Econometric Theory, **25**, 2009, 873–890. Printed in the United States of America.
doi:10.1017/S0266466609090707

A SIMPLE EFFICIENT INSTRUMENTAL VARIABLE ESTIMATOR FOR PANEL AR(p) MODELS WHEN BOTH N AND T ARE LARGE

KAZUHIKO HAYAKAWA
Hiroshima University

In this paper, we show that for panel AR(p) models, an instrumental variable (IV) estimator with instruments deviated from past means has the same asymptotic distribution as the infeasible optimal IV estimator when both N and T, the dimensions of the cross section and time series, are large. If we assume that the errors are normally distributed, the asymptotic variance of the proposed IV estimator is shown to attain the lower bound when both N and T are large. A simulation study is conducted to assess the estimator.

1. INTRODUCTION

In estimating a dynamic panel data model, several studies propose to use instrumental variables (IVs).[1] However, if we use IVs, a well-known problem of the trade-off between the bias and efficiency of the estimator arises, i.e., although the IV estimator is less biased than the generalized method of moments (GMM) estimator in finite sample, it is generally less efficient than the GMM estimator.

The purpose of this paper is to address the efficiency problem of the IV estimator. In particular, we show that for panel AR(p) models, a simple one-step IV estimator using instruments deviated from past means has the same asymptotic distribution as the infeasible optimal IV estimator derived by Arellano (2003a) when both N and T are large.[2] If normality is assumed with regard to the errors, the proposed IV estimator is shown to be asymptotically efficient. This implies that the proposed IV estimator addresses the problem of the trade-off between the

The author is deeply grateful to the co-editor, two anonymous referees, Kaddour Hadri, Chirok Han, Cheng Hsiao, Naoto Kunitomo, Eiji Kurozumi, Kosuke Oya, Peter Phillips, Donggyu Sul, Taku Yamamoto, and the participants of the 14th International Conference of Panel Data at Xiamen University, the Fall meeting of the Japanese Economic Association at Nihon University, the Hitotsubashi Conference on Econometrics 2007, and the special 18th meeting of the New Zealand Econometric Study Group at the University of Auckland for helpful comments. The author also acknowledges Ryo Okui, who posed a question that inspired this paper. The research benefited from the JSPS fellowship and a Grant-in-Aid for Scientific Research (KAKENHI 20830056) of the JSPS. All remaining errors are mine. Finally, this paper is dedicated to the late Professor Satoru Kanoh, who provided useful comments on an early version of this paper. Address correspondence to Kazuhiko Hayakawa, Department of Economics, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8525, Japan; e-mail: kazuhaya@hiroshima-u.ac.jp.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

bias and efficiency. In comparison to the existing estimators, there are two advantages in the proposed IV estimator. The first is that although the within groups (WG) estimator is consistent only when T is large and the GMM estimator is consistent only when N is large, the proposed IV estimator is consistent under large N and fixed T, fixed N and large T, or large N and large T asymptotics. This implies that the proposed IV estimator can be used for large N and small T, small N and large T, or large N and large T panel data.[3] The second advantage is that the proposed IV estimator is more efficient than the Anderson and Hsiao (1981) IV estimator and as efficient as the WG and GMM estimators when both N and T are large. Simulation results reveal that the proposed IV estimator is almost unbiased and the difference in dispersions between the feasible optimal IV estimator and the proposed IV estimator is small when T is large.

The remainder of this paper is organized as follows. Section 2 provides the setup and the main results. Section 3 presents a Monte Carlo simulation and assesses the theoretical result. Finally, Section 4 concludes.

A word on notation. For a vector $\mathbf{x}$ and a matrix $\mathbf{A}$, we define $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ and $\|\mathbf{A}\|^2 = \mathrm{tr}(\mathbf{A}'\mathbf{A})$ where $\mathrm{tr}(\cdot)$ denotes the trace operator.

2. SETUP AND RESULT

2.1. The Model and Assumptions

Let us consider the following panel AR(p) model:

$$\begin{aligned} y_{it} &= \alpha_1 y_{i,t-1} + \alpha_2 y_{i,t-2} + \cdots + \alpha_p y_{i,t-p} + \eta_i + \upsilon_{it} \\ &= \boldsymbol{\alpha}'\mathbf{x}_{it} + \eta_i + \upsilon_{it} \qquad (i = 1, \ldots, N, \qquad t = 1, \ldots, T), \end{aligned} \tag{1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, $\mathbf{x}_{it} = (y_{i,t-1}, \ldots, y_{i,t-p})'$, υ_{it} has zero mean given η_i, $y_{i,1-p}, \ldots, y_{i,t-1}$, and p is fixed and known.[4] For convenience, we assume that $y_{i,0}, \ldots, y_{i,1-p}$ are observed.

Expression (1) can be written in a companion form as

$$\mathbf{x}_{i,t+1} = \boldsymbol{\Pi}\mathbf{x}_{it} + \mathbf{d}_1(\eta_i + \upsilon_{it}), \tag{2}$$

where $\mathbf{d}_1 = (1, 0, \ldots, 0)'$ of dimension p and $\boldsymbol{\Pi}$ is the $p \times p$ matrix given by

$$\boldsymbol{\Pi} = \begin{pmatrix} \alpha_1 & \cdots & \alpha_p \\ \mathbf{I}_{p-1} & | & \mathbf{O}_{(p-1)\times 1} \end{pmatrix},$$

where $\mathbf{I}_k$ is an identity matrix of order k and $\mathbf{O}_{k\times\ell}$ is a $k \times \ell$ matrix of zeros.

We make the following assumptions, which are part of the assumptions made by Lee (2007).

Assumption 1. $\{\upsilon_{it}\}$ $(t = 1, \ldots, T, i = 1, \ldots, N)$ are independent and identically distributed (i.i.d.) over i and t and independent of η_i and $\mathbf{x}_{i1}$, with $\mathrm{E}(\upsilon_{it}) = 0$, $\mathrm{var}(\upsilon_{it}) = \sigma_\upsilon^2$, and finite fourth-order moment. Individual effects $\{\eta_i\}$ $(i = 1, \ldots, N)$ are i.i.d. over i with $\mathrm{E}(\eta_i) = 0$ and $\mathrm{var}(\eta_i) = \sigma_\eta^2$.

