] Two-step GMM estimators use Windmeijer (2005)'s standard errors and the continuous updating GMM estimators use Newey and Windmeijer (2009)'s standard errors.