Federal Reserve Bank of Dallas

An Augmented Anderson-Hsiao Estimator for Dynamic Short-*T* Panels

Alexander Chudik and M. Hashem Pesaran

Globalization Institute Working Paper 327 **March 2020**

Research Department

https://doi.org/10.24149/gwp327r1

An Augmented Anderson-Hsiao Estimator for Dynamic Short-*T* Panels[*]

Alexander Chudik[†] and M. Hashem Pesaran[‡]

March 13, 2020

Abstract

This paper introduces the idea of self-instrumenting endogenous regressors in settings when the correlation between these regressors and the errors can be derived and used to bias-correct the moment conditions. The resulting bias-corrected moment conditions are less likely to be subject to the weak instrument problem and can be used on their own and/or augmented with other available moment conditions (if any) to obtain more efficient estimators. This approach can be applied to estimation of a variety of models such as spatial and dynamic panel data models. This paper focuses on the latter, and proposes a new estimator for short-*T* dynamic panels by augmenting Anderson and Hsiao (AAH) estimator with bias-corrected quadratic moment conditions in first differences which substantially improve the small sample performance of the AH estimator without sacrificing on the generality of its underlying assumptions regarding the fixed effects, initial values, and heteroskedasticity of error terms. Using Monte Carlo experiments it is shown that the AAH estimator represents a substantial improvement over the AH estimator and more importantly it performs well even when compared to Arellano and Bond and Blundell and Bond (BB) estimators that are based on more restrictive assumptions, and continues to have satisfactory performance in cases where the standard GMM estimators are inconsistent. Finally, to decide between AAH and BB estimators, we also propose a Hausman type test which is shown to work well when *T* is small and n sufficiently large.

Keywords: Short-*T* Dynamic Panels, GMM, Bias-Corrected Moment Conditions, BMM, Self-Instrumenting, Nonlinear Moment Conditions, Panel VARs, Hausman Test, Monte Carlo Evidence.

JEL Classification: C12, C13, C23.

---

[*] We would like to thank Seung Ahn, Maurice Bun, Geert Dhaene, Brian Finley, Everett Grant, Kazuhiko Hayakawa, Cheng Hsiao, Andrea Nocera, Vasilis Sarafidis, Vanessa Smith, Ron Smith, Martin Weidner, conference participants at the June 2016 and 2017 IAAE annual conferences, participants of the Workshop on Recent Developments in Time Series and Panel Econometrics at the University of Koln (5 February 2020) and participants at the Federal Reserve Bank of Dallas research department seminar for helpful comments on earlier versions of this paper. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas. This research was supported in part through computational resources provided by the Big-Tex High Performance Computing Group at the Federal Reserve Bank of Dallas. A previous version of this paper was circulated under the title "A Bias-Corrected Method of Moments Approach to Estimation of Dynamic Short-*T* Panels," Federal Reserve Bank of Dallas, Globalization and Monetary Policy Working Paper No. 327, https://doi.org/10.24149/gwp327, and CESifo Working Paper Series No. 6688, available at SSRN: https://ssrn.com/abstract=3072243.

[†] Alexander Chudik, Federal Reserve Bank of Dallas, Alexander.Chudik@dal.frb.org.

[‡] M. Hashem Pesaran, University of Southern California, USA, and Trinity College, Cambridge, UK.

1 Introduction

Analysis of linear dynamic panel data models where the time dimension (T) is short relative to the cross section dimension (n), plays an important role in applied 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 advanced in the literature.[2] The GMM methods apply to autoregressive (AR) panels as well as to AR panels augmented with strictly or weakly exogenous regressors and are developed under fairly general conditions, which is important for applied work. However, the GMM methods are subject to a number of well-known drawbacks. Anderson and Hsiao (1981 and 1982)'s estimator of AR(1) panels has poor small sample performance due to weak correlations between the regressors and the instruments when the autoregressive coefficient is moderately large. Subsequently proposed GMM estimators have better small sample performance but at the cost of more restricted assumptions. The popular first-difference GMM estimator due to Arellano and Bond (1991) uses lagged levels rather than first-differences as instruments, and the system GMM approach by Blundell and Bond (1998) considers additional moment conditions that help identification but impose strong requirements on the initialization of the dynamic processes. In particular, as discussed in Section 2, the system GMM approach does not allow for the initial values to systematically differ from the long-run means.

This paper proposes a novel idea of self-instrumenting the endogenous regressors in settings where the correlation between the regressors and the errors can be derived instead of searching for instruments that are uncorrelated with the error terms. The resulting 'bias-corrected' moment conditions are less likely to be subject to the weak instrument problem and can be used on their own and/or augmented with other available moment conditions (if any) to obtain more efficient estimators. Our idea differs from the wide variety of the bias-corrected estimation methods in the literature, which correct a first-stage estimator for small-T bias and tend to be applicable under

[1] Other approaches in the literature include the likelihood-based methods (Hsiao et al., 2002, Lancaster, 2002, Moral-Benito, 2013, Hayakawa and Pesaran, 2015, and Dhaene and Jochmans, 2016), X-differencing method (Han et al., 2014), factor-analytical method (Bai, 2013), and bias-correction methods mentioned below.

[2] Anderson and Hsiao (1981 and 1982), Holtz-Eakin et al. (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), and Hayakawa (2012), among others. A recent contribution by Breitung, Hayakawa, and Kripfganz (2019) is also an interesting addition to the GMM literature. Their bias-corrected methods of moments estimator requires homoskedastic errors over time.

more restrictive assumptions.[3] Instead of correcting the bias of standard GMM estimators, we consider correcting the 'bias' of the moment conditions *before* estimation, when possible. The idea of self-instrumenting has wide-ranging applications for robust estimation and inference in settings where the correlation between the regressors and the errors can be derived. This paper focuses on dynamic panels. Another application is the estimation of spatial panel data models which is pursued in Pesaran and Yang (2020).

By self-instrumenting lagged differences, we develop a simple bias-corrected methods of moment (BMM) estimator under general conditions on initialization of the underlying dynamics, individual effects, with (possibly) heteroskedastic error variances over time as well as cross-sectionally. The resultant moment conditions turn out to be quadratic, and only reduce to linear moment conditions if the underlying AR processes are stationary. In this special case we show the BMM estimator to be identical to the first difference least square estimator proposed by Han and Phillips (2010). We establish consistency and asymptotic normality of the BMM estimator under general conditions and discuss its relation to a variety of GMM estimators proposed in the literature. These results help illustrate the important role played by the initialization of the AR processes in the case of short T panels.

We also consider augmenting the bias-corrected moment conditions with other moment conditions available in the literature, and for maximum robustness to assumptions regrading individual effects and initial values we focus on Anderson and Hsiao type moment conditions. Accordingly, we propose a new augmented Anderson and Hsiao (AAH) estimator which substantially improve the small sample performance of the AH estimator without sacrificing on the generality of its underlying assumptions. The AAH estimator holds under less restrictive conditions imposed by other prevalent GMM estimators proposed by Arellano and Bond (AB), and Blundell and Bond (BB) in the literature, and is more generally applicable. For testing the validity of the BB moment conditions, we consider a Hausman type test based on the difference between BB and AAH estimators.

Monte Carlo (MC) experiments document AAH's good small sample performance in comparison

[3] See, for example, methods based on exact analytical bias formula or its approximation, Bruno (2005), Bun (2003), Bun and Carree (2005, 2006), Bun and Kiviet (2003), Hahn and Kuersteiner (2002), Hahn and Moon (2006), Juodis (2013), and Kiviet (1995, 1999); simulation-based bias-correction methods by Everaert and Ponzi (2007), and Phillips and Sul (2003, 2007); the jackknife bias corrections by Dhaene and Jochmans (2015), and Chudik, Pesaran, and Yang (2018); or the recursive mean adjustment correction procedures, Choi et al. (2010)). 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 GMM estimators considered in this paper and bias correction estimators is a welcome addition to the literature but lies beyond the scope of the present paper.

with a number of GMM alternatives. Perhaps not surprisingly the AAH estimator represents a substantial improvement over the AH estimator across all designs considered. When compared to AB and BB estimators, the AAH is less efficient in designs that satisfy the more restrictive assumption that underlie BB estimators, but continues to perform well even in cases where the system-GMM type estimators are inconsistent. The robustness of the AAH estimator is an important advantage since in practice it is not known if the additional restrictions of the AB and BB estimators are met, and it is therefore desirable to consider estimation procedures that are robust to violation of such restrictive assumptions. In this regard, the AAH estimator is a useful addition to the literature.

The remainder of this paper is organized as follows. Section 2 sets up the baseline panel AR(1) model and discusses AH and subsequent GMM moment conditions. Section 3 introduces the main idea and presents a simple BMM estimator. Section 4 introduces the AAH estimator and discusses the related literature, in particular Ahn and Schmidt (1995, 1997). Section 5 discusses extensions of AAH estimator to ARX and VAR short-T panel data models. Section 6 discusses the problem of moment proliferation and adopts the One Covariate at the time Multiple Testing approach by Chudik, Kapetanios, and Pesaran (2018) for selection of relevant subset of AAH moments for estimation and inference. Section 7 presents MC evidence, and the last section concludes and discusses avenues for future research. Further results and discussions are provided in an Appendix.

2 Panel AR(1) model and assumptions

We begin with a simple panel AR(1) model to set out the main idea. 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.*[4]

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

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

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, $E(u_{it}^2) = \sigma_{it}^2$, and there exist positive constants*

[4] 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.

c and K such that $0 < c < \sigma^2_{it} < K$. Moreover, $\bar{\sigma}^2_{tn} \equiv n^{-1}\sum_{i=1}^{n}\sigma^2_{it} \to \bar{\sigma}^2_t$ as $n \to \infty$, and $\sup_{it} E\left|u_{it}\right|^{4+\epsilon} < K$ for some $\epsilon > 0$. For each t, u_{it} is independently distributed over i. For each i, u_{it} is serially uncorrelated over t.

ASSUMPTION 3 *(Initialization and individual effects) Let $b_i \equiv -\phi^{m_i}\left[(1-\phi)\, y_{i,-m_i} - \alpha_i\right]$ and $\varsigma^2_i = E\left(b^2_i\right)$. Then $\bar{\varsigma}^2_n \equiv n^{-1}\sum_{i=1}^{n}\varsigma^2_i \to \bar{\varsigma}^2$ as $n \to \infty$, 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 $t = -m_j + 1, -m_j + 2, ..., 1, 2, ..., T$, 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.*

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

Remark 3 *Assumption 3 allows for $E\left(b_i\right)$ to vary across i, and therefore, in view of (3)-(4), $E\left(\Delta y_{it}\right)$ 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 the 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.[5]

In addition to (8), the system GMM approach considered by Arellano and Bover (1995) and Blundell and Bond (1998) also requires that[6]

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

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

[5] 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 $E\left(v_i\right)=0$, and $E\left(v_i\Delta u_{it}\right)=0$, 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.

[6] 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.

have[7]

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

Therefore, when m_i is finite for the moment conditions (9) to hold 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 violated 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 a 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 estimators, it is also worth noting from (10) that if $|\phi|<1$ and $\{y_{it}\}$ are initialized in a distant past (with $m_i\rightarrow\infty$), then Δy_{it} will no longer depend on α_i and renders the BMM and Anderson-Hsiao 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 ratio $\sum_{i=1}^{n}Var\left(\alpha_i\right)/\sum_{i=1}^{n}Var\left(u_{it}\right)$. 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 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

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

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,

$$E\left(\Delta u_{it}\Delta y_{i,t-1}\right) + \sigma^2_{i,t-1} = 0, \text{ for } i = 1, 2, ..., n, \text{ and } t = 2, 3, ..., T-1. \tag{12}$$

To solve for σ^2_{it}, we note that $E\left(\Delta u_{it}\right)^2 = \sigma^2_{i,t-1} + \sigma^2_{it}$ and $E\left(\Delta u_{i,t+1}\Delta y_{it}\right) = -\sigma^2_{it}$. Hence, $\sigma^2_{i,t-1} = E\left(\Delta u_{it}\right)^2 + E\left(\Delta u_{i,t+1}\Delta y_{it}\right)$, 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 $\sigma^2_{i,t-1} = E\left(\Delta u_{it}\right)^2 + E\left(\Delta u_{i,t+1}\Delta y_{it}\right)$ 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 $\sigma^2_{it} = \sigma^2_{i,t-1} = \sigma^2_i$. Another possibility is to assume covariance stationarity of y_{it}, which will lead to a linear moment condition solution, discussed in Remark 5 below.[8]

Initially, we use the QM condition (13) alone to obtain an estimator of ϕ. We propose averaging (13) over i and t, which will deliver a simple exactly identified moment estimator. In Section 4, we consider optimally weighting the moment conditions in (13), and 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

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

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

where

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

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 $\hat{\phi}_{nT}$, 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

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

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

$$Q_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta y_{i,t-1}^2,\ Q_{iT}^{+} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta y_{it}^2, \text{ and } H_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Delta u_{it}\Delta y_{i,t-1}. \tag{20}$$

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

$$\left(\hat{\phi}_{nT} - \phi_0\right)\left[\left(\hat{\phi}_{nT} - \phi_0\right)\bar{Q}_{nT} - \bar{B}_{nT}\right] = -\bar{V}_{nT}, \tag{21}$$

where $\bar{V}_{nT} = n^{-1}\sum_{i=1}^{n} V_{iT}$, $\bar{Q}_{nT} = n^{-1}\sum_{i=1}^{n} Q_{iT}$, and

$$\bar{B}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{iT} + Q_{iT}^{+} + 2H_{iT}\right). \tag{22}$$

Using results (A.4)-(A.5) of Lemma A.1 in the appendix, under Assumptions 1-3, we have (for a fixed T)

$$\bar{Q}_{nT} = E\left(\bar{Q}_{nT}\right) + O_p\left(n^{-1/2}\right), \text{ and } \bar{B}_{nT} = E\left(\bar{B}_{nT}\right) + O_p\left(n^{-1/2}\right), \tag{23}$$

where

$$E\left(\bar{Q}_{nT}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(Q_{iT}\right) > 0. \tag{24}$$

In addition, using result (A.6) of Lemma A.2 in the appendix, we have

$$\bar{V}_{nT} = O_p\left(n^{-1/2}\right). \tag{25}$$

We now use (21) 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 (21), we have

$$\sqrt{n}\left(\hat{\phi}_{nT} - \phi_0\right)^2 \bar{Q}_{nT} - \sqrt{n}\left(\hat{\phi}_{nT} - \phi_0\right)\bar{B}_{nT} = -\sqrt{n}\bar{V}_{nT}. \tag{26}$$

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

$$\bar{B}_{nT}\sqrt{n}\left(\hat{\phi}_{nT} - \phi_0\right) = \sqrt{n}\bar{V}_{nT} + O_p\left(n^{-1/2}\right). \tag{27}$$

Also, using (23) 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 (25), $\sqrt{n}\bar{V}_{nT} = O_p(1)$. If

$$\bar{B}_{T} = \lim_{n\to\infty} E\left(\bar{B}_{nT}\right) \neq 0, \tag{28}$$

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.7) of Lemma A.2 in the appendix, we have $\sqrt{n}\bar{V}_{nT} \rightarrow_d N(0, S_T)$, where $S_T = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(V_{iT}^2\right)$, and it follows that $\sqrt{n}\left(\hat{\phi}_{nT}-\phi_0\right) \rightarrow_d N(0,\Sigma_T)$ 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 $\bar{B}_T \neq 0$. But using (20) in (22) we have $\bar{B}_T = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} B_{iT}$, where

$$B_{iT} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta y_{i,t-1}^2 + \Delta y_{it}^2 + 2\Delta u_{it}\Delta y_{i,t-1}\right). \tag{29}$$

It is now 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

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

where $\bar{\sigma}^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_i^2$. 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. As a simple example consider the case where $T = 3$, and note that (see Section A.1 of the Appendix)

$$\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2 + (1-\phi)^2\bar{\sigma}_1^2 + \left(1+\phi^2\right)(1-\phi)\psi_0. \tag{30}$$

where $\bar{\sigma}_t^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sigma_{it}^2$, and

$$\psi_0 = (1-\phi)\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left(y_{i0}-\mu_i\right)^2 - 2\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left[u_{i1}\left(y_{i0}-\mu_i\right)\right]. \tag{31}$$

If $\phi = 1$, then $\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2$, and $\bar{B}_3 \neq 0$, if and only if $\bar{\sigma}_1^2 \neq \bar{\sigma}_2^2$. When $|\phi| < 1$, $\bar{B}_3 \neq 0$ even if $\bar{\sigma}_1^2 = \bar{\sigma}_2^2$, except for when $(1-\phi)\left(1+\phi^2\right)\psi_0 = \phi(2-\phi)\bar{\sigma}_1^2 - \bar{\sigma}_2^2$. 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.

The following theorem summarizes the main results established above.