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

Assumption 2. The initial observations satisfy

$$\mathbf{x}_{i1} = (\mathbf{I}_p - \mathbf{\Pi})^{-1}\mathbf{d}_1\eta_i + \mathbf{w}_{i0},$$

where $\mathbf{w}_{i0} = \left(\sum_{j=0}^{\infty}\mathbf{\Pi}^j v_{i,-j}\right)\mathbf{d}_1$.

Assumption 3. $\det\left[\mathbf{I}_p - \mathbf{\Pi}z\right] \neq 0$ for all $|z| \leq 1$.

Assumption 4. Let $m_j(i,t) = \mathbf{\Pi}^j\mathbf{d}_1 v_{i,t-1-j}$. For all i, t, and for any $r_1, \ldots, r_4 \in \{1, 2, \ldots, p\}$,

$$\sum_{j_1,\ldots,j_4=0}^{\infty} |\mathrm{cum}_{r_1,\ldots,r_4}(m_{j_1}(i,t), m_{j_2}(i,t), m_{j_3}(i,t), m_{j_4}(i,t))| < \infty,$$

where $\mathrm{cum}_{r_1,\ldots,r_4}(m_{j_1}(i,t), m_{j_2}(i,t), m_{j_3}(i,t), m_{j_4}(i,t))$ denotes the fourth-order cross-cumulant of $(m_{j_1}^{r_1}(i,t), m_{j_2}^{r_2}(i,t), m_{j_3}^{r_3}(i,t), m_{j_4}^{r_4}(i,t))$ with $m_j^r(i,t)$ being the rth element of $m_j(i,t)$.

Unlike Lee (2007), we do not need to impose the asymptotic relative ratio between N and T. Assumptions 1 and 2 are standard in the literature.[5] Although Assumption 2 can be relaxed to nonstationary initial conditions, we do not pursue this here for the purpose of simplicity.[6] However, the main result of this paper is expected to hold because the initial conditions are negligible when T is large and because we do not use moment conditions that rely on stationary initial conditions as Blundell and Bond (1998) do. Assumption 3 is the stability condition, and Assumption 4 is used to show that the generalized Lindeberg–Feller condition in the central limit theorem for double indexed processes is satisfied.[7]

Under Assumptions 2 and 3, $\mathbf{x}_{it}$ can be written as

$$\mathbf{x}_{i,t+1} = (\mathbf{I}_p - \mathbf{\Pi})^{-1}\mathbf{d}_1\eta_i + \mathbf{w}_{it},$$

where $\mathbf{w}_{it} = \left(\sum_{j=0}^{\infty}\mathbf{\Pi}^j v_{i,t-j}\right)\mathbf{d}_1$.

To remove the individual effects, η_i, we use the forward orthogonal deviation (FOD) because the errors transformed by the FOD are serially uncorrelated and homoskedastic if the original errors are.[8] Specifically, the model to be estimated is given by

$$y_{it}^* = \boldsymbol{\alpha}'\mathbf{x}_{it}^* + v_{it}^* \qquad (i = 1, \ldots, N, \qquad t = 1, \ldots, T-1), \tag{3}$$

where $y_{it}^* = c_t[y_{it} - (y_{i,t+1} + \cdots + y_{iT})/(T-t)]$, $\mathbf{x}_{it}^* = c_t[\mathbf{x}_{it} - (\mathbf{x}_{i,t+1} + \cdots + \mathbf{x}_{iT})/(T-t)]$, $v_{it}^* = c_t[v_{it} - (v_{i,t+1} + \cdots + v_{iT})/(T-t)]$, and $c_t^2 = (T-t)/(T-t+1)$.

2.2. The Instrumental Variable Estimators

2.2.1. The Infeasible Optimal Instruments. Following Arelano (2003a, 2003b), the infeasible optimal IV estimator in a large N and small T context takes the form

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

$$\widehat{\boldsymbol{\alpha}}_{IV}^{OPT} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\mathbf{h}_{it}\mathbf{x}_{it}^{*\prime}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\mathbf{h}_{it}y_{it}^{*}\right) = \boldsymbol{\alpha} + \left(\widehat{\mathbf{A}}_{IV}^{OPT}\right)^{-1}\widehat{\mathbf{b}}_{IV}^{OPT},$$

where $\mathbf{h}_{it} = \mathrm{E}(\mathbf{x}_{it}^{*}|\mathbf{y}_{i}^{t-1})$ and $\mathbf{y}_{i}^{t-1} = (y_{i,t-1},\ldots,y_{i,1-p})'$. Estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{OPT}$ is an ideal estimator because it is consistent and asymptotically efficient when N is large and T is fixed. However, the drawback of this estimator is that it is infeasible because the optimal instruments $\mathbf{h}_{it}$ are unknown. A standard approach to obtain a feasible optimal IV estimator is to use a sample linear projection of $\mathbf{x}_{it}^{*}$ on $\mathbf{y}_{i}^{t-1}$, which is given by

$$\widehat{\mathbf{h}}_{it} = \left(\sum_{i=1}^{N}\mathbf{x}_{it}^{*}\mathbf{y}_{i}^{t-1\prime}\right)\left(\sum_{i=1}^{N}\mathbf{y}_{i}^{t-1}\mathbf{y}_{i}^{t-1\prime}\right)^{-1}\mathbf{y}_{i}^{t-1}.$$

In this case, the feasible optimal IV estimator is equivalent to the GMM estimator using $\mathbf{y}_{i}^{t-1}$ as instruments:

$$\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV} = \left(\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{X}_{t}^{*\prime}\mathbf{M}_{t}^{LEV}\mathbf{X}_{t}^{*}\right)^{-1}\left(\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{X}_{t}^{*\prime}\mathbf{M}_{t}^{LEV}\mathbf{y}_{t}^{*}\right), \tag{4}$$

where $\mathbf{X}_{t}^{*} = (\mathbf{x}_{1t}^{*},\ldots,\mathbf{x}_{Nt}^{*})'$, $\mathbf{M}_{t}^{LEV} = \mathbf{Z}_{t}^{LEV}\left(\mathbf{Z}_{t}^{LEV\prime}\mathbf{Z}_{t}^{LEV}\right)^{-1}\mathbf{Z}_{t}^{LEV\prime}$, $\mathbf{Z}_{t}^{LEV} = (\mathbf{y}_{1}^{t-1},\ldots,\mathbf{y}_{N}^{t-1})'$, and $\mathbf{y}_{t}^{*} = (y_{1t}^{*},\ldots,y_{Nt}^{*})'$.

However, one problem of $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ is that if N and T increase at the same rate, the estimate of $\mathbf{h}_{it}$ is asymptotically biased (see Arellano, 2003b, p. 170). This causes a bias in $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$. In fact, for the case of $p = 1$, Alvarez and Arellano (2003) show that $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ has a bias of the order $O(1/N)$.[9]

Thus, in this paper, we propose an alternative approach. Instead of estimating the optimal instruments, we propose to use an *observable* variable that has the same structure as the optimal instruments $\mathbf{h}_{it}$. Hence, we need to investigate the structure of $\mathbf{h}_{it}$. Arellano (2003a) shows that, under the assumption that $\mathrm{E}(\mu_i|\mathbf{y}_{i}^{t-1})$ coincides with the linear projection, the infeasible optimal instruments can be rewritten in the following form:

$$\begin{aligned}\mathrm{E}(\mathbf{x}_{it}^{*}|\mathbf{y}_{i}^{t-1}) &= c_t\left[\mathbf{I}_p - \frac{\boldsymbol{\Pi}(\mathbf{I}_p - \boldsymbol{\Pi}^{T-t})(\mathbf{I}_p - \boldsymbol{\Pi})^{-1}}{T-t}\right]\left[\mathbf{x}_{it} - \boldsymbol{\iota}_p\mathrm{E}\left(\mu_i|\mathbf{y}_{i}^{t-1}\right)\right] \\ &= c_t\left[\mathbf{I}_p - \frac{\boldsymbol{\Pi}(\mathbf{I}_p - \boldsymbol{\Pi}^{T-t})(\mathbf{I}_p - \boldsymbol{\Pi})^{-1}}{T-t}\right]\left[\mathbf{w}_{i,t-1}\right. \\ &\qquad \left. + \boldsymbol{\iota}_p\left\{\mu_i - \mathrm{E}(\mu_i|\mathbf{y}_{i}^{t-1})\right\}\right] \qquad (5) \\ &= c_t\left[\mathbf{I}_p - O\left(\frac{1}{T-t}\right)\right]\left[\mathbf{w}_{i,t-1} + O_p\left(\frac{1}{\sqrt{t}}\right)\right], \qquad (6)\end{aligned}$$

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

where the second equality arises from the fact that $\mathbf{x}_{it} = \boldsymbol{\iota}_p \mu_i + \mathbf{w}_{i,t-1}, \mu_i = \eta_i/(1-\boldsymbol{\alpha}'\boldsymbol{\iota}_p)$, and the third equality is proved in Lemma A (see the Appendix).

From (5) and (6), we find that (a) the individual effect μ_i is demeaned in (5) and (b) when t is large, $\mathbf{w}_{i,t-1}$ is the dominating term in (6).

Our next task is to find an observable variable that has the same structure as (6).

2.2.2. Instruments Deviated from Past Means. We consider instruments $\mathbf{z}_{it(\ell)} = \left(z_{it(\ell)}^{t-1}, z_{it(\ell)}^{t-2}, \ldots, z_{it(\ell)}^{t-p}\right)'$ where

$$z_{it(\ell)}^{t-m} = c_t \left[y_{i,t-m} - \frac{y_{i,t-m-\ell} + \cdots + y_{i,-p+1}}{t+p-\ell-m} \right] \qquad (m = 1, \ldots, p)$$

with $\ell \geq 0$ being fixed.[10] Because $\mathbf{z}_{it(\ell)}$ is deviated from past means, it can be seen as a modification of the recursive mean adjustment (RMA) method of So and Shin (1999).[11]

At this point, we show that $\mathbf{z}_{it(\ell)}$ meets the preceding two requirements. For the first requirement, it is straightforward to show that the individual effects are demeaned because $\mathbf{z}_{it(\ell)}$ is deviated from past means. For the second requirement, we show in the Appendix that $\mathbf{z}_{it(\ell)}$ can be written as

$$\mathbf{z}_{it(\ell)} = c_t \left[\mathbf{w}_{i,t-1} + O_p\left(\frac{1}{\sqrt{t}}\right) \right]. \tag{7}$$

Thus, comparing (6) and (7), we find that unobservable $\mathbf{h}_{it}$ and observable $\mathbf{z}_{it(\ell)}$ have the same structure, i.e., (a) demeaning individual effects and (b) $\mathbf{w}_{i,t-1}$ is dominating.

The IV estimator using $\mathbf{z}_{it(\ell)}$ as instruments is given by

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell} &= \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=t_0}^{T-1} \mathbf{z}_{it(\ell)} \mathbf{x}_{it}^{*\prime} \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=t_0}^{T-1} \mathbf{z}_{it(\ell)} y_{it}^{*} \right) \\ &= \boldsymbol{\alpha} + \left(\widehat{\mathbf{A}}_{IV}^{RMA\ell} \right)^{-1} \widehat{\mathbf{b}}_{IV}^{RMA\ell}, \end{aligned} \tag{8}$$

where $t_0 = 2$ for $\ell = 0, 1$ and $t_0 = \ell + 1$ for $\ell \geq 2$. Note that a similar estimator is also proposed by Moon and Phillips (2000, pp. 944–945) to avoid bias.