Theorem 1 *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 $y_{i,-m_i}$, and the true value of the parameter of interest ϕ_0. Let Assumptions 1-3 hold, and suppose $\bar{B}_T \neq 0$ and $n^{-1}\sum_{i=1}^n E\left(V_{iT}^2\right) \to S_T > 0$, where $\bar{B}_T$ is given by (28) and V_{iT} is defined in (19). Consider the BMM estimator $\hat{\phi}_{nT}$ given by (16). Let T be fixed and $n \to \infty$. Then, the unique $\sqrt{n}$-consistent estimator $\hat{\phi}_{nT}$ satisfies

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

where

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

Remark 4 *When $\bar{B}_T = 0$, from (21) 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 $\bar{Q}_{nT} \to \bar{Q}_T > 0$ as $n \to \infty$, 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.

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

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

where

$$\hat{\bar{B}}_{nT} = \frac{1}{n}\sum_{i=1}^n \left(Q_{iT} + Q_{iT}^{+} + 2\hat{H}_{i,nT}\right), \tag{35}$$

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

$$\hat{V}_{i,nT} = -\frac{1}{T-2}\sum_{t=2}^{T-1}\left(\Delta\hat{u}_{it}\Delta y_{i,t-1} + \Delta\hat{u}_{it}^2 + \Delta\hat{u}_{i,t+1}\Delta y_{it}\right). \tag{36}$$

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 $E\left(\Delta y_{it}^2\right) = 2\sigma_i^2/\left(1+\phi\right)$ is time-invariant. Under these restrictions $\sigma_i^2 = (1+\phi)\,E\left(\Delta y_{i,t-1}^2\right)/2$, $E\left(\Delta u_{it}\Delta y_{i,t-1}\right) = E\left(\Delta u_{i,t+1}\Delta y_{it}\right)$, and using (12)* *the quadratic moment condition, (13), simplifies to the following linear moment condition:*

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

which yields the associated BMM estimator

$$\hat{\phi}_n=\frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\left(2\Delta y_{it}\Delta y_{i,t-1}+\Delta y_{i,t-1}^{2}\right)}{\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{i,t-1}^{2}}. \tag{37}$$

Note that in this case ϕ is identified even when $T=2$. Interestingly enough, the above linear BMM estimator is identical to the first-difference least square (FDLS) estimator proposed by Han and Phillips (2010).[9] *As discussed by Han and Phillips (2010), $\hat{\phi}_n$ given by (37) has standard Gaussian asymptotics for all values of $\phi\in(-1,1]$ and does not suffer from the weak instrument problem. 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 it is not possible to rule out the heteroskedasticity of error variances over t.*

4 Augmented Anderson Hsiao (AAH) estimator

The BMM estimator above is useful for illustrative purposes, but it is not asymptotically efficient partly due to averaging of moment conditions over t, and more importantly due to not exploring additional readily available moment conditions that hold under the same set of assumptions. As noted above, amongst the moment conditions proposed in the literature, only the ones proposed by AH are sufficiently general, and accordingly, we propose to augment the $T-2$ QM condition (13) with the $(T-2)(T-1)/2$ AH moment conditions (7). These provide $(T-2)(T-1)/2+T-2$ AAH moment conditions in total. As usual, we can obtain first, second and cumulative updating GMM estimators based on these quadratic-linear moment conditions.

Remark 6 *It is clear that conditions (7) and (13) do not imply conditions (8) and/or (9) since (7) and (13) rely only on first differences, whereas (8) and (9) also rely on levels. Hence, it is possible that (7) and (13) can hold whilst (8) and/or (9) might not hold. An example of this case*

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

is discussed and explored in Monte Carlo section below.

The set of AAH moment conditions (7) and (13) is a subset of the conditions listed in Ahn and Schmidt (1995, 1997), who explored a complete set of moments conditions under stronger set of assumptions than are necessary for AAH alone, see their Assumptions SA1-SA3. Sufficient set of assumptions that give rise to AAH are the following 'basic' assumptions:

(BA1) For all i, the u_{it} are mutually uncorrelated.

(BA2) $E\left[\left(y_{i0}-\mu_i\right)\Delta u_{it}\right]=0$ for all i and $t=2,3,...,T$,

where $\mu_i=\alpha_i/\left(1-\phi\right)$ is the long-run mean. Assumption BA1 on its own has been considered as Case H of Ahn and Schmidt (1997), which implies $T\left(T-3\right)/3$ moment conditions. Assumption BA2 is implied by assumptions SA1-SA2 of Ahn and Schmidt (1995), but not *vice versa*. The full set of moment conditions based on BA1 and BA2 is the union of AH moment conditions given by (7) and QM moment conditions given by (13). Derivation of the asymptotic distribution and conducting inference requires additional standard high-level regularity conditions routinely used in the GMM literature.[10]

It is of interest to consider the efficiency loss that arises when using AAH moment conditions, whilst in fact the more restrictive system GMM conditions (8)-(9) hold. To shed light on this, we report the ratios of asymptotic variances of the AH, first-difference GMM and system GMM estimators, all relative to that of the AAH estimator. We illustrate the asymptotic efficiency gains and losses in Table 1 in the same way as in Ahn and Schmidt (1995). We are interested in two questions: (i) How much is gained by adding QM conditions to AH, and (ii) how much is lost by not utilizing the additional moment conditions assuming that the DGP satisfies all of the restrictions in (8) and (9). Following Ahn and Schmidt (1995), we tabulate the asymptotic variance ratios for the stationary homoskedastic case for different values of ϕ, and different ratios of $E\left(\alpha_i^2\right)/E\left(u_{it}^2\right)=\sigma_\alpha^2/\sigma_u^2$, for all i and t. The results, computed by simulations, are summarized in Table 1.

As can be seen from Table 1, augmenting AH moment conditions with the quadratic moment conditions (13) results in substantial efficiency gains for all values of ϕ, $\sigma_\alpha^2/\sigma_u^2$ and the three choices

[10] These are listed, for example, in Pesaran (2015). In particular, assumptions for consistency are given by Assumptions A1 and A2 in Chapter 10 of Pesaran (2015) and the additional assumptions for asymptotic normality are given by Assumptions A3-A5 of the same chapter. See also Assumptions 1-3 for a set of low-level assumptions required for consistency and asymptotic normality.

of $T = 3, 6$ and 10, being considered. The efficiency gains are particularly pronounced for values of ϕ close to unity. Also as to be expected the two estimators perform equally well for all values of $\sigma_\alpha^2/\sigma_u^2$ since both use first-differences as instruments and hence are not affected by σ_α^2. The efficiency gain of AAH over AH reduces somewhat when T is increased.

Table 1: Asymptotic efficiency of AH, AB and BB estimators relative to the AAH estimator under stationarity

	$var(AH)/var(AAH)$			$var(AB)/var(AAH)$			$var(BB)/var(AAH)$		
	$\sigma_\alpha^2/\sigma_u^2$			$\sigma_\alpha^2/\sigma_u^2$			$\sigma_\alpha^2/\sigma_u^2$		
ϕ	0.5	1	4	0.5	1	4	0.5	1	4
$T=3$									
-0.9	1.5	1.5	1.5	1.0	1.0	1.1	0.9	1.0	1.0
-0.8	1.7	1.7	1.7	1.0	1.1	1.2	0.9	1.0	1.0
-0.5	2.3	2.3	2.3	1.1	1.2	1.6	0.9	0.9	1.0
-0.3	2.9	2.9	2.9	1.2	1.3	1.9	0.9	0.9	1.0
0	4.0	4.0	4.0	1.2	1.5	2.4	0.8	0.9	0.9
0.3	5.3	5.3	5.3	1.2	1.7	3.0	0.6	0.7	0.9
0.5	6.4	6.4	6.4	1.2	1.7	3.4	0.4	0.5	0.7
0.8	8.2	8.2	8.2	1.2	1.7	4.1	0.1	0.1	0.3
0.9	9.2	9.2	9.2	1.9	2.6	3.8	0.05	0.05	0.06
$T=6$									
-0.9	1.3	1.3	1.3	1.0	1.0	1.1	1.0	1.0	1.0
-0.8	1.3	1.3	1.3	1.1	1.1	1.1	1.0	1.0	1.0
-0.5	1.6	1.6	1.6	1.2	1.2	1.4	1.0	1.0	1.0
-0.3	1.8	1.8	1.8	1.3	1.4	1.6	1.0	1.0	1.0
0	2.2	2.2	2.2	1.5	1.6	2.0	1.0	1.0	1.0
0.3	3.0	3.0	3.0	1.7	2.0	2.6	0.9	1.0	1.0
0.5	3.9	3.9	3.9	2.0	2.4	3.3	0.9	0.9	1.0
0.8	6.1	6.1	6.1	2.5	3.5	5.1	0.5	0.6	0.8
0.9	7.6	7.6	7.6	2.5	4.0	6.3	0.2	0.3	0.5
$T=10$									
-0.9	1.2	1.2	1.2	1.0	1.0	1.1	1.0	1.0	1.0
-0.8	1.2	1.2	1.2	1.1	1.1	1.1	1.0	1.0	1.0
-0.5	1.3	1.3	1.3	1.1	1.2	1.2	1.0	1.0	1.0
-0.3	1.5	1.5	1.5	1.2	1.3	1.4	1.0	1.0	1.0
0	1.7	1.7	1.7	1.4	1.5	1.6	1.0	1.0	1.0
0.3	2.2	2.2	2.2	1.6	1.8	2.0	1.0	1.0	1.0
0.5	2.7	2.7	2.7	1.9	2.2	2.5	1.0	1.0	1.0
0.8	4.8	4.8	4.8	2.9	3.6	4.3	0.8	0.9	0.9
0.9	6.5	6.5	6.5	3.6	4.6	5.8	0.5	0.6	0.8

Notes: This table reports ratios of asymptotic variance of the Anderson and Hsiao (AH), Arellano and Bond (AB) and Blundell and Bond (BB) estimators relative to the asymptotic variance of the augmented AH (AAH) estimator in a stationary design with $E(\alpha_i^2) = \sigma_\alpha^2$ and $E(u_{it}^2) = \sigma_u^2$, and for different values of the AR coefficient, ϕ. Asymptotic variances are computed by simulations.

Turning now to the second issue, namely efficiency loss of AAH relative to AB and BB estimators, we first note that interestingly enough, the expected efficiency gain of AB over AAH does not materialize and AAH is in fact generally more efficient than the AB estimator, with efficiency gain of AAH increasing substantially as larger values of ϕ and $\sigma_\alpha^2/\sigma_u^2$ are considered. Increasing T does not seem to have much effect on the relative efficiency of the AB estimator. The results in Table 1 also confirm the sensitivity of the AB estimator to the ratio, $\sigma_\alpha^2/\sigma_u^2$. In contrast, the BB estimator performs favorably relative to the AAH estimator (and by implication relative to the AB estimator) particularly, for values of ϕ close to unity. However, this efficiency gain is achieved assuming that $E\left[\mu_i\left(y_{i,-m_i}-\mu_i\right)\right]=0$, for $i=1,2,...,n$, which might not hold in practice. (see (11) and the related discussions). The cost of using BB estimator is inconsistency if condition (11) is not met. Further evidence on this is provided in the Monte Carlo section.

The above simulations suggest that AAH estimator cannot be more efficient than BB estimator when all BB moment conditions are met. This can be seen formally by investigating more closely the relation between the BB condition (9) and the QM moment condition (12), or equivalently (13). Substituting $u_{it}=\Delta u_{it}+\alpha_i+u_{i,t-1}$ in (9) we have

$$E\left[\Delta y_{i,t-1}\left(\alpha_i+u_{it}\right)\right]=E\left(\Delta y_{i,t-1}\Delta u_{it}\right)+E\left[\Delta y_{i,t-1}\left(\alpha_i+u_{i,t-1}\right)\right],$$

and since $\Delta y_{i,t-1}=\phi\Delta y_{i,t-2}+\Delta u_{i,t-1}$, then

$$E\left[\Delta y_{i,t-1}\left(\alpha_i+u_{it}\right)\right]=E\left(\Delta y_{i,t-1}\Delta u_{it}\right)+\phi E\left[\Delta y_{i,t-2}\left(\alpha_i+u_{i,t-1}\right)\right]+E\left(\Delta u_{i,t-1}\alpha_i\right)+E\left(\Delta u_{i,t-1}u_{i,t-1}\right). \tag{38}$$

But under BB moment conditions $E\left(\Delta u_{i,t-1}\alpha_i\right)=0$, as well as

$$E\left[\Delta y_{i,t-2}\left(\alpha_i+u_{i,t-1}\right)\right]=0. \tag{39}$$

Using these results in (38) we now have

$$\begin{aligned} E\left[\Delta y_{i,t-1}\left(\alpha_i+u_{it}\right)\right] &= E\left(\Delta y_{i,t-1}\Delta u_{it}\right)+E\left(\Delta u_{i,t-1}u_{i,t-1}\right) \\ &= E\left(\Delta y_{i,t-1}\Delta u_{it}\right)+\sigma_{i,t-1}^2=0, \end{aligned}$$

which is the same as the QM condition given by (12). Namely, the QM condition is implied by the BB moment conditions and not *vice versa*, and hence under BB conditions the AAH estimator cannot be more efficient than the BB estimator. Note that (39) is the same as (9) and it is satisfied if $E(\Delta u_{i,t-1}\alpha_i) = 0$ and $E[\mu_i(y_{i,-m_i} - \mu_i)] = 0$, as discussed in Section 2. However, when the BB conditions (8) and/or (9) are not met the BB estimator becomes inconsistent contrary to the AAH estimator that continues to be consistent. Therefore, the main two conditions underlying the Hausman test (Hausman, 1978) are met and the validity of BB moment conditions can be tested using the Hausman procedure. Denoting the AAH and BB estimators by $\hat{\phi}_{nT}^{aah}$ and $\hat{\phi}_{nT}^{bb}$, respectively, the Hausman test statistic is defined by

$$H_n = \left(\hat{\phi}_{nT}^{aah} - \hat{\phi}_{nT}^{bb}\right)^2 \left[\widehat{Var}\left(\hat{\phi}_{nT}^{aah}\right) - \widehat{Var}\left(\hat{\phi}_{nT}^{bb}\right)\right]^{-1}, \tag{40}$$

assuming that $\widehat{Var}\left(\hat{\phi}_{nT}^{aah}\right) - \widehat{Var}\left(\hat{\phi}_{nT}^{bb}\right) > 0$, where $\widehat{Var}\left(\hat{\phi}_{nT}^{aah}\right)$ and $\widehat{Var}\left(\hat{\phi}_{nT}^{bb}\right)$ are consistent estimators of the asymptotic variances of $\hat{\phi}_{nT}^{aah}$, and $\hat{\phi}_{nT}^{bb}$, respectively. Under the null hypothesis that the BB conditions are met, H_n is asymptotically distributed as $\chi^2(1)$, for a fixed T and as $n \to \infty$.

5 Extensions

There are two important extensions of model (1). The first extension is to allow for additional regressors. Let $\mathbf{x}_{it}$ be $k-1$ additional regressors, and consider the ARX model

$$y_{it} = \alpha_i + \phi y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}, \text{ for } i = 1, 2, ..., n, \; t = 1, 2, ..., T. \tag{41}$$

The regressors in $\mathbf{x}_{it}$ can be strictly or weakly exogenous. AAH moment conditions (7) and (13) can be augmented by the standard orthogonality for the regressors $\mathbf{x}_{it}$, as is standard in the GMM literature. This paper does not have anything new to add regarding instrumenting the regressors $\mathbf{x}_{it}$.

Remark 7 *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 consider below. It is also possible to derive the conditional model (41) 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 (41) can be obtained only under very restrictive conditions derived in the Appendix. Specifically, $\theta_{it} = \boldsymbol{\Omega}_{xx,it}^{-1}\boldsymbol{\omega}_{xy,it}$ must be time invariant, where $\boldsymbol{\omega}_{xy,it} = E(\mathbf{u}_{x,it}u_{y,it})$, $\boldsymbol{\Omega}_{xx,it}^{-1} = E\left(\mathbf{u}_{x,it}\mathbf{u}_{x,it}'\right)$, and $\mathbf{u}_{it} = \left(u_{y,it}, \mathbf{u}_{x,it}'\right)'$ are the idiosyncratic innovations in the panel VAR representation of $\mathbf{z}_{it} =$*

$(y_{it}, \mathbf{x}'_{it})'$.