The following proposition establishes the asymptotic equivalence of the infeasible optimal IV estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{OPT}$ and the proposed IV estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ in the sense that both estimators have the same asymptotic distribution.

PROPOSITION 1. *Let Assumptions 1–3 hold. Then, for fixed $\ell \geq 0$, as both N and T tend to infinity, the infeasible optimal IV estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{OPT}$ and the feasible IV estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ are consistent. If we further assume that Assumption 4 holds,*

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

then, as both N and T tend to infinity, we have

$$\sqrt{NT}(\widehat{\boldsymbol{\alpha}}_{IV}-\boldsymbol{\alpha})\xrightarrow{d}\mathcal{N}\left(0,\sigma_v^2[\mathrm{E}(\boldsymbol{w}_{i,t-1}\boldsymbol{w}_{i,t-1}')]^{-1}\right), \tag{9}$$

where $\widehat{\boldsymbol{\alpha}}_{IV}$ *denotes* $\widehat{\boldsymbol{\alpha}}_{IV}^{OPT}$ *and* $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$.

The proof is given in the Appendix. Note that the asymptotic variance $\sigma_v^2[\mathrm{E}(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}')]^{-1}$ is of the same form as the WG estimator derived by Lee (2007).

Remark 1. For the case of $p=1$, Alvarez and Arellano (2003) show that $\widehat{\alpha}_{GMM}^{LEV}$ and the WG estimator, $\widehat{\alpha}_{WG}$, have the following asymptotic distribution:

$$\sqrt{NT}\left[\widehat{\alpha}_{GMM}^{LEV}-\left(\alpha_1-\frac{1}{N}(1+\alpha_1)\right)\right]\xrightarrow{d}\mathcal{N}(0,1-\alpha_1^2), \tag{10}$$

$$\sqrt{NT}\left[\widehat{\alpha}_{WG}-\left(\alpha_1-\frac{1}{T}(1+\alpha_1)\right)\right]\xrightarrow{d}\mathcal{N}(0,1-\alpha_1^2). \tag{11}$$

Moreover, from Proposition 1, we have

$$\sqrt{NT}\left[\widehat{\alpha}_{IV}^{RMA\ell}-\alpha_1\right]\xrightarrow{d}\mathcal{N}\left(0,1-\alpha_1^2\right). \tag{12}$$

Comparing (10), (11), and (12), we find that although all estimators have the same asymptotic variance, $\widehat{\alpha}_{GMM}^{LEV}$ and $\widehat{\alpha}_{WG}$ have asymptotic biases of the order $O(1/N)$ and $O(1/T)$, respectively, whereas the distribution of $\widehat{\alpha}_{IV}^{RMA\ell}-\alpha_1$ is centered at zero. This is because $\widehat{\alpha}_{GMM}^{LEV}$ and $\widehat{\alpha}_{WG}$ suffer from the "many instruments problem" and "incidental parameter problem," respectively, whereas $\widehat{\alpha}_{IV}^{RMA\ell}$ does not.[12]

Remark 2. Hahn and Kuersteiner (2002) show that if we further assume normality on v_{it}, then $\sigma_v^2\left[\mathrm{E}(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}')\right]^{-1}$ is equal to the lower bound under large N and T asymptotics.[13] Hence, $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ is an efficient IV estimator under large N and T asymptotics without an asymptotic bias when v_{it} is normally distributed.

Remark 3. Another feature of $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ is that because the individual effects are completely eliminated from both the model and instruments under stationary initial conditions, the performance of $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ is not affected by the variance ratio of the individual effects to the disturbances although the typical GMM estimators using instruments in levels are.[14]

Remark 4. Although we use large N and T asymptotics to derive the asymptotic properties, consistency and asymptotic normality are also obtained under large N and fixed T or fixed N and large T asymptotics. Especially, under fixed N and large T asymptotics, the same expression as (9) is expected to hold. This is in marked contrast to the GMM estimator, where large N is required. Furthermore, although the GMM estimator can be computed only when $T-1\leq N$, the proposed IV estimator can be computed for any N and T.

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

3. MONTE CARLO SIMULATION

In this section, we compare $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ with other estimators by Monte Carlo simulation. We consider AR(1) and AR(2) models. The terms v_{it} and η_i are drawn from $\mathcal{N}(0,1)$ independently. We consider the cases of $(T,N) = (10,100)$, $(10,500)$, $(15,100)$, $(15,300)$, $(20,100)$, $(20,200)$, $(50,100)$, and $(100,100)$. For the AR(1) model, we set $\alpha_1 = 0.5, 0.9$, and for the AR(2) model, we set $(\alpha_1,\alpha_2) = (0.6, -0.1), (0.6, 0.3)$. We generate $T+p+50$ observations for each i and discard the first 50 periods to diminish the effect of initial conditions. We compute the median (Median), the interquartile range (IQR), and the median absolute error (MAE). The number of replications is 5,000 for all cases. The simulation results for the AR(1) and AR(2) models are provided in Tables 1 and 2, respectively.

We consider the GMM and IV estimators using instruments in levels or deviated from past means. The GMM estimator using $\mathbf{y}_i^{t-1}$ as instruments is defined as (4). The GMM estimator using $\mathbf{z}_{it(1)}$ as instruments is defined by $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$.[15] The estimator $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ does not share the problem with $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ that the number of instruments is too large. In fact, the number of instruments used in $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ is $O(T)$, whereas that in $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ is $O(T^2)$.

For the proposed IV estimators, we consider $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA0},\ldots,\widehat{\boldsymbol{\alpha}}_{IV}^{RMA4}$ as defined by (8). Moreover, for the purpose of comparison, we consider an IV estimator $\widehat{\boldsymbol{\alpha}}_{IV}^{LEV}$ that uses $\mathbf{x}_{it}$ as instruments. Note that $\widehat{\boldsymbol{\alpha}}_{IV}^{LEV}$ is not exactly the same IV estimator as the one suggested by Anderson and Hsiao (1981, 1982) because they used the first difference to remove the individual effects from the model.