The second extension we consider is a panel VAR model in the $k \times 1$ vector of variables $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$,

$$\mathbf{z}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\Phi}\mathbf{z}_{i,t-1} + \mathbf{u}_{it}, \tag{42}$$

for $t = 0, 1, 2, ..., T$, and $i = 1, 2, ..., n$, with the initial values given by $\mathbf{z}_{i,0}$, where $\boldsymbol{\alpha}_i$ is a $k \times 1$ vector of individual effects, $\boldsymbol{\Phi}$ is a $k \times k$ matrix of slope coefficients, and $\mathbf{u}_{it} = (u_{i1t}, u_{i2t}, ..., u_{ikt})'$ is a $k \times 1$ vector of idiosyncratic errors. Similarly, to the univariate case, the set of linear AH moment conditions is given by:

$$E\left(\Delta\mathbf{z}_{is}\Delta\mathbf{u}'_{it}\right) = \mathbf{0}_{k\times k}, \text{ for } i = 1, 2, ..., n,\ s = 1, 2, ..., t-2, \text{ and } t = 3, 4, ..., T, \tag{43}$$

and the QM moment conditions are given by:

$$E\left(\Delta\mathbf{u}_{it}\Delta\mathbf{z}'_{i,t-1}\right) + E\left[\Delta\mathbf{u}_{it}\Delta\mathbf{u}'_{it}\right] + E\left(\Delta\mathbf{u}_{i,t+1}\Delta\mathbf{z}'_{it}\right) = \mathbf{0}_{k\times k}, \text{ for } i = 1, 2, ..., n, \text{ and } t = 2, 3, ..., T-1. \tag{44}$$

AAH estimation of the panel VAR model can proceed based on (43) and (44), which replace (7) and (13), respectively.

6 Problem of proliferation of moment conditions

As it is well known, the number of moment conditions that underlie any of the GMM based estimation techniques discussed above (AH, AAH, AB, or BB) grow at the quadratic rate in T. Consequently, the number of moments can get quite large even for moderate values of T. Under their respective set of assumptions, these are all valid moments and their relevance (strength) varies, some of which could be weakly identifying. Unless the number of cross-section dimension, n, is sufficiently large, as compared to the number of moment conditions, $h = h(T)$, the proliferation of moments will have negative consequences for estimation and inference in finite samples. See, for instance, Anderson and Sorenson (1996), Clark (1996), and Hansen, Heaton, and Yaron (1996). The many moment problem often occurs together with the weak moment problem, but they are not necessarily the same. Han and Phillips (2006) provide a number of asymptotic theoretical results for GMM estimation that allow for the number of moments to increase with sample size, whilst moment conditions may only be weakly identifying, encompassing earlier contributions by Bekker

(1994), Staiger and Stock (1997), Stock and Wright (2000), and Chao and Swanson (2003), among others. GMM estimators utilizing many weak moment conditions may not be consistent and the rate of convergence could depend not only on the sample size, but also on the number and quality of the moment conditions.

Hsiao and Zhang (2015) show that the AB estimator is asymptotically biased if $T/n \to c$, for some $0 < c < \infty$, as $n, T \to \infty$. This bias can be reduced using jackknife instrumental variables estimation (JIVE), which has been considered in a general GMM framework by Angrist, Imbens, and Krueger (1999), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Hansen and Kozbur (2014), Lee, Moon, and Zhou (2017), Phillips and Hale (1977) and Zhang and Zhou (2020).[11] Koenker and Machado (1999) and Donald, Imbens, and Newey (2003) consider GMM estimation under a large number of strong moments, and provide conditions on the number of moments that permits the usual asymptotic theory and inference. In particular, Koenker and Machado (1999) show $h^3/n \to 0$ is sufficient for validity of conventional GMM asymptotic inference.

There are two approaches to dealing with a large number of valid moments. One is to use them all, but combine them in such a way that allows for the number of moments to be large relative to the sample size so that consistency and valid inference are achieved. The second approach is to select and use only a subset of available moments. Contributions to this strand of the literature includes Donald and Newey (2001), Kuersteiner (2002), Hall and Peixe (2003), Inoue (2006), Hall, Inoue, Jana, and Shin (2007), and Donald, Imbens, and Newey (2009).[12] In what follows we propose a new sub-set selection procedure by adapting the One Covariate at the time Multiple Testing (OCMT) recently developed by Chudik, Kapetanios, and Pesaran (2018) for variable selection to the problem of moment selection in the case of the AAH estimator.

6.1 Moment selection using OCMT approach

In the case of AH moments listed in (7), there are $t-2$ instruments for $\Delta y_{i,t-1}$, for $t = 3, 4, ..., T$. We collect them in the set $\mathcal{S}_{i,t-2} = \{\Delta y_{i,1}, \Delta y_{i,2}, ..., \Delta y_{i,t-2}\}$. In general, it is not possible to derive

[11] Monte Carlo findings reported in Zhang and Zhou (2020) suggest very good size performance of JIVE corrected AB GMM estimator. However, the size reported in Zhang and Zhou (2020) is computed using standard deviation of the estimated slope coefficients across Monte Carlo replications, as opposed to conducting the empirically relevant tests, where standard deviation of slope coefficients is estimated for each replication. Hence the findings in Zhang and Zhou (2020) are not indicative of inference that can be conducted in empirical applications.

[12] In addition to the literature on selecting relevant moment from a set valid moments, there is a vast literature on moment validity, and the selection of valid moments, including Andrews (1999), Andrews and Lu (2001), Chatelain (2007), and Liao (2013). A problem of selecting valid as well as relevant moments has been consideed by Cheng and Liao (2015).

analytical expressions for the correlation of the target variable $\Delta y_{i,t-1}$ and individual instruments in $\mathcal{S}_{i,t-2}$ in the case where the underlying dynamic processes are initialized from finite pasts, and little is known about the data generating processes for the initial values. It is, nevertheless, possible to show that $corr\left(\Delta y_{i,t-1}, \Delta y_{i,t-\ell}\right)$ declines in ℓ at an exponential rate in the case of stationary initial values. This is illustrated in the following example.

Example 1 *Let $y_{it} = \alpha_i + \phi y_{i,t-1} + u_{it}$, for $t = ..., -1, 0, 1, ..., T$ and $i = 1, 2, ..., n$, where $|\phi| < 1$. Then $y_{it} = \mu_i + \sum_{\ell=0}^{\infty} \phi^{\ell} u_{i,t-\ell}$, and*

$$\Delta y_{it} = u_{it} + \sum_{\ell=1}^{\infty} \phi^{\ell-1} \left(\phi - 1\right) u_{i,t-\ell},$$

where $\mu_i = \alpha_i / \left(1 - \phi\right)$. Provided $E\left(u_{it} u_{it'}\right) = 0$ for $t \neq t'$ and $E\left(u_{it}^2\right)$ is bounded, it follows that $\left|corr\left(\Delta y_{i,t-1}, \Delta y_{i,t-\ell}\right)\right| < K \phi^{|\ell-1|}$.

Hence it could be the case that some of the $t-2$ instruments in $\mathcal{S}_{i,t-2}$ are rather weak and consequently not very useful in improving the asymptotic variance of the resulting GMM estimator. Our suggestion is, for each $t = 4, 5, ..., T$, to apply OCMT method to select the relevant instruments from the set $\mathcal{S}_{i,t-2}$. It is desirable to always include $\Delta y_{i,t-2}$, which is likely to have the largest correlation with the target variable $\Delta y_{i,t-1}$, as a conditioning (or pre-selected) variable in the OCMT procedure, as described below.

OCMT algorithm for selecting AH instruments for a given t $(= 4, 5, ..., T)$

1. Estimate the $(t-1)$ individual first stage regressions

$$\Delta y_{i,t-1} = a_{\ell} + \beta_{\ell} \Delta y_{i,t-2} + \theta_{\ell} \Delta y_{i,\ell}, \text{ for } \ell = t-3, t-4, ..., 1 \tag{45}$$

   by least squares and compute the associated t-ratios for the coefficients θ_{ℓ} in the above regression, denoted as $t_{\hat{\theta}_{\ell}(s)} = \hat{\theta}_{\ell} / s.e.\left(\hat{\theta}_{\ell}\right)$ for stage $s = 1$. The first stage OCMT selection indicator is given by

$$\widehat{\mathcal{J}}_{\ell,(1)} = I[|t_{\hat{\theta}_{\ell}(1)}| > c_p\left(t-1, \delta\right)], \text{ for } \ell = 1, 2, ..., t-2, \tag{46}$$

where $c_p(t,\delta)$ is a critical value function defined by

$$c_p\left(t,\delta\right)=\Phi^{-1}\left(1-\frac{p}{2t^{\delta}}\right), \tag{47}$$

$\Phi^{-1}\left(.\right)$ is the inverse of standard normal distribution function, $0<p<1$, and $\delta>0$. Following Chudik, Kapetanios, and Pesaran (2018), we set $p=0.05$ and $\delta=1$ in the first stage, while another value, $\delta^{*}=2$, is used in subsequent stages of OCMT described below. Variables with $\widehat{\mathcal{J}}_{i,(1)}=1$ are selected as instruments in the first stage. If no variables are selected in the first stage, then OCMT procedure stops. Otherwise, increase s by one.

2. The next stage ($s>1$) is computed by regressing $\Delta y_{i,t-1}$ on a constant, $\Delta y_{i,t-2}$, all instruments selected from the previous stages, and, one-at-time, the remaining instruments not yet selected. Let $t_{\hat{\theta}_{\ell,(s)}}$ denote the corresponding t-ratio of the instruments considered for selection in the stage $s>1$. Then the instruments are added to the selected set if the indicator $\widehat{\mathcal{J}}_{\ell,(s)}=I[|t_{\hat{\theta}_{\ell,(s)}}|>c_p\left(t-1,\delta^{*}\right)]$ is one. If no instruments are selected in stage s, then the OCMT procedure stops. Otherwise s is increased by one.

3. Step 2 is repeated until no further instruments are selected.

The outcome of this data-dependent selection of moments is $\hat{h}_{nT}$ selected AH moments, $T-2\leq\hat{h}_{nT}\leq\left(T-2\right)\left(T-1\right)/2$.[13]

7 Monte Carlo Evidence

We now provide some evidence on the small sample performance of the AAH estimator as compared to AH, and the two popular AB and BB estimators (also known as first-difference and the system GMM estimators). In addition, we also investigate the small sample performance of the AAH estimator using the subset of AAH moments selected by the OCMT procedure.

7.1 Data generating process (DGP)

The dependent variable is generated as

$$y_{it}=\alpha_i+\phi y_{i,t-1}+u_{it}, \tag{48}$$

[13] This idea can be applied to the any of the GMM estimators considered in this paper. Our focus is on the AAH estimator.

for $i = 1, 2, ..., n$, and $t = 1, 2, ..., T$. We consider $\phi = 0.4, 0.6, 0.8, 0.9$ and report results for $\phi = 0.4$ and 0.8 in the body of the paper.[14] Individual effects are generated as

$$\alpha_i = \sum_{t=1}^{T} \rho^t u_{it} + \pi_i,\ \pi_i \sim IIDN\,(1, 1)\,. \tag{49}$$

We consider two values for $\rho = 0$ or 0.8. When $\rho \neq 0$ then the individual effects are correlated with errors u_{it}, and AB and BB restrictions implicit in (8)-(9), respectively, are not satisfied. The processes are initialized as

$$y_{i,0} = \mu_i + \kappa\pi_i + \upsilon_i,\ \upsilon_i \sim IIDN\,(0, 1)\,, \tag{50}$$

where $\mu_i = \alpha_i/\,(1-\phi)$. We consider two values for $\kappa = 0$ or 1. When $\kappa \neq 0$ the individual effects are correlated with the deviations of initial values from their long-run means μ_i, and BB restrictions implicit in (9) are not satisfied. But setting $\kappa \neq 0$ on its own does not invalidate the AB restrictions implicit in (8).

Restriction $\kappa = 0$ rules out any systematic deviations of initial values from their long-run means. It is less likely to hold in empirical applications, where individual dynamic processes over i might have been initialized from a recent past and possibly from non-stationary initial value distributions. In contrast, the restriction $\rho = 0$ appears much less restrictive, since it would be satisfied whenever fixed effects are uncorrelated with innovations.

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 $\sigma^2_{ia} \sim IIDU\,(0.25, 0.75)$, $\sigma^2_{ib} \sim IIDU\,(1, 2)$, and $e_{it} \sim IID\chi^2\,(2)$, where $[T/2]$ is the integer part of $T/2$. σ^2_{ia} and σ^2_{ib} are generated independently of e_{it}. This ensures that the errors have zero means, and are conditionally heteroskedastic, in particular, $V\left(u_{it}\right|\sigma_{ia}) = \sigma^2_{ia}$ for $t \leq [T/2]$, and $V\left(u_{it}\right|\sigma_{ib}) = \sigma^2_{ib}$ for $t > [T/2]$. We consider comprehensive choices of sample sizes $T = 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20$ and $n = 100, 200, 500, 1000, 2000, 4000, 8000$. Findings for selected sample choices are reported below, whilst the full set of results is available from authors upon request. 2000 replications were carried out for each experiment.

Besides the parameter of interest ϕ, the key parameters of the MC design are κ and ρ. AH and AAH estimators are valid for all values of κ and ρ. AB estimators require $\rho = 0$, and the

[14] Findings for the remaining values of ϕ are available from authors upon request.

BB estimator requires $\rho = 0$ and $\kappa = 0$. Consequently, we consider the following three sets of experiments, based on values of ρ and κ: (i) experiments with $\rho = 0$ and $\kappa = 0$ labeled as experiments where both AB and BB restrictions are met; (ii) experiments with $\rho = 0$ and $\kappa \neq 0$ labeled as experiments where BB restrictions are not met whilst AB restrictions are met, and (iii) experiments with $\rho \neq 0$ and $\kappa \neq 0$ labeled as experiments where neither AB nor BB restrictions are met.

7.2 Estimation methods

We consider 2-step GMM estimators based on the AH moment conditions given by (7), AAH moment conditions given by (7) and (13), Arellano and Bond's first-difference moment conditions given by (8), and the Arellano and Bover's and Blundell and Bond's system moment conditions given by (8)-(9).[15] These estimators are labeled below as AH, AAH, AB, and BB, respectively. Inference is conducted using the conventional standard errors. In addition to two-step GMM estimator based on AAH moments, we also consider using OCMT to select relevant AAH moments, as discussed in Subsection 6.1, denoted as AAH-O. In particular, our AAH-O estimator is based on the union of $T-2$ quadratic moments in (13) and $\hat{h}_{nT}$ selected subset of AH moments using the OCMT procedure described in Subsection 6.1, noting that $T-2 \leq \hat{h}_{nT} \leq (T-2)(T-1)/2$, the number of moments for the AAH-O estimator lies between $2(T-2)$ and $(T-2)(T-1)/2 + T - 2$.

7.3 Monte Carlo findings

7.3.1 Comparison of AH and AAH estimators

We first focus on the comparison of AH and AAH estimators in experiments where both AB and BB restrictions are met ($\rho = 0$ & $\kappa = 0$).[16] Results for bias and RMSE (both $\times 100$) of estimating ϕ are reported in Table 2, and size and power of the tests at the 5% nominal level are reported in Table 3 and Figure 1. Table 2 shows very large RMSE values for the AH estimator, especially when $T = 3$. Once the set of AH moment conditions (7) is augmented by the quadratic moment conditions in (13), we see a substantial drop in the reported RMSE values. The small sample improvements in RMSE are more than 10 fold for $T = 3$, and about four to five-fold for $T = 4, 5$,

[15] We found that cumulative updating (CU) estimators exhibit often worse performance than the 2-step estimators in our experiments. A comaprison of two-step and CU estimators is available in an earlier version of this paper, Chudik and Pesaran (2017).

[16] Findings for the relative performance of AH and AAH estimators are similar for other experiments, available from authors upon request.

and smaller but still substantial for larger values of T, all regardless of n. The relative RMSE differences are somewhat more pronounced when $\phi = 0.8$, as compared to the ones obtained for $\phi = 0.4$. Compared to the AH estimator, the AAH estimator is less biased in almost all reported cases, and has a smaller RMSE even for $T = 14$. This suggests that correcting for bias will be unimportant for the reported sample choices.

In line with the bias and RMSE findings, we see in Table 3 that there are substantial gains in power from the augmentation of the AH moments with the new quadratic moment conditions in (13). These differences can be seen more clearly in Figure 1, shown for the sample combinations, $n = 1000$, $T = 4$ and 6. The empirical power functions of the AAH estimator are rather flat when $\phi = 0.8$, and $T = 4$. As to be expected, the results for the AH estimator improve with a decrease in ϕ (as AH instruments become stronger), and/or a rise in T. In contrast, the empirical power function of the AAH estimator is much more satisfactory. The size of the AH and AAH estimators reported in Table 3 are close to their nominal value of 0.05, in cases where T/n is sufficiently small. For $T = 3$ the reported size is close to 5% for all values of n considered, whereas for $T = 4$, size is close to 5 percent only for $n \geq 500$. Size clearly deteriorates in the case where the number of moments is not sufficiently small relative to the number of cross-section units, n, which is a well-known problem in the GMM literature.

7.3.2 Comparison of AAH and AAH-O estimators

With an increase in T, the number of moments becomes large, many of which could be relatively weak. In such a case, using a well chosen sub-set of moments could improve the small sample performance. We investigate the small sample benefits and drawbacks of using OCMT procedure described in Section 6.1 to select a subset or relevant AH moments. The AAH-O estimator is based on the union of $T - 2$ quadratic moments (13) and the selected subset of AH moments. We expect that for a fixed T and as $n \to \infty$, all relevant moments will be selected by OCMT procedure and therefore asymptotically AAH, and AAH-O achieve the same variance (for a fixed T), although AAH-O could have lower or higher RMSE compared with AAH in finite samples. These expectations are in line with the reported findings in Tables 4-5 and Figure 2. First, the average number of moments (reported in the last columns of Table 4) increases in n for a fixed T, since all of the AAH moments are relevant albeit with a varying degree of strength. The differences in RMSE values between AAH and AAH-O estimators are negligible for the largest value of n, as expected. Second, AAH-O outperforms AAH in cases with highest value of the ratio T/n, for example when $n = 100$, and $T > 12$. However, for intermediate cases with more moderate T/n ratio, AAH-O tends to perform less well as compared to the AAH estimator in terms of RMSEs. The size distortions of AAH-O are not as serious as the size distortions of AAH, but still quite substantial in the case of experiments where $T \geq 10$ and $n < 2000$.

7.3.3 Comparison of AAH with AB and BB estimators

We now turn to the small sample performance of the AAH estimator compared to the AB and BB estimators. Comparisons for experiments where both AB and BB restrictions are met are reported in Tables 6-7 and Figure 3. In these experiments AAH is asymptotically less efficient than BB, and this is reflected in the lower values of RMSEs obtained for the BB estimator; although it is interesting to note that these differences are not large in many cases. This result is also in line with the asymptotic relative efficiency of the BB estimator reported in Table 1. The situation is very different when the AAH estimator is compared to the AB estimator. As can be seen from Table 6, in all cases the AAH estimator performs better (in many cases substantially) than the AB estimator. Size of the tests based on the individual estimators is close to 5% when n is sufficiently large relative to T, otherwise when T is large relative to n inference could be unsafe with substantial over-rejections.[17]

To investigate the factors behind the better performance of the BB estimator, we now consider experiments where individual effects are correlated with the deviations of initial values $y_{i0} - \mu_i$, by setting $\kappa = 1$. In these experiments, reported in Tables 8-9 and Figure 4, the restrictions underlying the BB estimator are not met. Hence, BB is estimator is no longer consistent, which shows the BB estimator having large biases and close to 100 percent size over-rejections. The remaining two estimators (AAH and AB) are consistent and their relative performance is very similar to the previous experiments reported in Tables 6-7, with the proposed AAH estimator generally dominating the AB estimator.

In the last set of experiments, reported in Tables 10-11 and Figure 5, we also allow for correlation of errors and fixed effects (by setting the parameter $\rho = 0.8$), in addition to $\kappa = 1$. In these experiments AAH continues to be valid, but the moment conditions of AB and BB are both violated.

[17] The size performance can be improved upon by considering alternative estimates of standard errors, such as Windmeijer (2005) finite sample corrections for the standard errors of two-step GMM estimators, or Newey and Windmeijer (2009) standard errors for the CU-GMM estimators. These or other alternative estimators of standard errors are not pursued in this paper.

As a result both of these estimators perform very poorly, and exhibit large biases and substantial size distortions even when $T = 3$. In contrast, the MC findings for the AAH estimator perform well, and in fact are numerically identical to those reported in Tables 8-9, since due to first-differencing the AAH estimator is not affected by changes in ρ and κ.[18]

Overall, the MC findings show that the AAH estimator is robust and outperform its 'cousin', the AH estimator by a wide margin. The AB and BB estimators are not robust to $\rho \neq 0$, and BB is also not robust to $\kappa \neq 0$. In the case of experiments with $\rho = 0$ & $\kappa = 0$, the AAH estimator continues to outperform the AB estimator, but performs less well when compared to the BB estimator, which is obtained under a much stronger set of restrictions (given by (11)). In practice it is not known whether these additional restrictions on the initialization of dynamic processes are satisfied, and violation of these conditions renders the BB estimator inconsistent with large biases and substantial over-rejections.

7.3.4 Hausman test for a comparison of AAH and BB estimators

We now consider the small sample performance of the Hausman test proposed in Section 4. This test compares AAH and BB estimators. As already noted, under the null hypothesis of BB conditions holding, we have $Var\left(AAH\right) \leq Var\left(BB\right)$, whereas BB estimator will be inconsistent if BB conditions are not met. Table 12 shows the rejection rates of Hausman test (defined by (40)) at 5 per cent nominal level under the null that BB conditions are met (namely $H_0 : \rho = \kappa = 0$), as well as the rejection rates under the alternative hypothesis $H_1 : \rho = 0$ and $\kappa = 1$, under which the BB conditions do not hold. For $T = 3$ and all choices of n considered, our findings suggest that the Hausman test has relatively good size. However, rejection rates increase well beyond the 5 per cent nominal level as T increases and n is not sufficiently large. These distortions can be observed in sample sizes where $Var\left(AAH\right)$ and $Var\left(BB\right)$ are not well estimated due to large number of moments and n not being sufficiently large. Under the null hypothesis we also observe a large incidence of cases (reported in the right part of Table 2) where $\widehat{Var}\left(AAH\right) < \widehat{Var}\left(BB\right)$ and Hausman test is therefore not applicable. Large incidence of these cases relate to very small differences in RMSE values reported earlier in Table 6, in particular for larger values of T.

Rejection rates under the alternative hypothesis ($H_1 : \rho = 0$ and $\kappa = 1$) are quite large and quickly approach one as n increases, suggesting relatively good power of the Hausman test for this

[18] To make the results in Tables 10 & 11 and Tables 8 & 9 comparable we have used the same seed for generating the random numbers.

design. Overall, Hausman test seems to work well when T is sufficiently small but as T is increased the size distortion quickly appear in a same pattern as the reported size distortions observed in Table 7.

8 Concluding remarks

Instead of searching for instruments that are uncorrelated with the errors, this paper proposes to use the regressors (target variables) themselves in cases where the correlation between the target variables and the errors can be derived. This approach will lead to possibly nonlinear bias-corrected moment conditions. In this paper this idea is applied to the estimation of short-T dynamic panel data models, and a new augmented Anderson-Hsiao (AAH) estimator is proposed without making additional restrictions. The basic idea has potential applications 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, which we leave for future research.

The proposed AAH estimator is 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 AAH estimators can be less efficient asymptotically when the stricter requirements of the system GMM estimator proposed by Blundell and Bond hold. The robustness of the AAH estimators is likely to be an advantage in practice where it is not possible to know if the stronger requirements of the system-GMM estimators are met, and thus avoid possible estimation bias and incorrect inference.

To decide between AAH and BB estimators in empirical applications we also propose a Hausman type test which is shown to work well when T is small and n sufficiently large.

This paper only considered panels with a fixed T. In panels with $n, T \to \infty$ jointly, there is an important issue that pertains to the GMM approach, namely the problem of combining a large number of moment conditions. We have briefly discussed this topic and proposed using OCMT to select a subset of relevant moment conditions as a simple way to mitigate the adverse effects of moments proliferation.

Table 2: **Bias and RMSE of AH and AAH estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met**

		Bias (×100)				**RMSE**(×100)			
		$\phi_0 = 0.4$		$\phi_0 = 0.8$		$\phi_0 = 0.4$		$\phi_0 = 0.8$	
T	n	AH	AAH	AH	AAH	AH	AAH	AH	AAH
3	100	< -100	4.63	32.37	4.98	>1000	21.32	>1000	29.09
3	200	< -100	3.22	< -100	4.88	>1000	13.62	>1000	23.48
3	500	8.85	1.49	< -100	2.18	227.47	6.85	>1000	10.14
3	1000	8.91	0.65	>100	1.76	88.28	4.25	>1000	7.40
3	2000	2.34	0.04	14.91	0.52	25.58	2.63	108.21	4.08
3	8000	0.99	-0.06	3.15	-0.06	11.97	1.23	23.40	1.68
4	100	-11.18	1.22	-39.91	0.22	51.22	12.05	94.44	14.56
4	200	-3.70	0.36	-24.55	0.21	34.55	7.44	75.82	10.65
4	500	-2.43	-0.01	-10.60	-0.01	19.97	4.05	41.66	6.40
4	1000	-0.83	0.13	-3.78	0.13	13.78	2.88	26.43	4.33
4	2000	-0.71	-0.10	-2.21	-0.10	9.78	2.03	17.50	3.04
4	8000	-0.08	-0.01	-0.15	-0.01	4.65	1.06	8.40	1.53
5	100	-4.68	1.18	-26.99	-0.17	21.26	10.74	49.53	10.77
5	200	-2.43	0.59	-15.29	-0.28	14.65	8.00	34.44	6.81
5	500	-0.95	-0.02	-5.73	-0.23	9.42	3.03	20.15	4.00
5	1000	-0.13	-0.02	-2.53	-0.20	6.58	2.09	13.57	2.72
5	2000	-0.33	-0.09	-1.69	-0.16	4.60	1.47	9.49	1.88
5	8000	-0.04	-0.02	-0.36	-0.04	2.26	0.75	4.51	0.95
6	100	-5.01	1.08	-23.59	0.14	17.29	9.86	38.13	9.27
6	200	-2.42	0.24	-12.94	-0.06	11.74	6.08	25.62	6.51
6	500	-1.05	-0.14	-5.20	-0.39	7.13	2.64	14.30	3.70
6	1000	-0.37	-0.01	-2.42	-0.14	5.09	1.83	9.91	2.49
6	2000	-0.19	-0.07	-1.28	-0.14	3.51	1.32	6.71	1.75
6	8000	-0.03	-0.02	-0.20	-0.05	1.75	0.63	3.23	0.86
10	100	-3.40	0.48	-14.08	0.22	8.84	5.66	19.59	6.38
10	200	-1.49	0.17	-7.05	-0.01	5.59	3.03	11.61	3.92
10	500	-0.60	-0.03	-2.54	-0.04	3.37	1.82	6.05	2.32
10	1000	-0.22	-0.03	-1.20	-0.12	2.36	1.25	3.97	1.46
10	2000	-0.16	-0.03	-0.60	-0.07	1.65	0.84	2.74	0.98
10	8000	-0.04	-0.01	-0.14	-0.03	0.79	0.41	1.30	0.49

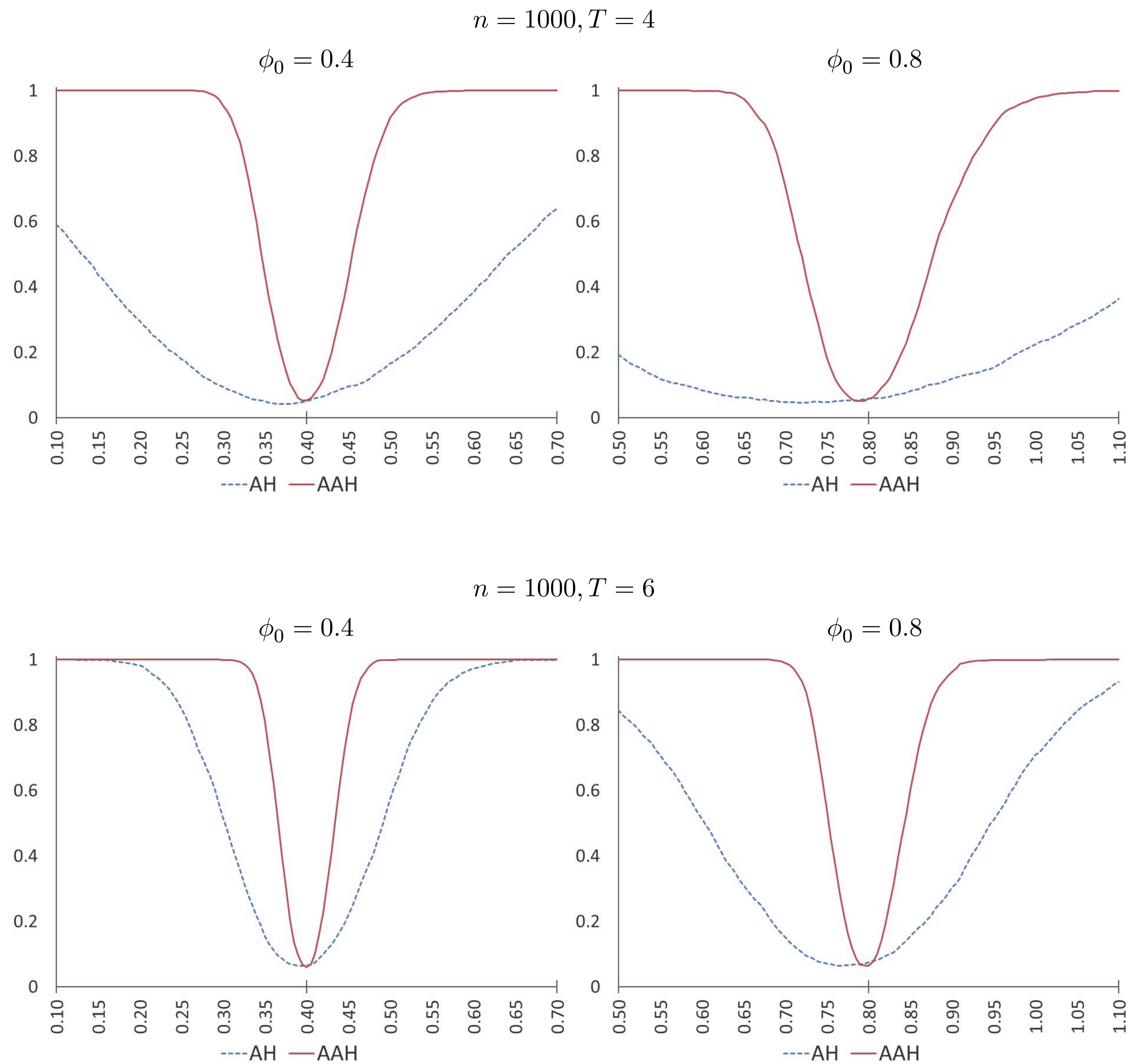
Notes: "AH" is the 2-step GMM estimator based on the $(T-2)(T-1)/2$ Anderson and Hsiao's moment conditions (7), "AAH" is the augmented Anderson and Hsiao 2-step GMM estimator based on the $(T-2)(T-1)/2+T-2$ moment conditions (7) and (13). The DGP is given by $y_{it} = \alpha_i + \phi y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., n$, and $t = 1, 2, ..., T$, with $y_{i,0} = \mu_i + \kappa\pi_i + \upsilon_i$, where $\mu_i = \alpha_i/(1-\phi)$, $\alpha_i = \sum_{t=1}^{T} \rho^t u_{it} + \pi_i$, $\pi_i \sim IIDN(1,1)$, and $\upsilon_i \sim IIDN(0,1)$. This table reports findings for experiments where $\kappa = \rho = 0$, namely AB and BB restrictions are met. BB restrictions are not satisfied when $\kappa \neq 0$, and AB restrictions are not satisfied when $\rho \neq 0$. Errors u_{it} 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 $T/2$. See Section 7 for a full description of the MC experiments.