For the choice of ℓ, we find from both tables that in terms of MAE, $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA1}$ performs the best in many cases. To describe the intuition behind this result, we consider the AR(1) model and $\ell = 0,1,2$. In this case, the instruments are

$$z_{it(\ell)} = c_t\left[y_{i,t-1} - \frac{y_{i,t-1-\ell}+\cdots+y_{i,0}}{t-\ell}\right], \qquad (\ell = 0,1,2).$$

For the case of $\ell = 2$, we find that $y_{i,t-2}$ is not used, and this causes an efficiency loss. The same result applies to the case $\ell \geq 2$. For the case of $\ell = 0,1$, although both instruments use all the past variables, $z_{it(1)}$ is more strongly correlated with the endogenous variable x_{it}^* than $z_{it(0)}$. In fact, after some algebra, we have $\mathrm{E}\left(x_{it}^* z_{it(1)}\right) > \mathrm{E}\left(x_{it}^* z_{it(0)}\right)$. Thus, $\mathbf{z}_{it(1)}$ should be used in practice.

First, we consider the AR(1) case. We find from Table 1 that, in terms of the bias, the IV estimators, $\widehat{\boldsymbol{\alpha}}_{IV}^{LEV}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$, have little bias for all cases, whereas the GMM estimators have nonnegligible bias when $\alpha = 0.9$. In particular, $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ has significant bias for all cases. However, with regard to the IQR, the feasible optimal IV estimator, $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$, has the smallest dispersion and $\widehat{\boldsymbol{\alpha}}_{IV}^{LEV}$ has the largest dispersion. Moreover, we find that the differences in the IQR of $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$, $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ become quite small when T is as large as 50. This result is consistent with Proposition 1, where $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ are shown to have the same asymptotic variance when N and T are large. Also, the asymptotic variances in (10) and (12) suggest that for given N and T, the dispersion of $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ should decrease as α_1 grows. However, the simulation results do not show such

TABLE 1. Simulation results for the AR(1) model

T	N	α_1	$\widehat{\alpha}_{GMM}^{LEV}$	$\widehat{\alpha}_{GMM}^{RMA1}$	$\widehat{\alpha}_{IV}^{LEV}$	$\widehat{\alpha}_{IV}^{RMA0}$	$\widehat{\alpha}_{IV}^{RMA1}$	$\widehat{\alpha}_{IV}^{RMA2}$	$\widehat{\alpha}_{IV}^{RMA3}$	$\widehat{\alpha}_{IV}^{RMA4}$
						Median				
10	100	0.5	0.467	0.493	0.500	0.499	0.499	0.498	0.497	0.497
10	500	0.5	0.493	0.499	0.501	0.500	0.500	0.500	0.499	0.500
15	100	0.5	0.472	0.495	0.499	0.498	0.498	0.498	0.498	0.499
15	300	0.5	0.490	0.499	0.501	0.500	0.500	0.499	0.499	0.500
20	100	0.5	0.476	0.498	0.500	0.500	0.500	0.499	0.499	0.499
20	200	0.5	0.488	0.499	0.500	0.500	0.500	0.500	0.500	0.500
50	100	0.5	0.482	0.499	0.500	0.500	0.500	0.500	0.500	0.500
100	100	0.5	0.484	0.500	0.500	0.500	0.500	0.500	0.500	0.500
10	100	0.9	0.691	0.805	0.906	0.891	0.894	0.889	0.884	0.882
10	500	0.9	0.827	0.878	0.903	0.897	0.897	0.899	0.898	0.898
15	100	0.9	0.767	0.862	0.902	0.896	0.896	0.897	0.896	0.897
15	300	0.9	0.833	0.885	0.902	0.899	0.899	0.899	0.900	0.900
20	100	0.9	0.802	0.881	0.900	0.899	0.899	0.898	0.898	0.898
20	200	0.9	0.837	0.892	0.902	0.900	0.900	0.900	0.900	0.899
50	100	0.9	0.860	0.897	0.899	0.900	0.900	0.899	0.899	0.899
100	100	0.9	0.875	0.899	0.900	0.900	0.900	0.900	0.900	0.900
						IQR				
10	100	0.5	0.075	0.084	0.117	0.088	0.086	0.092	0.103	0.118
10	500	0.5	0.034	0.038	0.050	0.039	0.039	0.041	0.046	0.052
15	100	0.5	0.050	0.055	0.082	0.057	0.056	0.057	0.062	0.067
15	300	0.5	0.030	0.032	0.048	0.033	0.032	0.034	0.035	0.038
20	100	0.5	0.040	0.043	0.069	0.044	0.044	0.044	0.046	0.047
20	200	0.5	0.028	0.030	0.048	0.030	0.031	0.031	0.032	0.033
50	100	0.5	0.020	0.021	0.038	0.021	0.021	0.021	0.021	0.021
100	100	0.5	0.013	0.014	0.025	0.014	0.014	0.014	0.014	0.014
10	100	0.9	0.156	0.222	0.407	0.291	0.278	0.289	0.314	0.367
10	500	0.9	0.086	0.114	0.174	0.127	0.125	0.130	0.139	0.159
15	100	0.9	0.085	0.117	0.237	0.143	0.139	0.137	0.142	0.150
15	300	0.9	0.058	0.073	0.136	0.082	0.079	0.078	0.081	0.086
20	100	0.9	0.059	0.078	0.176	0.088	0.087	0.086	0.088	0.090
20	200	0.9	0.045	0.057	0.120	0.064	0.062	0.062	0.062	0.065
50	100	0.9	0.018	0.021	0.066	0.023	0.023	0.023	0.023	0.023
100	100	0.9	0.008	0.010	0.037	0.011	0.011	0.011	0.011	0.011