Table 3: Size and Power of AH and AAH estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met

		Size (5% level, $\times 100$)				**Power (5% level, $\times 100$, $H_1 : \phi = \phi_0 + 0.1$)**			
		$\phi_0 = 0.4$		$\phi_0 = 0.8$		$\phi_0 = 0.4$		$\phi_0 = 0.8$	
T	n	AH	AAH	AH	AAH	AH	AAH	AH	AAH
3	100	5.8	6.2	7.7	7.0	8.3	21.2	10.0	18.7
3	200	6.2	4.6	7.8	5.8	9.1	26.0	9.8	21.2
3	500	5.9	4.5	7.3	3.9	9.8	47.3	10.2	30.5
3	1000	5.8	4.6	6.3	4.5	10.9	73.4	9.8	45.4
3	2000	4.5	5.3	5.8	4.6	11.2	95.8	9.9	76.6
3	8000	5.0	5.0	4.8	4.6	16.5	100.0	11.5	99.9
4	100	11.0	10.2	19.2	12.7	16.0	35.0	23.2	30.6
4	200	7.5	7.4	13.0	9.5	12.3	45.6	17.3	34.0
4	500	6.3	6.1	7.9	6.6	13.9	71.7	12.8	47.6
4	1000	5.2	5.3	5.8	5.9	16.6	91.7	11.9	65.9
4	2000	6.3	5.3	5.3	5.8	23.5	99.5	13.2	88.1
4	8000	5.0	5.6	4.9	5.8	56.8	100.0	23.3	100.0
5	100	13.0	15.5	25.7	17.3	24.5	51.1	33.6	46.0
5	200	8.3	10.9	16.3	10.8	22.5	65.6	24.5	54.0
5	500	6.8	7.1	8.4	7.8	29.2	90.9	19.7	77.7
5	1000	5.9	6.5	5.9	6.4	38.8	99.5	20.2	95.1
5	2000	5.2	5.4	5.1	5.9	61.3	100.0	27.7	99.9
5	8000	4.0	5.8	4.0	5.3	99.3	100.0	60.5	100.0
6	100	18.3	20.5	31.1	20.3	33.9	58.0	43.4	53.9
6	200	11.2	12.0	19.1	13.7	31.4	76.1	31.8	61.8
6	500	7.4	8.0	9.5	9.5	40.0	96.9	26.8	84.7
6	1000	6.5	6.1	7.5	6.5	57.4	100.0	30.5	96.4
6	2000	4.8	5.5	5.5	6.3	82.1	100.0	42.3	100.0
6	8000	4.6	4.1	4.5	5.2	100.0	100.0	87.2	100.0
10	100	40.5	47.3	58.9	49.3	76.6	88.4	83.3	85.4
10	200	20.5	24.1	32.6	27.9	77.9	97.2	74.6	93.4
10	500	10.4	12.1	14.4	15.3	93.7	100.0	75.8	99.8
10	1000	8.3	9.6	8.8	11.3	99.4	100.0	88.8	100.0
10	2000	7.1	6.2	7.4	7.3	100.0	100.0	98.5	100.0
10	8000	5.3	5.2	5.0	5.9	100.0	100.0	100.0	100.0

See the notes to Table 2

Figure 1: Rejection frequencies (at 5% nominal level) for AH and AAH estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met



See the notes to Table 2.

Table 4: Bias and RMSE of AAH and AAH-O estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met

		Bias ($\times 100$)				**RMSE**($\times 100$)				**Average number of moments**		
		$\phi_0 = 0.4$		$\phi_0 = 0.8$		$\phi_0 = 0.4$		$\phi_0 = 0.8$			AAH-O	
T	n	AAH	AAH-O	AAH	AAH-O	AAH	AAH-O	AAH	AAH-O	AAH	$\phi_0 = 0.4$	$\phi_0 = 0.8$
10	100	0.48	0.12	0.22	-0.39	5.66	5.53	6.38	7.21	44	20	17
10	200	0.17	-0.05	-0.01	-0.56	3.03	3.12	3.92	4.29	44	23	18
10	500	-0.03	-0.07	-0.04	-0.25	1.82	1.90	2.32	2.57	44	28	22
10	1000	-0.03	-0.05	-0.12	-0.26	1.25	1.28	1.46	1.55	44	33	25
10	2000	-0.03	-0.05	-0.07	-0.14	0.84	0.85	0.98	1.02	44	39	29
10	8000	-0.01	-0.01	-0.03	-0.03	0.41	0.41	0.49	0.49	44	43	44
12	100	0.70	0.04	0.75	-0.20	5.34	4.79	6.39	6.76	65	25	22
12	200	0.19	-0.05	0.09	-0.52	2.95	2.98	3.71	4.38	65	29	23
12	500	0.00	-0.03	-0.06	-0.24	1.63	1.72	1.97	2.39	65	37	28
12	1000	-0.01	-0.03	-0.09	-0.21	1.10	1.15	1.20	1.61	65	45	34
12	2000	-0.02	-0.04	-0.04	-0.12	0.77	0.78	0.83	0.88	65	54	40
12	8000	0.00	-0.01	-0.01	-0.01	0.37	0.37	0.40	0.40	65	63	63
14	100	0.98	0.03	1.11	-0.23	6.68	4.26	7.59	6.25	90	30	26
14	200	0.22	-0.03	0.21	-0.51	2.81	2.73	3.55	4.19	90	35	28
14	500	-0.01	-0.07	-0.03	-0.26	1.52	1.60	1.68	2.24	90	47	34
14	1000	0.02	-0.01	-0.03	-0.18	1.01	1.06	1.06	1.22	90	57	43
14	2000	-0.01	-0.03	-0.03	-0.13	0.68	0.70	0.71	0.78	90	69	52
14	8000	0.00	-0.01	-0.01	-0.02	0.34	0.34	0.35	0.35	90	86	84
16	100	0.73	0.09	-0.59	-0.11	7.93	4.16	9.24	6.02	119	35	30
16	200	0.22	-0.12	0.42	-0.57	2.92	2.63	3.69	4.00	119	41	32
16	500	0.06	-0.02	0.04	-0.32	1.44	1.45	1.58	1.83	119	56	39
16	1000	0.03	0.00	-0.01	-0.20	0.93	0.97	0.96	1.15	119	71	51
16	2000	-0.03	-0.04	-0.05	-0.16	0.62	0.64	0.63	0.70	119	86	65
16	8000	0.00	0.00	0.00	-0.01	0.31	0.31	0.31	0.31	119	112	103
18	100	-0.33	-0.01	-5.41	-0.22	7.57	3.83	10.81	5.86	152	40	35
18	200	0.35	-0.08	0.77	-0.54	3.20	2.55	4.08	3.97	152	47	37
18	500	0.09	-0.03	0.11	-0.31	1.40	1.39	1.57	1.93	152	64	44
18	1000	0.03	0.00	0.01	-0.21	0.87	0.89	0.87	1.08	152	83	58
18	2000	-0.01	-0.02	-0.03	-0.14	0.59	0.61	0.58	0.64	152	103	79
18	8000	0.00	0.00	0.00	-0.01	0.29	0.29	0.28	0.29	152	141	121
20	100	-1.73	-0.06	-9.91	-0.34	8.22	3.73	14.22	5.72	189	45	39
20	200	0.55	-0.09	1.07	-0.67	4.63	2.39	5.32	3.51	189	53	41
20	500	0.07	-0.04	0.08	-0.36	1.38	1.33	1.46	1.86	189	72	49
20	1000	0.02	-0.01	0.00	-0.22	0.84	0.86	0.82	1.04	189	96	64
20	2000	0.00	-0.02	0.00	-0.12	0.55	0.57	0.53	0.60	189	121	92
20	8000	0.00	0.00	0.00	-0.02	0.27	0.27	0.25	0.26	189	171	139

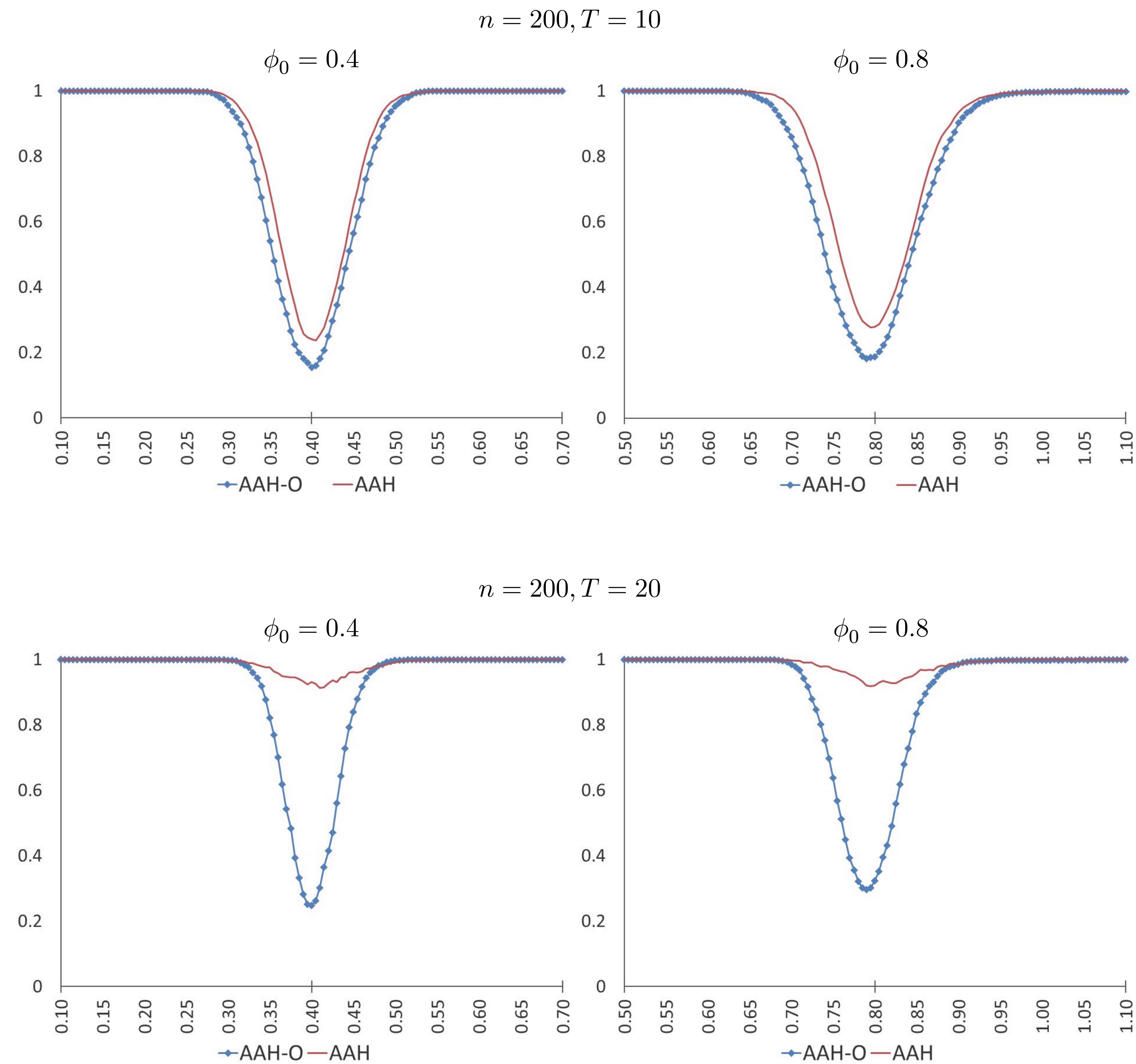
Notes: See the notes to Table 2. "AAH-O" estimator is the two-step GMM estimator based on $T-2$ quadratic moment conditions (13) and a subset of $(T-2)(T-1)/2$ AAH moment conditions (7) selected by OCMT. See section 7 for a full description of the MC experiments.

Table 5: Size and power of AAH and AAH-O estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met

		Size (5% level, ×100)				Power (5% level, ×100, $H_1: \phi = \phi_0 + 0.02$)			
		$\phi_0 = 0.4$		$\phi_0 = 0.8$		$\phi_0 = 0.4$		$\phi_0 = 0.8$	
T	n	AAH	AAH-O	AAH	AAH-O	AAH	AAH-O	AAH	AAH-O
10	100	47.3	25.6	49.3	29.6	49.5	32.6	52.3	36.3
10	200	24.1	15.4	27.9	18.7	31.6	25.0	36.4	28.5
10	500	12.1	10.7	15.3	14.6	34.8	30.5	34.2	33.0
10	1000	9.6	9.5	11.3	12.4	46.2	44.1	42.4	44.7
10	2000	6.2	7.0	7.3	8.8	71.0	71.2	61.4	64.9
10	8000	5.2	5.3	5.9	6.1	99.8	99.9	98.0	98.1
12	100	62.8	28.0	66.1	33.4	66.0	32.6	67.8	40.2
12	200	32.3	19.5	37.3	23.8	43.0	29.0	47.1	35.8
12	500	15.7	12.1	18.1	16.5	41.6	35.4	44.2	38.9
12	1000	9.9	8.9	11.4	13.1	56.7	51.5	54.6	56.6
12	2000	8.7	8.2	8.2	9.2	80.9	80.0	74.7	76.6
12	8000	6.1	6.0	5.5	5.7	100.0	100.0	99.9	99.9
14	100	89.2	32.2	88.0	39.0	87.2	38.3	88.4	45.4
14	200	42.8	19.5	46.3	27.4	53.2	33.4	58.4	38.7
14	500	18.6	13.4	23.6	18.8	51.2	42.1	52.0	44.8
14	1000	11.4	10.7	14.6	15.2	64.7	59.2	65.4	62.3
14	2000	7.9	7.2	8.9	9.9	87.5	85.6	86.5	86.8
14	8000	6.2	6.2	5.8	6.6	100.0	100.0	100.0	100.0
16	100	85.9	34.6	85.4	41.9	86.7	41.1	86.3	48.1
16	200	59.4	22.3	61.9	28.5	63.4	36.3	66.6	43.3
16	500	22.0	12.9	27.7	18.8	56.6	44.6	59.3	47.0
16	1000	13.8	10.8	16.1	16.0	71.8	66.3	74.7	69.5
16	2000	8.6	7.9	10.0	11.1	93.1	91.8	93.4	92.8
16	8000	5.7	5.8	6.2	7.2	100.0	100.0	100.0	100.0
18	100	74.5	38.0	78.1	44.3	75.0	42.9	80.2	53.1
18	200	72.0	24.4	74.4	30.8	76.7	39.0	77.9	45.9
18	500	28.1	14.0	31.7	20.6	62.8	49.7	67.3	51.0
18	1000	14.8	9.7	17.6	16.6	78.1	70.9	81.2	72.1
18	2000	9.7	8.7	10.9	11.5	95.7	94.1	96.8	95.6
18	8000	6.4	6.4	6.3	7.7	100.0	100.0	100.0	100.0
20	100	66.8	38.9	78.2	46.1	70.1	48.2	82.8	53.8
20	200	93.1	24.8	92.0	32.3	92.8	41.6	92.7	49.1
20	500	33.4	15.1	39.1	20.3	69.6	54.2	72.0	54.7
20	1000	18.7	11.2	20.8	15.9	82.5	75.7	87.0	75.7
20	2000	10.4	9.0	11.9	11.7	96.5	95.4	98.4	96.7
20	8000	7.3	6.8	6.4	7.5	100.0	100.0	100.0	100.0

See the notes to Tables 2 and 4.

Figure 2: Rejection frequencies (at 5% nominal level) for AAH and AAH-O estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met



See the notes to Tables 2 and 4.

Table 6: Bias and RMSE of AAH, AB and BB estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met

		Bias ($\times 100$)						**RMSE**($\times 100$)					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	4.63	-8.77	2.41	4.98	-59.00	2.32	21.32	33.60	11.69	29.09	112.54	13.30
3	200	3.22	-4.27	1.48	4.88	-43.32	1.38	13.62	21.86	8.28	23.48	96.01	8.53
3	500	1.49	-1.48	0.67	2.18	-24.42	0.55	6.85	12.81	4.97	10.14	68.48	4.94
3	1000	0.65	-0.48	0.49	1.76	-9.78	0.45	4.25	8.74	3.46	7.40	46.31	3.61
3	2000	0.04	-0.18	0.15	0.52	-5.11	0.06	2.63	6.28	2.47	4.08	30.92	2.61
3	8000	-0.06	-0.18	-0.02	-0.06	-1.14	-0.01	1.23	3.07	1.19	1.68	15.11	1.30
4	100	1.22	-4.92	2.01	0.22	-45.29	2.44	12.05	20.08	9.84	14.56	73.89	11.16
4	200	0.36	-2.18	0.99	0.21	-29.61	1.10	7.44	13.84	6.57	10.65	55.67	7.64
4	500	-0.01	-0.85	0.40	-0.01	-14.09	0.25	4.05	8.65	4.07	6.40	32.95	4.55
4	1000	0.13	-0.41	0.32	0.13	-6.41	0.21	2.88	6.13	2.85	4.33	21.54	3.26
4	2000	-0.10	-0.26	0.01	-0.10	-3.46	-0.07	2.03	4.36	2.01	3.04	14.81	2.40
4	8000	-0.01	-0.03	0.02	-0.01	-0.55	-0.02	1.06	2.13	1.03	1.53	7.25	1.16
5	100	1.18	-4.25	1.06	-0.17	-29.96	1.93	10.74	15.00	7.60	10.77	46.23	8.33
5	200	0.59	-2.20	0.49	-0.28	-18.54	0.83	8.00	10.20	4.97	6.81	32.12	5.40
5	500	-0.02	-0.77	0.17	-0.23	-8.00	0.20	3.03	6.31	3.03	4.00	18.72	3.37
5	1000	-0.02	-0.18	0.08	-0.20	-3.74	0.04	2.09	4.25	2.09	2.72	12.37	2.35
5	2000	-0.09	-0.38	-0.05	-0.16	-2.45	-0.02	1.47	3.09	1.45	1.88	8.66	1.63
5	8000	-0.02	-0.07	-0.01	-0.04	-0.56	0.00	0.75	1.56	0.74	0.95	4.23	0.84
6	100	1.08	-4.36	1.00	0.14	-25.52	2.72	9.86	12.77	6.92	9.27	37.03	7.99
6	200	0.24	-2.00	0.44	-0.06	-15.32	1.31	6.08	8.56	4.44	6.51	25.24	5.49
6	500	-0.14	-0.83	0.04	-0.39	-6.53	0.26	2.64	5.16	2.68	3.70	14.21	3.24
6	1000	-0.01	-0.26	0.08	-0.14	-3.03	0.13	1.83	3.63	1.84	2.49	9.43	2.23
6	2000	-0.07	-0.16	-0.03	-0.14	-1.61	-0.03	1.32	2.47	1.31	1.75	6.25	1.54
6	8000	-0.02	-0.06	-0.01	-0.05	-0.36	-0.03	0.63	1.24	0.63	0.86	3.00	0.77
10	100	0.48	-3.14	0.65	0.22	-14.73	2.53	5.66	8.24	5.23	6.38	19.91	6.00
10	200	0.17	-1.44	0.32	-0.01	-7.79	1.40	3.03	5.02	3.04	3.92	11.92	3.84
10	500	-0.03	-0.62	0.03	-0.04	-2.97	0.33	1.82	3.04	1.83	2.32	6.11	2.16
10	1000	-0.03	-0.22	0.00	-0.12	-1.38	0.07	1.25	2.07	1.26	1.46	3.92	1.44
10	2000	-0.03	-0.14	-0.02	-0.07	-0.70	0.01	0.84	1.42	0.84	0.98	2.66	0.95
10	8000	-0.01	-0.04	-0.01	-0.03	-0.17	-0.01	0.41	0.70	0.41	0.49	1.26	0.46

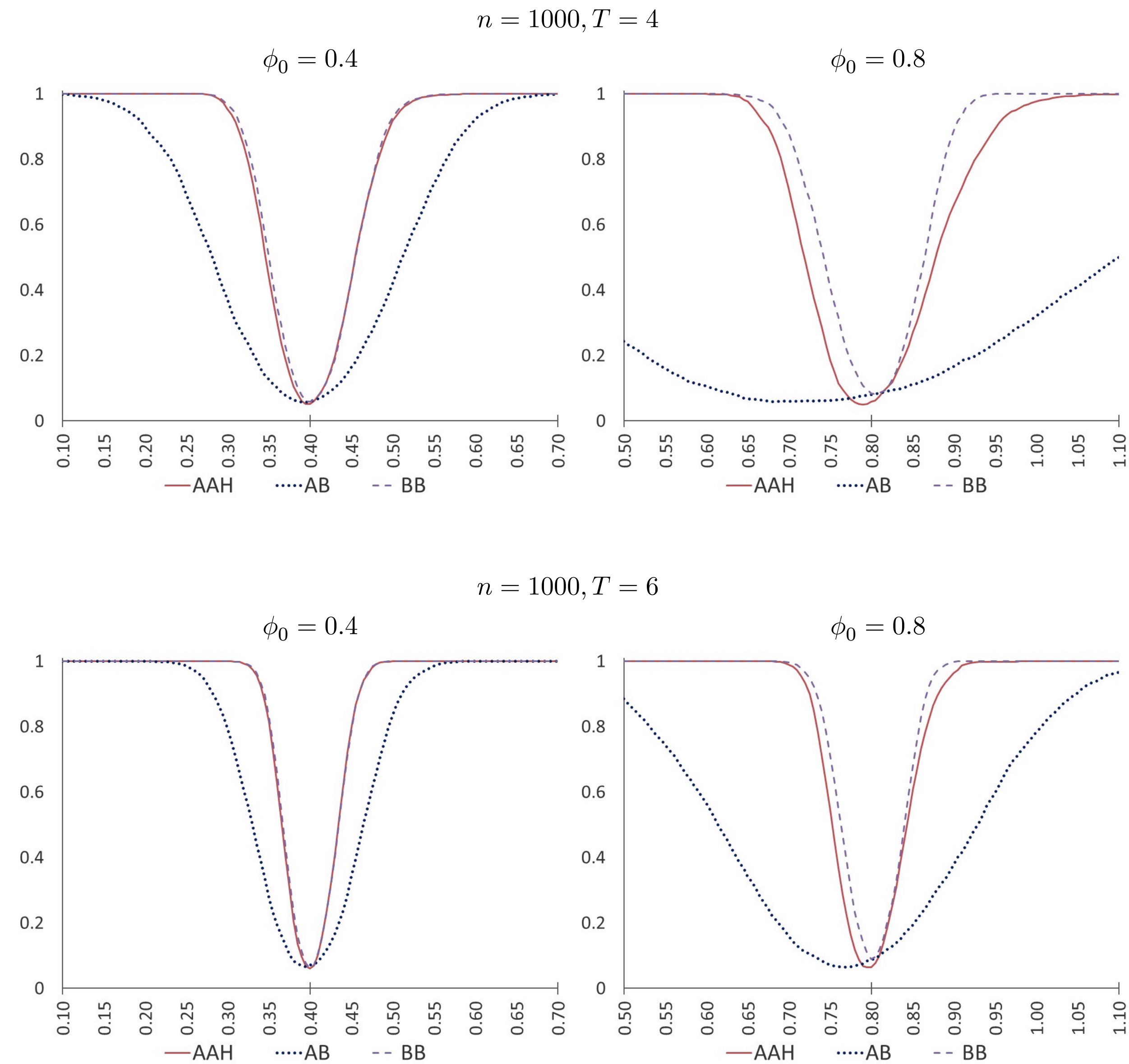
Notes: See the notes to Table 2. "AB" is the 2-step GMM estimator based on the Arellano and Bond's first-difference moment conditions (8), and "BB" is the 2-step GMM estimator based on the Arellano and Bover's and Blundell and Bond's system moment conditions (8)-(9).

Table 7: Size and power of AAH, AB and BB estimators when both Arellano and Bond (AB) and Blundell and Bond (BB) restrictions are met

		Size (5% level, ×100)						Power (5% level, ×100, $H_1 : \phi = \phi_0 + 0.1$)					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	6.2	9.2	11.3	7.0	21.9	18.3	21.2	14.6	22.7	18.7	25.6	20.6
3	200	4.6	7.5	8.9	5.8	19.2	12.8	26.0	14.7	32.2	21.2	23.0	26.7
3	500	4.5	5.9	6.5	3.9	13.1	8.0	47.3	16.3	52.9	30.5	17.2	46.1
3	1000	4.6	4.8	5.3	4.5	8.9	5.9	73.4	23.5	77.2	45.4	13.8	72.3
3	2000	5.3	5.9	5.4	4.6	7.2	6.8	95.8	39.7	97.1	76.6	12.6	96.7
3	8000	5.0	5.5	4.9	4.6	5.8	4.9	100.0	91.0	100.0	99.9	16.0	100.0
4	100	10.2	12.0	15.9	12.7	30.1	28.7	35.0	21.4	32.5	30.6	36.5	32.4
4	200	7.4	9.0	10.3	9.5	21.0	19.8	45.6	21.5	43.5	34.0	27.3	39.0
4	500	6.1	6.5	7.0	6.6	12.0	10.9	71.7	27.9	71.1	47.6	19.8	65.1
4	1000	5.3	6.2	6.0	5.9	8.1	8.5	91.7	41.9	92.7	65.9	16.7	88.7
4	2000	5.3	5.9	5.1	5.8	6.5	7.6	99.5	67.2	99.6	88.1	18.8	99.5
4	8000	5.6	4.7	5.7	5.8	5.8	5.2	100.0	99.8	100.0	100.0	32.6	100.0
5	100	15.5	15.7	21.9	17.3	33.7	35.3	51.1	33.1	51.3	46.0	45.6	49.4
5	200	10.9	10.6	13.3	10.8	21.9	20.6	65.6	34.4	65.6	54.0	34.6	63.6
5	500	7.1	6.8	8.2	7.8	10.8	11.4	90.9	46.3	91.4	77.7	25.9	90.1
5	1000	6.5	4.5	6.7	6.4	6.9	8.7	99.5	66.4	99.6	95.1	24.7	99.4
5	2000	5.4	5.5	4.9	5.9	6.7	6.0	100.0	91.8	100.0	99.9	33.4	100.0
5	8000	5.8	4.9	5.6	5.3	4.8	5.6	100.0	100.0	100.0	100.0	70.9	100.0
6	100	20.5	21.1	27.8	20.3	41.2	46.1	58.0	45.9	61.8	53.9	56.2	56.3
6	200	12.0	14.0	16.5	13.7	25.9	30.1	76.1	44.8	76.5	61.8	43.8	68.2
6	500	8.0	7.9	9.2	9.5	13.4	14.8	96.9	62.6	96.9	84.7	35.3	94.6
6	1000	6.1	7.1	6.2	6.5	8.8	9.0	100.0	83.4	100.0	96.4	37.9	99.8
6	2000	5.5	5.2	6.0	6.3	5.8	7.5	100.0	97.7	100.0	100.0	50.1	100.0
6	8000	4.1	4.6	4.6	5.2	4.6	5.4	100.0	100.0	100.0	100.0	92.6	100.0
10	100	47.3	48.0	56.9	49.3	67.8	72.8	88.4	82.7	90.2	85.4	88.6	86.0
10	200	24.1	23.0	26.2	27.9	39.9	45.1	97.2	87.3	98.1	93.4	81.1	95.1
10	500	12.1	11.9	14.0	15.3	16.8	22.9	100.0	97.1	100.0	99.8	82.4	100.0
10	1000	9.6	8.5	10.3	11.3	10.2	13.7	100.0	100.0	100.0	100.0	91.5	100.0
10	2000	6.2	6.8	6.8	7.3	7.5	9.1	100.0	100.0	100.0	100.0	99.1	100.0
10	8000	5.2	5.3	5.2	5.9	5.3	5.4	100.0	100.0	100.0	100.0	100.0	100.0

See the notes to Tables 2 and 6.

Figure 3: Rejection frequencies (at 5% nominal level) for AAH, AB, and BB estimators when AB and BB restrictions are met



See notes to Tables 2 and 6.

Table 8: Bias and RMSE of AAH, AB and BB estimators when Arellano and Bond (AB) restrictions are met and Blundell and Bond (BB) restrictions are not met

		Bias (×100)						RMSE(×100)					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	1.17	-1.35	26.32	2.70	-12.59	18.48	14.13	11.98	28.35	23.20	43.64	23.41
3	200	0.59	-0.83	26.88	2.85	-5.68	19.51	8.41	8.37	27.97	18.57	26.62	21.68
3	500	0.56	-0.21	27.48	1.56	-1.75	20.41	4.99	5.12	27.89	8.97	14.66	21.05
3	1000	0.47	-0.03	27.69	1.58	-0.70	20.83	3.45	3.64	27.90	6.14	10.27	21.10
3	2000	0.23	0.01	27.68	0.82	-0.31	20.86	2.42	2.58	27.78	4.01	7.17	20.98
3	8000	0.13	-0.06	27.71	0.19	-0.21	20.89	1.19	1.28	27.74	1.89	3.59	20.92
4	100	1.20	-0.80	23.34	0.54	-7.56	18.94	11.70	8.62	24.96	14.23	23.42	20.44
4	200	0.31	-0.33	23.96	0.32	-3.21	20.34	6.95	6.15	24.80	10.33	15.30	20.93
4	500	-0.02	-0.12	24.75	-0.02	-1.24	21.41	3.30	3.84	25.10	5.89	9.46	21.60
4	1000	0.06	-0.07	24.88	0.09	-0.75	21.82	2.36	2.76	25.05	4.03	6.81	21.91
4	2000	-0.08	-0.03	24.96	-0.10	-0.31	22.03	1.65	1.97	25.05	2.84	4.78	22.07
4	8000	-0.02	0.00	25.08	-0.02	-0.04	22.11	0.85	0.96	25.10	1.44	2.35	22.12
5	100	1.02	-1.28	15.44	-0.18	-8.38	15.66	9.96	7.72	17.31	10.03	19.41	17.39
5	200	0.39	-0.66	15.85	-0.29	-4.21	17.60	7.15	5.39	16.78	6.45	12.52	18.36
5	500	0.01	-0.14	16.37	-0.23	-1.39	19.06	2.58	3.28	16.77	3.75	7.18	19.32
5	1000	0.01	0.00	16.53	-0.18	-0.57	19.64	1.80	2.24	16.74	2.56	4.85	19.77
5	2000	-0.10	-0.13	16.55	-0.16	-0.55	19.87	1.26	1.61	16.66	1.77	3.49	19.94
5	8000	-0.02	-0.02	16.70	-0.05	-0.12	20.10	0.65	0.83	16.73	0.91	1.79	20.11
6	100	0.75	-1.41	12.91	0.50	-7.39	15.90	8.12	6.78	14.62	9.35	15.46	16.80
6	200	0.23	-0.57	13.12	0.10	-3.44	17.71	5.41	4.58	14.03	6.33	9.73	18.18
6	500	-0.08	-0.21	13.44	-0.36	-1.31	19.18	2.24	2.78	13.83	3.37	5.65	19.36
6	1000	0.02	-0.02	13.78	-0.11	-0.51	19.87	1.58	2.00	13.98	2.30	3.93	19.95
6	2000	-0.05	-0.04	13.76	-0.12	-0.29	20.14	1.12	1.35	13.86	1.63	2.66	20.18
6	8000	-0.02	-0.02	13.89	-0.05	-0.11	20.33	0.55	0.69	13.91	0.81	1.36	20.34
10	100	0.48	-1.68	6.17	0.54	-6.44	10.01	4.82	5.68	7.92	6.22	10.79	11.13
10	200	0.16	-0.72	6.11	0.08	-2.99	11.15	2.72	3.46	6.95	3.73	6.19	11.80
10	500	-0.02	-0.30	6.21	-0.07	-1.10	12.27	1.67	2.11	6.57	2.01	3.33	12.57
10	1000	-0.01	-0.09	6.30	-0.12	-0.49	12.54	1.13	1.44	6.49	1.36	2.27	12.71
10	2000	-0.02	-0.06	6.36	-0.07	-0.24	12.78	0.77	1.00	6.45	0.92	1.58	12.87
10	8000	-0.01	-0.02	6.46	-0.02	-0.06	12.94	0.38	0.49	6.48	0.45	0.76	12.96

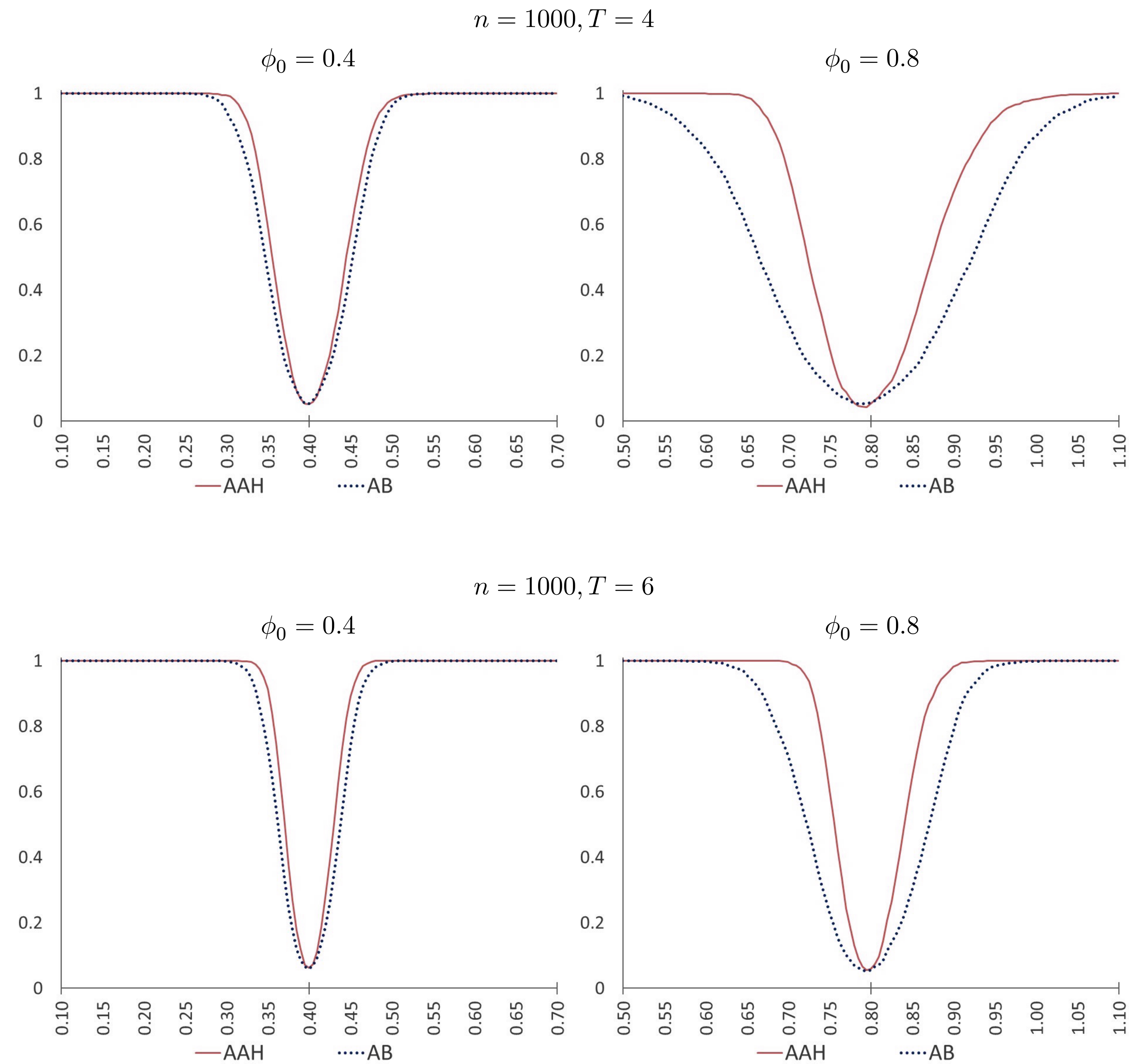
See the notes to Tables 2 and 6.