continued

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

TABLE 1. *Continued.*

T	N	α_1	$\widehat{\alpha}_{GMM}^{LEV}$	$\widehat{\alpha}_{GMM}^{RMA1}$	$\widehat{\alpha}_{IV}^{LEV}$	$\widehat{\alpha}_{IV}^{RMA0}$	$\widehat{\alpha}_{IV}^{RMA1}$	$\widehat{\alpha}_{IV}^{RMA2}$	$\widehat{\alpha}_{IV}^{RMA3}$	$\widehat{\alpha}_{IV}^{RMA4}$
						MAE				
10	100	0.5	0.043	0.042	0.058	0.044	0.043	0.046	0.051	0.059
10	500	0.5	0.018	0.019	0.025	0.019	0.020	0.020	0.023	0.026
15	100	0.5	0.032	0.028	0.040	0.028	0.028	0.028	0.031	0.033
15	300	0.5	0.016	0.016	0.024	0.016	0.016	0.017	0.017	0.019
20	100	0.5	0.027	0.022	0.034	0.022	0.022	0.022	0.023	0.024
20	200	0.5	0.016	0.015	0.024	0.015	0.015	0.015	0.016	0.017
50	100	0.5	0.018	0.010	0.019	0.010	0.010	0.011	0.011	0.011
100	100	0.5	0.016	0.007	0.012	0.007	0.007	0.007	0.007	0.007
10	100	0.9	0.209	0.131	0.203	0.146	0.139	0.146	0.158	0.185
10	500	0.9	0.075	0.059	0.087	0.064	0.063	0.065	0.070	0.080
15	100	0.9	0.133	0.064	0.118	0.072	0.070	0.069	0.071	0.075
15	300	0.9	0.067	0.037	0.067	0.041	0.039	0.039	0.040	0.043
20	100	0.9	0.098	0.040	0.087	0.044	0.043	0.043	0.044	0.045
20	200	0.9	0.063	0.029	0.060	0.032	0.031	0.031	0.031	0.033
50	100	0.9	0.040	0.011	0.032	0.012	0.011	0.012	0.012	0.011
100	100	0.9	0.025	0.005	0.018	0.005	0.005	0.005	0.005	0.005

a tendency when $T \leq 50$. This may be due to the weak identification problem as α_1 approaches unity.[16] For the median absolute error, we find that $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ has the smallest MAE in many cases. However, the difference in the MAE between $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ is fairly small.

Next, we discuss the results for the AR(2) case. The IV estimators are virtually median unbiased, and $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ has the largest bias. In terms of the IQR, $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$ has the smallest dispersion in all cases. Also, we find that the difference in the IQR between $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$, $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$, and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ becomes small when T is large. In terms of the MAE, although $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ performs best in many cases, the difference between $\widehat{\boldsymbol{\alpha}}_{GMM}^{RMA1}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ is quite small.

The simulation results suggest that, in terms of the bias and MAE, the IV and GMM estimators using the proposed instruments perform better than the commonly used IV and GMM estimators, $\widehat{\boldsymbol{\alpha}}_{IV}^{LEV}$ and $\widehat{\boldsymbol{\alpha}}_{GMM}^{LEV}$, even when T is as large as 10.

4. CONCLUSION

In this paper, we showed that the infeasible optimal IV estimator and the IV estimator using instruments deviated from past means are asymptotically equivalent in the sense that both estimators have the same asymptotic distribution when both N and T are large. We further showed that if we assume normality on the errors, the proposed IV estimator is asymptotically efficient when both N and T