Table 9: Size and power of AAH, AB and BB estimators in experiments when Arellano and Bond (AB) restrictions are met and Blundell and Bond (BB) restrictions are not met

		Size (5% level, ×100)						**Power (5% level, ×100, $H_1 : \phi = \phi_0 + 0.1$)**					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	5.4	7.0	93.2	7.4	11.0	72.6	24.4	21.4	81.6	19.3	16.5	41.7
3	200	5.3	6.3	98.7	6.2	7.7	86.0	32.0	29.5	90.1	22.2	14.3	56.4
3	500	4.2	5.1	100.0	3.7	5.5	98.1	54.6	52.5	99.4	32.5	15.5	84.9
3	1000	4.1	4.8	100.0	3.7	5.6	99.8	71.3	78.8	100.0	44.7	20.8	97.1
3	2000	4.4	4.5	100.0	4.1	4.9	100.0	82.3	97.4	100.0	69.2	32.8	100.0
3	8000	3.2	5.0	100.0	4.7	5.0	100.0	86.2	100.0	100.0	93.3	80.2	100.0
4	100	9.3	9.4	95.3	12.1	11.9	92.9	40.7	33.3	82.6	30.8	21.4	73.4
4	200	6.9	8.7	99.6	8.8	8.4	98.1	56.0	45.4	91.7	35.7	20.0	90.4
4	500	6.3	6.4	100.0	6.3	6.0	100.0	84.8	76.7	99.5	50.7	23.8	99.3
4	1000	5.2	5.4	100.0	5.3	5.8	100.0	98.0	96.1	100.0	69.7	38.0	100.0
4	2000	4.8	6.1	100.0	5.6	5.9	100.0	100.0	100.0	100.0	91.4	58.9	100.0
4	8000	5.6	5.1	100.0	5.5	4.8	100.0	100.0	100.0	100.0	100.0	98.7	100.0
5	100	14.3	13.2	89.3	16.5	17.0	91.1	57.3	46.5	65.2	48.3	33.9	72.9
5	200	10.7	9.5	98.2	10.8	11.3	97.1	75.6	61.3	68.4	57.1	32.5	84.6
5	500	6.8	5.8	100.0	6.5	6.7	100.0	96.8	88.4	86.0	81.4	40.1	96.4
5	1000	6.2	3.6	100.0	6.4	4.2	100.0	99.9	99.7	95.5	96.6	58.5	99.6
5	2000	4.8	4.9	100.0	5.7	5.2	100.0	100.0	100.0	99.7	100.0	85.6	99.9
5	8000	5.3	5.5	100.0	5.7	5.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
6	100	17.5	16.8	90.0	19.9	23.2	96.6	67.4	60.4	61.4	55.3	47.0	78.8
6	200	10.4	10.0	97.5	12.8	12.6	99.0	84.5	73.9	59.0	64.3	44.6	90.4
6	500	7.1	6.3	100.0	8.4	6.4	100.0	99.5	97.3	68.4	87.2	57.3	98.6
6	1000	6.4	6.0	100.0	5.9	5.6	100.0	100.0	99.8	81.9	98.2	78.5	100.0
6	2000	5.4	4.5	100.0	5.8	4.3	100.0	100.0	100.0	93.9	100.0	96.6	100.0
6	8000	4.1	4.6	100.0	4.9	5.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	100	46.1	44.1	84.8	48.1	53.3	95.3	90.4	90.5	77.6	85.0	87.5	77.7
10	200	22.1	21.4	89.4	26.4	26.5	98.8	98.8	97.5	76.7	95.3	88.9	72.1
10	500	12.9	10.7	98.6	14.0	10.9	100.0	100.0	100.0	86.6	99.7	97.8	72.8
10	1000	8.9	8.3	99.9	10.3	7.6	100.0	100.0	100.0	95.6	100.0	99.9	81.5
10	2000	6.7	6.6	100.0	7.1	6.3	100.0	100.0	100.0	99.5	100.0	100.0	91.9
10	8000	5.0	5.0	100.0	5.2	5.7	100.0	100.0	100.0	100.0	100.0	100.0	99.9

See the notes to Tables 2 and 6.

Figure 4: Rejection frequencies (at 5% nominal level) for AAH and AB estimators when AB restrictions are met and BB restrictions are not met



See the notes to Tables 2 and 6.

Table 10: Bias and RMSE of AAH, AB and BB estimators when Arellano and Bond (AB) restrictions are met and Blundell and Bond (BB) restrictions are not met

		Bias ($\times$100)						**RMSE**($\times$100)					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	1.17	-27.22	12.78	2.70	-82.80	7.92	14.13	30.67	16.15	23.20	108.71	21.86
3	200	0.59	-26.09	12.59	2.85	-74.35	8.66	8.41	27.98	14.48	18.57	88.08	16.75
3	500	0.56	-24.65	12.53	1.56	-68.10	9.24	4.99	25.58	13.36	8.97	75.64	11.54
3	1000	0.47	-24.05	12.47	1.58	-63.24	9.18	3.45	24.59	12.88	6.14	68.08	10.45
3	2000	0.23	-23.84	12.29	0.82	-62.46	9.06	2.42	24.13	12.52	4.01	64.99	9.79
3	8000	0.13	-23.48	12.30	0.19	-59.44	9.19	1.19	23.57	12.36	1.89	60.21	9.40
4	100	1.20	-13.32	14.81	0.54	-49.05	12.90	11.70	16.12	16.92	14.23	56.54	16.03
4	200	0.31	-11.97	14.90	0.32	-40.60	14.06	6.95	13.77	15.99	10.33	44.93	15.38
4	500	-0.02	-11.00	15.29	-0.02	-34.86	15.01	3.30	11.83	15.75	5.89	36.98	15.52
4	1000	0.06	-10.57	15.32	0.09	-32.28	15.43	2.36	11.07	15.55	4.03	33.64	15.66
4	2000	-0.08	-10.24	15.33	-0.10	-30.79	15.76	1.65	10.53	15.45	2.84	31.56	15.88
4	8000	-0.02	-9.99	15.42	-0.02	-29.56	15.93	0.85	10.07	15.45	1.44	29.78	15.95
5	100	1.02	-10.51	10.92	-0.18	-35.45	11.22	9.96	13.01	13.09	10.03	40.12	13.81
5	200	0.39	-9.59	11.07	-0.29	-30.79	12.01	7.15	11.02	12.10	6.45	33.16	13.26
5	500	0.01	-8.64	11.46	-0.23	-26.83	12.61	2.58	9.34	11.88	3.75	27.93	13.17
5	1000	0.01	-8.37	11.53	-0.18	-25.52	12.80	1.80	8.73	11.76	2.56	26.10	13.12
5	2000	-0.10	-8.20	11.53	-0.16	-24.67	12.89	1.26	8.41	11.65	1.77	25.00	13.06
5	8000	-0.02	-7.88	11.65	-0.05	-23.68	13.01	0.65	7.94	11.68	0.91	23.79	13.06
6	100	0.75	-7.79	9.94	0.50	-24.23	12.63	8.12	10.26	11.87	9.35	27.95	13.92
6	200	0.23	-6.59	10.05	0.10	-19.63	13.64	5.41	8.05	11.09	6.33	21.65	14.42
6	500	-0.08	-5.89	10.27	-0.36	-16.82	14.47	2.24	6.56	10.71	3.37	17.70	14.80
6	1000	0.02	-5.49	10.55	-0.11	-15.52	14.93	1.58	5.89	10.78	2.30	16.02	15.10
6	2000	-0.05	-5.42	10.54	-0.12	-15.08	15.08	1.12	5.63	10.65	1.63	15.35	15.16
6	8000	-0.02	-5.24	10.64	-0.05	-14.53	15.20	0.55	5.30	10.67	0.81	14.61	15.22
10	100	0.48	-3.18	5.53	0.54	-9.76	9.00	4.82	6.28	7.42	6.22	13.10	10.21
10	200	0.16	-2.28	5.52	0.08	-6.46	9.78	2.72	4.07	6.41	3.73	8.44	10.48
10	500	-0.02	-1.85	5.63	-0.07	-4.51	10.57	1.67	2.77	6.02	2.01	5.48	10.89
10	1000	-0.01	-1.62	5.74	-0.12	-3.88	10.69	1.13	2.15	5.94	1.36	4.43	10.87
10	2000	-0.02	-1.58	5.80	-0.07	-3.63	10.84	0.77	1.86	5.90	0.92	3.93	10.93
10	8000	-0.01	-1.53	5.91	-0.02	-3.44	10.93	0.38	1.60	5.93	0.45	3.52	10.96

See the notes to Tables 2 and 6.

Table 11: Size and power of AAH, AB and BB estimators in experiments when AB and BB restrictions are not met

		Size (5% level, ×100)						Power (5% level, ×100, $H_1 : \phi = \phi_0 + 0.1$)					
		$\phi_0 = 0.4$			$\phi_0 = 0.8$			$\phi_0 = 0.4$			$\phi_0 = 0.8$		
T	n	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB	AAH	AB	BB
3	100	5.4	51.1	68.1	7.4	48.1	56.0	24.4	77.5	40.0	19.3	54.7	33.6
3	200	5.3	74.0	79.7	6.2	54.1	59.6	32.0	94.5	43.6	22.2	62.1	34.2
3	500	4.2	95.0	94.8	3.7	71.3	68.2	54.6	99.7	44.7	32.5	79.0	35.7
3	1000	4.1	99.3	99.8	3.7	84.2	79.1	71.3	99.9	49.4	44.7	91.0	38.7
3	2000	4.4	100.0	100.0	4.1	95.9	92.1	82.3	100.0	55.4	69.2	98.6	41.5
3	8000	3.2	100.0	100.0	4.7	100.0	100.0	86.2	100.0	83.7	93.3	100.0	44.8
4	100	9.3	36.0	84.6	12.1	59.3	81.2	40.7	75.3	55.1	30.8	73.0	50.8
4	200	6.9	46.4	93.8	8.8	63.5	89.3	56.0	89.4	58.3	35.7	79.5	58.7
4	500	6.3	71.5	99.7	6.3	80.7	97.9	84.8	99.7	73.9	50.7	94.3	70.7
4	1000	5.2	90.8	100.0	5.3	93.8	100.0	98.0	100.0	87.0	69.7	98.8	86.4
4	2000	4.8	99.2	100.0	5.6	99.3	100.0	100.0	100.0	96.7	91.4	100.0	96.0
4	8000	5.6	100.0	100.0	5.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5	100	14.3	37.5	79.9	16.5	62.9	82.5	57.3	83.1	51.2	48.3	83.5	56.8
5	200	10.7	46.5	91.3	10.8	72.0	89.7	75.6	95.8	44.2	57.1	92.0	53.0
5	500	6.8	69.2	99.5	6.5	90.4	97.9	96.8	100.0	44.2	81.4	99.0	59.1
5	1000	6.2	90.8	100.0	6.4	98.5	99.8	99.9	100.0	49.5	96.6	100.0	68.2
5	2000	4.8	99.1	100.0	5.7	100.0	100.0	100.0	100.0	58.5	100.0	100.0	77.9
5	8000	5.3	100.0	100.0	5.7	100.0	100.0	100.0	100.0	87.7	100.0	100.0	97.3
6	100	17.5	34.9	83.0	19.9	60.8	91.5	67.4	87.3	53.9	55.3	85.6	66.0
6	200	10.4	37.3	91.4	12.8	62.6	95.3	84.5	96.6	49.6	64.3	92.3	71.3
6	500	7.1	53.7	99.4	8.4	82.8	99.5	99.5	100.0	45.5	87.2	99.6	81.4
6	1000	6.4	74.8	100.0	5.9	95.2	100.0	100.0	100.0	46.1	98.2	100.0	91.2
6	2000	5.4	94.0	100.0	5.8	99.6	100.0	100.0	100.0	45.5	100.0	100.0	97.7
6	8000	4.1	100.0	100.0	4.9	100.0	100.0	100.0	100.0	56.3	100.0	100.0	100.0
10	100	46.1	47.2	83.0	48.1	63.8	93.6	90.4	94.7	78.8	85.0	93.8	77.3
10	200	22.1	27.1	87.1	26.4	44.5	98.0	98.8	99.6	80.8	95.3	97.0	68.4
10	500	12.9	22.8	97.3	14.0	37.5	100.0	100.0	100.0	91.1	99.7	100.0	64.3
10	1000	8.9	24.8	99.9	10.3	46.4	100.0	100.0	100.0	98.0	100.0	100.0	65.1
10	2000	6.7	36.8	100.0	7.1	66.2	100.0	100.0	100.0	99.9	100.0	100.0	67.9
10	8000	5.0	86.9	100.0	5.2	99.6	100.0	100.0	100.0	100.0	100.0	100.0	79.8

See the notes to Tables 2 and 6.

Table 12: Empirical size and power of Hausman test applied to the difference between BB and AAH estimators at the 5% nominal level

		Rejection rates (×100)				Fraction of replications (×100) where Hausman test was not applicable due to $\widehat{Var}\left(\hat{\phi}^{aah}\right) - \widehat{Var}\left(\hat{\phi}^{bb}\right) < 0$			
		under H_0		under H_1		under H_0		under H_1	
T	n	$\phi_0 = 0.4$	$\phi_0 = 0.8$	$\phi_0 = 0.4$	$\phi_0 = 0.8$	$\phi_0 = 0.4$	$\phi_0 = 0.8$	$\phi_0 = 0.4$	$\phi_0 = 0.8$
3	100	6.64	9.03	68.67	30.28	20.95	13.60	0.40	8.70
3	200	5.80	7.09	81.28	44.49	20.65	12.60	0.10	5.60
3	500	4.71	6.03	84.45	72.06	24.60	11.30	0.00	1.95
3	1000	6.60	6.17	89.75	78.79	23.45	7.65	0.00	0.30
3	2000	8.31	5.95	98.05	82.85	20.55	3.35	0.00	0.00
3	8000	7.22	6.05	100.00	93.25	7.15	0.00	0.00	0.00
4	100	15.27	17.60	93.54	54.08	28.60	11.65	0.10	2.65
4	200	9.92	13.08	99.65	70.67	28.95	11.70	0.00	0.95
4	500	9.41	8.91	100.00	93.25	27.20	9.05	0.00	0.05
4	1000	7.26	8.34	100.00	99.45	26.30	4.65	0.00	0.00
4	2000	8.88	7.28	100.00	100.00	22.30	2.45	0.00	0.00
4	8000	8.41	5.50	100.00	100.00	9.05	0.05	0.00	0.00
5	100	22.13	25.25	87.50	71.27	27.25	9.90	0.00	2.20
5	200	13.10	18.03	99.05	89.37	32.80	11.00	0.00	0.75
5	500	9.59	13.24	100.00	99.50	35.85	11.65	0.00	0.00
5	1000	8.40	10.73	100.00	100.00	36.90	7.70	0.00	0.00
5	2000	6.55	7.92	100.00	100.00	33.55	4.65	0.00	0.00
5	8000	8.26	7.01	100.00	100.00	26.75	0.15	0.00	0.00
6	100	23.32	33.62	88.25	77.15	25.40	6.30	0.00	0.45
6	200	18.06	23.41	98.70	92.80	30.80	7.95	0.00	0.00
6	500	9.90	15.79	100.00	99.85	36.90	9.75	0.00	0.00
6	1000	6.60	14.47	100.00	100.00	39.40	8.45	0.00	0.00
6	2000	6.38	9.78	100.00	100.00	36.55	3.35	0.00	0.00
6	8000	5.97	7.07	100.00	100.00	30.50	0.30	0.00	0.00
10	100	44.92	58.80	79.58	86.42	20.20	4.50	0.10	0.25
10	200	21.62	41.71	91.05	97.10	34.10	9.25	0.00	0.00
10	500	9.29	25.83	99.55	99.95	39.20	18.70	0.00	0.00
10	1000	5.65	20.95	100.00	100.00	45.10	22.90	0.00	0.00
10	2000	4.29	13.21	100.00	100.00	42.95	21.65	0.00	0.00
10	8000	3.63	11.40	100.00	100.00	44.90	14.45	0.00	0.00

Notes: Reported rejection rates under the null correspond to DGP with $\rho = 0$ and $\kappa = 0$, i.e. both AB and BB conditions are met. Reported rejection rate under alternative correspond to DGP with $\rho = 0$ and $\kappa \neq 0$, i.e. BB conditions are not met, whilst AB conditions still hold. See also the notes to Tables 2 and 4.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27. https://doi.org/10.1016/0304-4076(94)01641-c.