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

TABLE 2. Simulation results for an AR(2) model

T	N	α_1	α_2	$\widehat{\alpha}_{GMM}^{LEV}$	$\widehat{\alpha}_{GMM}^{RMA1}$	$\widehat{\alpha}_{IV}^{LEV}$	$\widehat{\alpha}_{IV}^{RMA0}$	$\widehat{\alpha}_{IV}^{RMA1}$	$\widehat{\alpha}_{IV}^{RMA2}$	$\widehat{\alpha}_{IV}^{RMA3}$	$\widehat{\alpha}_{IV}^{RMA4}$
						Median					
10	100	0.6		0.565	0.586	0.599	0.598	0.597	0.598	0.597	0.598
10	100		−0.1	−0.117	−0.108	−0.101	−0.102	−0.102	−0.103	−0.104	−0.102
10	500	0.6		0.593	0.598	0.600	0.600	0.600	0.600	0.600	0.599
10	500		−0.1	−0.104	−0.102	−0.101	−0.101	−0.101	−0.101	−0.101	−0.101
15	100	0.6		0.577	0.595	0.601	0.600	0.600	0.600	0.600	0.599
15	100		−0.1	−0.114	−0.103	−0.101	−0.101	−0.101	−0.101	−0.101	−0.101
15	300	0.6		0.591	0.598	0.600	0.600	0.600	0.600	0.600	0.599
15	300		−0.1	−0.105	−0.101	−0.100	−0.100	−0.100	−0.100	−0.100	−0.100
20	100	0.6		0.590	0.599	0.601	0.600	0.600	0.600	0.600	0.600
20	100		−0.1	−0.106	−0.101	−0.100	−0.100	−0.100	−0.100	−0.100	−0.100
20	200	0.6		0.590	0.599	0.601	0.600	0.600	0.600	0.600	0.600
20	200		−0.1	−0.106	−0.101	−0.100	−0.100	−0.100	−0.100	−0.100	−0.100
50	100	0.6		0.588	0.600	0.601	0.601	0.601	0.600	0.601	0.601
50	100		−0.1	−0.110	−0.101	−0.100	−0.100	−0.100	−0.100	−0.100	−0.100
100	100	0.6		0.590	0.600	0.600	0.600	0.600	0.600	0.600	0.600
100	100		−0.1	−0.109	−0.100	−0.101	−0.100	−0.100	−0.100	−0.100	−0.100
						IQR					
10	100	0.6		0.067	0.077	0.101	0.079	0.079	0.084	0.094	0.111
10	100		−0.1	0.051	0.056	0.064	0.058	0.059	0.063	0.068	0.076
10	500	0.6		0.032	0.035	0.046	0.036	0.036	0.038	0.042	0.050
10	500		−0.1	0.024	0.026	0.029	0.027	0.026	0.029	0.031	0.035
15	100	0.6		0.046	0.050	0.070	0.050	0.050	0.051	0.054	0.059
15	100		−0.1	0.038	0.041	0.050	0.043	0.043	0.045	0.047	0.051
15	300	0.6		0.028	0.029	0.041	0.030	0.030	0.031	0.032	0.034
15	300		−0.1	0.023	0.024	0.029	0.024	0.025	0.026	0.027	0.029
20	100	0.6		0.027	0.029	0.041	0.029	0.029	0.029	0.030	0.031
20	100		−0.1	0.024	0.025	0.031	0.026	0.026	0.026	0.027	0.028
20	200	0.6		0.027	0.029	0.041	0.029	0.029	0.029	0.030	0.031
20	200		−0.1	0.024	0.025	0.031	0.026	0.026	0.026	0.027	0.028
50	100	0.6		0.020	0.021	0.031	0.021	0.021	0.021	0.022	0.022
50	100		−0.1	0.019	0.020	0.026	0.020	0.020	0.020	0.021	0.021
100	100	0.6		0.014	0.014	0.020	0.014	0.014	0.014	0.014	0.014
100	100		−0.1	0.014	0.014	0.019	0.014	0.014	0.014	0.014	0.014
						MAE					
10	100	0.6		0.043	0.040	0.050	0.040	0.040	0.042	0.047	0.055
10	100		−0.1	0.028	0.029	0.032	0.029	0.029	0.032	0.034	0.038
10	500	0.6		0.016	0.017	0.023	0.018	0.018	0.019	0.021	0.025
10	500		−0.1	0.012	0.013	0.014	0.013	0.013	0.014	0.016	0.018
15	100	0.6		0.029	0.025	0.035	0.025	0.025	0.026	0.027	0.030
15	100		−0.1	0.021	0.021	0.025	0.021	0.021	0.022	0.023	0.025
15	300	0.6		0.015	0.015	0.021	0.015	0.015	0.015	0.016	0.017
15	300		−0.1	0.012	0.012	0.014	0.012	0.012	0.013	0.014	0.015
20	100	0.6		0.015	0.014	0.021	0.015	0.015	0.015	0.015	0.015
20	100		−0.1	0.013	0.013	0.015	0.013	0.013	0.013	0.014	0.014
20	200	0.6		0.015	0.014	0.021	0.015	0.015	0.015	0.015	0.015
20	200		−0.1	0.013	0.013	0.015	0.013	0.013	0.013	0.014	0.014
50	100	0.6		0.014	0.011	0.015	0.011	0.011	0.011	0.011	0.011
50	100		−0.1	0.012	0.010	0.013	0.010	0.010	0.010	0.010	0.010
100	100	0.6		0.011	0.007	0.010	0.007	0.007	0.007	0.007	0.007
100	100		−0.1	0.010	0.007	0.010	0.007	0.007	0.007	0.007	0.007

continued

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

TABLE 2. *Continued.*

T	N	α_1	α_2	$\widehat{\alpha}_{GMM}^{LEV}$	$\widehat{\alpha}_{GMM}^{RMA1}$	$\widehat{\alpha}_{IV}^{LEV}$	$\widehat{\alpha}_{IV}^{RMA0}$	$\widehat{\alpha}_{IV}^{RMA1}$	$\widehat{\alpha}_{IV}^{RMA2}$	$\widehat{\alpha}_{IV}^{RMA3}$	$\widehat{\alpha}_{IV}^{RMA4}$
						Median					
10	100	0.6		0.357	0.430	0.600	0.583	0.583	0.586	0.580	0.577
10	100		0.3	0.196	0.225	0.298	0.285	0.287	0.286	0.284	0.285
10	500	0.6		0.482	0.541	0.597	0.595	0.595	0.593	0.596	0.593
10	500		0.3	0.250	0.274	0.299	0.297	0.297	0.296	0.297	0.296
15	100	0.6		0.445	0.524	0.608	0.596	0.597	0.597	0.597	0.597
15	100		0.3	0.227	0.263	0.303	0.295	0.295	0.296	0.296	0.295
15	300	0.6		0.508	0.566	0.602	0.600	0.600	0.600	0.600	0.600
15	300		0.3	0.258	0.284	0.300	0.299	0.299	0.299	0.299	0.299
20	100	0.6		0.521	0.575	0.605	0.600	0.600	0.600	0.600	0.600
20	100		0.3	0.261	0.288	0.301	0.299	0.299	0.298	0.299	0.299
20	200	0.6		0.521	0.575	0.605	0.600	0.600	0.600	0.600	0.600
20	200		0.3	0.261	0.288	0.301	0.299	0.299	0.298	0.299	0.299
50	100	0.6		0.561	0.596	0.601	0.600	0.600	0.601	0.600	0.601
50	100		0.3	0.275	0.296	0.300	0.299	0.299	0.299	0.299	0.299
100	100	0.6		0.579	0.599	0.599	0.600	0.600	0.600	0.600	0.600
100	100		0.3	0.284	0.299	0.299	0.300	0.300	0.300	0.300	0.300
						IQR					
10	100	0.6		0.161	0.225	0.474	0.365	0.353	0.373	0.390	0.465
10	100		0.3	0.073	0.098	0.207	0.160	0.157	0.165	0.170	0.192
10	500	0.6		0.101	0.129	0.222	0.161	0.157	0.161	0.178	0.200
10	500		0.3	0.044	0.056	0.092	0.069	0.067	0.070	0.074	0.083
15	100	0.6		0.093	0.126	0.292	0.166	0.161	0.162	0.169	0.183
15	100		0.3	0.050	0.062	0.140	0.083	0.082	0.082	0.085	0.090
15	300	0.6		0.066	0.083	0.166	0.098	0.095	0.097	0.099	0.107
15	300		0.3	0.032	0.041	0.076	0.047	0.047	0.048	0.049	0.051
20	100	0.6		0.052	0.063	0.141	0.074	0.072	0.072	0.075	0.075
20	100		0.3	0.030	0.036	0.071	0.040	0.039	0.040	0.040	0.041
20	200	0.6		0.052	0.063	0.141	0.074	0.072	0.072	0.075	0.075
20	200		0.3	0.030	0.036	0.071	0.040	0.039	0.040	0.040	0.041
50	100	0.6		0.024	0.026	0.066	0.028	0.028	0.028	0.028	0.028
50	100		0.3	0.020	0.021	0.044	0.023	0.023	0.023	0.023	0.023
100	100	0.6		0.014	0.015	0.034	0.015	0.015	0.015	0.015	0.015
100	100		0.3	0.013	0.014	0.027	0.014	0.014	0.014	0.014	0.015
						MAE					
10	100	0.6		0.243	0.176	0.236	0.184	0.179	0.187	0.196	0.233
10	100		0.3	0.104	0.079	0.104	0.081	0.079	0.083	0.087	0.099
10	500	0.6		0.118	0.078	0.110	0.081	0.078	0.081	0.089	0.101
10	500		0.3	0.050	0.034	0.046	0.035	0.034	0.035	0.037	0.042
15	100	0.6		0.155	0.084	0.146	0.083	0.080	0.081	0.084	0.092
15	100		0.3	0.073	0.042	0.070	0.041	0.041	0.041	0.043	0.045
15	300	0.6		0.092	0.047	0.083	0.049	0.047	0.048	0.049	0.053
15	300		0.3	0.042	0.023	0.038	0.024	0.023	0.024	0.024	0.025
20	100	0.6		0.079	0.036	0.070	0.037	0.036	0.036	0.037	0.037
20	100		0.3	0.039	0.019	0.035	0.020	0.020	0.020	0.020	0.021
20	200	0.6		0.079	0.036	0.070	0.037	0.036	0.036	0.037	0.037
20	200		0.3	0.039	0.019	0.035	0.020	0.020	0.020	0.020	0.021
50	100	0.6		0.039	0.014	0.033	0.014	0.014	0.014	0.014	0.014
50	100		0.3	0.025	0.011	0.022	0.011	0.011	0.011	0.011	0.012
100	100	0.6		0.021	0.007	0.016	0.007	0.007	0.007	0.007	0.007
100	100		0.3	0.016	0.007	0.013	0.007	0.007	0.007	0.007	0.007