Ahn, S. C. and P. Schmidt (1997). Efficient estimation of models for dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics 76*, 309–321. https://doi.org/10.1016/0304-4076(95)01793-3.

Anderson, T. and B. Sorenson (1996). GMM estimation of a stochastic volatility model: A monte carlo study. *Journal of Business and Economic Statistics 14*, 328–352. https://doi.org/10.2307/1392446.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*, 598–606. https://doi.org/10.2307/2287517.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*, 47–82. https://doi.org/10.1016/0304-4076(82)90095-1.

Andrews, D. W. and B. Lu (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics 101*(1), 123 – 164. https://doi.org/10.1016/s0304-4076(00)00077-4.

Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica 67*(3), 543–563. https://doi.org/10.1111/1468-0262.00036.

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics 14*(1), 57–67. https://doi.org/10.2307/223249.

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. https://doi.org/10.2307/2297968.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51. https://doi.org/10.1016/0304-4076(94)01642-d.

Bai, J. (2013). Fixed-effects dynamic panel models, a factor analytical method. *Econometrica 81*, 285–314. https://doi.org/10.3982/ecta9409.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica 62*, 657–681. https://doi.org/10.2307/2951662.

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. https://doi.org/10.1016/s0304-4076(98)00009-8.

Breitung, J., K. Hayakawa, and S. Kripfganz (2019, November). Asymptotically efficient method of moments estimators for dynamic panel data models. mimeo.

Bruno, G. S. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters 87*, 361–366. https://doi.org/10.1016/j.econlet.2005.01.005.

Bun, M. (2003). Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews 22*, 29–58. https://doi.org/10.1081/etc-120017973.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of business & economic statistics 23*, 200–210. ttps://doi.org/10.1198/073500104000000532.

Bun, M. J. and M. A. Carree (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters 92*, 220–227. https://doi.org/10.1016/j.econlet.2006.02.008.

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. https://doi.org/10.1016/s0165-1765(02)00299-9.

Chao, J. C. and N. R. Swanson (2003). Asymptotic normality of single-equation estimators for the case with a large number of weak instruments. available at: http://www.sas.rutgers.edu/virtual/snde/wp/2003-12.pdf.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012). Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments. *Econometric Theory 28*(1), 42–86. https://doi.org/10.2307/41426507.

Chatelain, J.-B. (2007). Improving consistent moment selection procedures for generalized method of moments estimation. *Economics Letters 95*(3), 380 – 385. https://doi.org/10.1016/j.econlet.2006.11.011.

Cheng, X. and Z. Liao (2015). Select the valid and relevant moments: An information-based LASSO for GMM with many moments. *Journal of Econometrics 186*(2), 443 – 464. https://doi.org/10.1016/j.jeconom.2015.02.019.

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. https://doi.org/10.1111/j.1468-0084.2010.00594.x.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica 86*(4), 1479–1512. https://doi.org/10.3982/ECTA14176.

Chudik, A. and M. H. Pesaran (2017). A bias-corrected method of moments approach for estimation of dynamic panels. Federal Reserve Bank of Dallas, Globalization and Monetary Policy Working Paper No. 327, https://doi.org/10.24149/gwp327, and CESifo Working Paper Series No. 6688, available at SSRN: https://ssrn.com/abstract=3072243.

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 (2018). Half-panel jackknife fixed effects estimation of panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*, 816–836. https://doi.org/10.1002/jae.2623.

Clark, T. E. (1996). Small-sample properties of estimators of nonlinear models of covariance structure. *Journal of Business & Economic Statistics 14*(3), 367–373. https://doi.org/10.2307/1392448.

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. https://doi.org/10.1093/restud/rdv007.

Dhaene, G. and K. Jochmans (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32*, 1178–1215. https://doi.org/10.1017/s0266466615000146.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics 117*, 55–93. https://doi.org/10.1016/s0304-4076(03)00118-0.

Donald, S. G., G. W. Imbens, and W. K. Newey (2009). Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics 152*(1), 28–36. https://doi.org/10.1016/j.jeconom.2008.10.013.

Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. *Econometrica 69*(5), 1161–1191. https://doi.org/10.1111/1468-0262.00238.

Everaert, G. and L. Ponzi (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control 31*, 1160–1184. https://doi.org/10.1016/j.jedc.2006.04.006.

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. https://doi.org/10.1111/1468-0262.00344.

Hahn, J. and H. Moon (2006). Reducing bias of MLE in a dynamic panel model. *Econometric Theory 22*, 499–512. https://doi.org/10.1017/s0266466606060245.

Hall, A. R., A. Inoue, K. Jana, and C. Shin (2007). Information in generalized method of moments estimation and entropy-based moment selection. *Journal of Econometrics 138*, 488–512. https://doi.org/10.1016/j.jeconom.2006.05.006.

Hall, A. R. and F. P. M. Peixe (2003). A consistent method for the selection of relevant instruments. *Econometric Reviews 22*(3), 269–287. https://doi.org/10.1081/ETC-120024752.

Han, C. and P. C. B. Phillips (2006). GMM with many moment conditions. *Econometrica 74*, 147–192. https://doi.org/10.1111/j.1468-0262.2006.00652.x.

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. https://doi.org/10.1017/s026646660909063x.

Han, C., P. C. B. Phillips, and D. Sul (2014). X-Differencing and dynamic panel model estimation. *Econometric Theory 30*, 201–251. https://doi.org/10.1017/s0266466613000170.

Hansen, C. and D. Kozbur (2014). Instrumental variables estimation with many weak instruments using regularized JIVE. *Journal of Econometrics 182*(2), 290 – 308. https://doi.org/10.1016/j.jeconom.2014.04.022.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics 14*(3), 262–280. https://doi.org/10.2307/1392442.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251 – 1271. https://doi.org/10.2307/1913827.

Hayakawa, K. (2012). GMM estimation of short dynamic panel data models with interactive fixed effects. *Journal of the Japan Statistical Society 42*, 109–123. https://doi.org/10.14490/jjss.42.109.

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. https://doi.org/10.1016/j.jeconom.2015.03.042.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395. https://doi.org/10.2307/1913103.

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. https://doi.org/10.1016/s0304-4076(01)00143-9.

Hsiao, C. and J. Zhang (2015). IV, GMM or likelihood approach to estimate dynamic panel models when either N or T or both are large. *Journal of Econometrics 187*(1), 312 – 322.

Inoue, A. (2006). A bootstrap approach to moment selection. *The Econometrics Journal 9*(1), 48–75. https://doi.org/10.1111/j.1368-423x.2006.00176.x.

Juodis, A. (2013). A note on bias-corrected estimation in dynamic panel data models. *Economics Letters 118*, 435–438. https://doi.org/10.1016/j.econlet.2012.12.013.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics 68*, 53–78. https://doi.org/10.1016/0304-4076(94)01643-e.

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. https://doi.org/10.1017/cbo9780511493140.010.

Koenker, R. and J. A. F. Machado (1999). GMM inference when the number of moment conditions is large. *Journal of Econometrics 93*, 327–344. https://doi.org/10.1016/S0304-4076(99)00014-7.

Kuersteiner, G. M. (2002). Mean square error reduction for GMM estimators of linear time series models. mimeo, available at http://econweb.umd.edu/ kuersteiner/Research/mrg.pdf, March 15, 2002.

Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies 69*, 647–666. https://doi.org/10.1111/1467-937x.t01-1-00025.

Lee, N., H. R. Moon, and Q. Zhou (2017). Many IVs estimation of dynamic panel regression models with measurement error. *Journal of Econometrics 200*(2), 251–259. https://doi.org/10.1016/j.jeconom.2017.06.009.

Liao, Z. (2013). Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory 29*(5), 857–904. https://doi.org/10.1017/s0266466612000783.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics 31*, 451–472. https://doi.org/10.1080/07350015.2013.818003.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*, 687–719. https://doi.org/10.3982/ecta6224.

Pesaran, M. H. (2015). *Time series and panel data econometrics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198736912.001.0001.

Pesaran, M. H. and C. F. Yang (2020). Estimation and inference in spatial models with dominant units. USC-INET Research Paper No. 19-06, March 2019, revised January 2020, revised version available at pesaran.com.

Phillips, G. D. A. and C. Hale (1977). The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review 18*(1), 219–228. https://doi.org/10.2307/2525779.

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. https://doi.org/10.1016/j.jeconom.2006.03.009.

Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal 6*, 217–259. https://doi.org/10.1111/1368-423x.00108.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586. https://doi.org/10.2307/2171753.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*, 1055–1096. https://doi.org/10.1111/1468-0262.00151.

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

Zhang, Y. and Q. Zhou (2020). Correction for the asymptotical bias of the arellano-bond type GMM estimation of dynamic panel models. *Advances in Econometrics, Essays in Honor of Cheng Hsiao 41*, 1–24. forthcoming.

A Appendix

This appendix is organized as follows. Section A.1 derives $\bar{B}_3$ given by (30) and (31). Section A.2 states and proves a number of lemmas used in the rest of this appendix. Additional propositions and proofs are given in Section A.3. Section A.4 provides derivation of conditional model for y_{it} when $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is generated from a panel VAR model with heteroskedastic errors.

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

Using (29), it readily follows that $\bar{B}_3 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E(B_{i3})$, where

$$E(B_{i3}) = E\left(\Delta y_{i1}^2\right) + E\left(\Delta y_{i2}^2\right) + 2E\left(\Delta u_{i2}\Delta y_{i1}\right). \tag{A.1}$$

Also recall that $\Delta y_{i1} = u_{i1} - (1-\phi)(y_{i0} - \mu_i)$, and $\Delta y_{i2} = \phi \Delta y_{i1} + \Delta u_{i2}$. Hence $E(\Delta u_{i2}\Delta y_{i1}) = -\sigma_{i1}^2$,

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

and

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

Using the above results in (A.1) now yields:

$$E(B_{i3}) = \left(\sigma_{i2}^2 - \sigma_{i1}^2\right) + (1-\phi)^2 \sigma_{i1}^2 + \left(1+\phi^2\right)\left\{(1-\phi)^2 E(y_{i0} - \mu_i)^2 - 2(1-\phi) E\left[u_{i1}(y_{i0} - \mu_i)\right]\right\}.$$

Hence, as required, we have

$$\bar{B}_3 = \bar{\sigma}_2^2 - \bar{\sigma}_1^2 + (1-\phi)^2 \bar{\sigma}_1^2 + \left(1+\phi^2\right)(1-\phi)\psi_0, \tag{A.3}$$

where $\bar{\sigma}_t^2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \sigma_{it}^2$, for $t = 1, 2$, and

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

A.2 Lemmas

Lemma A.1 *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*

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

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*

$$\bar{Q}_{nT}=E\left(\bar{Q}_{nT}\right)+O_p\left(n^{-1/2}\right), \tag{A.4}$$

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

Proof. Under Assumptions 1-3, the fourth moments of u_{it} and b_i are bounded, and hence, using Loève's inequality,[19] 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.4).

Result (A.5) 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.4). Using the same arguments as in the proof of (A.4), 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.5) 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*

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

[19] 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*

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

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*

$$\sqrt{n}\bar{V}_{nT} \to_d N\left(0, S_T\right). \tag{A.7}$$

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))

$$E\left(V_{iT}\right) = \frac{1}{T-2}\sum_{t=2}^{T-1} E\left(\Delta u_{it}\Delta y_{i,t-1} + \Delta u_{it}^2 + \Delta u_{i,t+1}\Delta y_{it}\right) = 0. \tag{A.8}$$

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,[20] 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

$$\sup_i E\left(\left|V_{iT}\right|^{2+\epsilon/2}\right) < K. \tag{A.9}$$

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.6) follows. To establish (A.7), we note that (A.9) 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 Propositions and Proofs

Theorem 1 is established in the main text. This section presents propositions for the consistency of $\hat{\Sigma}_{nT}$.

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

Proposition 1 *Suppose conditions of Theorem 1 hold, and consider $\hat{\Sigma}_{nT}$ defined by (34), 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,

$$\hat{\Sigma}_{nT} \to_p \Sigma_T, \tag{A.10}$$

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 $\Delta\hat{u}_{it} - \Delta u_{it} = -\left(\hat{\phi}_{nT} - \phi_0\right) \Delta y_{i,t-1} = O_p\left(n^{-1/2}\right)$. This implies $\hat{H}_{i,nT}$ is consistent, which in turn implies $\hat{\bar{B}}_{nT} - \bar{B}_{nT} \to_p 0$. But, using result (A.5) of Lemma A.1, we have $\bar{B}_{nT} \to_p E\left(\bar{B}_{nT}\right)$, and $E\left(\bar{B}_{nT}\right) \to B_T$. Therefore $\hat{\bar{B}}_{nT} \to_p \bar{B}_T$. Since $\bar{B}_T > 0$ by assumption, it follows that

$$\hat{\bar{B}}_{nT}^{-2} \to_p \bar{B}_T^{-2}. \tag{A.11}$$

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

$$n^{-1} \sum_{i=1}^{n} \left(\hat{V}_{i,nT} - V_{iT} \right)^2 \to_p 0, \text{ and } n^{-1} \sum_{i=1}^{n} \left(\hat{V}_{i,nT} - V_{iT} \right) V_{iT} \to_p 0. \tag{A.12}$$

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.9), we have $\sup_i E\left|V_{iT}\right|^{2+\epsilon/2} < K$ for some $\epsilon > 0$. It follows that $n^{-1} \sum_{i=1}^{n} \left[V_{iT}^2 - E\left(V_{iT}^2\right) \right] \to_p 0$, and therefore (noting that $n^{-1} \sum_{i=1}^{n} E\left(V_{iT}^2\right) \to S_T$ by assumption) we have

$$n^{-1} \sum_{i=1}^{n} V_{iT}^2 \to_p S_T. \tag{A.13}$$

Result (A.10) now follows from (A.11), (A.12), and (A.13). ■

A.4 Derivation of conditional model for y_{it} when $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is given by a panel VAR model

Suppose $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$ is generated from panel VAR(1) model given by equation (42) in the paper. Individual equations for y_{it} and $\mathbf{x}_{it}$ in (42) are

$$
\begin{aligned}
y_{it} &= \alpha_{iy} + \phi_{11} y_{i,t-1} + \boldsymbol{\phi}'_{yx}\mathbf{x}_{i,t-1} + u_{y,it}, & \text{(A.14)}\\
\mathbf{x}_{it} &= \boldsymbol{\alpha}_{ix} + \boldsymbol{\phi}_{xy} y_{i,t-1} + \boldsymbol{\Phi}_{xx}\mathbf{x}_{i,t-1} + \mathbf{u}_{x,it}, & \text{(A.15)}
\end{aligned}
$$

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:

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

Suppose that the errors, $\mathbf{u}_{it}$, are heteroskedastic over i and t, and 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

$$
u_{y,it} = \boldsymbol{\theta}'_{it}\mathbf{u}_{x,it} + \eta_{it}, \tag{A.16}
$$

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 (A.16) and (A.15) in (A.14), we have

$$
\begin{aligned}
y_{it} &= \alpha_{iy} + \phi_{11} y_{i,t-1} + \boldsymbol{\phi}'_{yx}\mathbf{x}_{i,t-1} + \boldsymbol{\theta}'_{it}\left(\mathbf{x}_{it} - \boldsymbol{\alpha}_{ix} - \boldsymbol{\phi}_{xy} y_{i,t-1} - \boldsymbol{\Phi}_{xx}\mathbf{x}_{i,t-1}\right) + \eta_{it},\\
&= \left(\alpha_{iy} - \boldsymbol{\theta}'_{it}\boldsymbol{\alpha}_{ix}\right) + \left(\phi_{11} - \boldsymbol{\theta}'_{it}\boldsymbol{\phi}_{xy}\right) y_{i,t-1} + \left(\boldsymbol{\phi}'_{yx} - \boldsymbol{\theta}'_{it}\boldsymbol{\Phi}_{xx}\right)\mathbf{x}_{i,t-1} + \eta_{it}, & \text{(A.17)}
\end{aligned}
$$

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 (A.17) will have homogeneous slopes only if $\boldsymbol{\theta}_{it} = \boldsymbol{\Omega}^{-1}_{xx,it}\boldsymbol{\omega}_{xy,it} = \boldsymbol{\theta}$ for all i and t.