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

are large. Simulation results demonstrated that in terms of the bias and MAE, the new IV estimator outperforms the GMM and IV estimators using instruments in levels, which are commonly used in the literature.

Finally, we note certain possible extensions. Although we considered an AR(p) model with i.i.d. errors, it is interesting to investigate whether the results obtained in this paper apply to more general models and errors, say, models that include additional regressors other than the lagged dependent variables (Arellano, 2003a) and/or heteroskedastic errors (Alvarez and Arellano, 2004). Moreover, it may be interesting to apply the Okui (2008) method, i.e., a procedure to select the number of moment conditions so as to minimize the mean square error of the estimators, to improve the IV/GMM estimators using instruments deviated from past means. Finally, investigating the behavior of the proposed IV estimator in a near unit root case is important because in such a case, instruments may become weak. Recently, Han and Phillips (2009) proposed a new estimator that addresses the weak instruments problem at unity.[17] They showed that the estimator does not suffer from the weak instruments problem at unity, is consistent as long as NT is large, i.e., either N or T can be fixed, and has a very small finite-sample bias.[18] However, their estimator is not efficient, and a numerical optimization is required for AR(p) models (see Chowdhury, 1987), whereas our estimator is efficient and very easy to compute. Hence it may be interesting to address the efficiency problem of their estimator and compare it with the IV estimator proposed in this paper. However, these tasks are beyond the scope of the present paper and are left for future research.

NOTES

1. See Anderson and Hsiao (1981, 1982), Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995, 1997), and Blundell and Bond (1998), Arellano (2003a), Hahn, Hausman, and Kuersteiner (2007), and so on.

2. As pointed out by a referee, the same idea is already advocated by Moon and Phillips (2000, pp. 944–945). The difference between their paper and this paper is that their focus is on reducing bias, whereas ours is on improving efficiency.

3. Note that the WG and GMM estimators do not work well for large N and small T and small N and large T panel data, respectively.

4. The problem of how to choose p is extensively discussed by Lee (2007) but is not resolved.

5. See Alvarez and Arellano (2003) for the AR(1) case.

6. See Hayakawa (2008) for the effect of nonstationary initial conditions.

7. See Phillips and Moon (1999), Hahn and Kuersteiner (2002), and Lee (2007).

8. Note that taking a first difference induces a serial correlation in the errors, and this correlation changes the form of the optimal instruments defined in the next section.

9. Also see Bun and Kiviet (2006).

10. For the choice of ℓ, we consider $\ell = 0, \ldots, 4$ in simulation studies (see Section 3).

11. The case $\ell = 1$ corresponds to the original RMA method.

12. This interpretation was suggested by a referee.

13. Note that the AR(p) model (1) can be written as the VAR(1) model (2).

14. See Bun and Kiviet (2006), Hayakawa (2007b), and Bun and Windmeijer (2007).

15. The reason why we consider the GMM estimator using $\mathbf{z}_{it(1)}$ as instruments is that, in terms of MAE, the GMM estimator may perform better than the IV estimator because the GMM estimator is

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

more efficient than the IV estimator under large N and fixed T asymptotics. Also, the reason why we choose $\ell = 1$ is that the IV estimator with $\ell = 1$ performs the best.

16. Note that similar results are also reported in Alvarez and Arellano (2003).

17. Although the same estimator is considered by Chowdhury (1987), Ramalho (2005), and Hayakawa (2007a), it is Han and Phillips (2009) who rigorously derived the asymptotic properties.

18. In a time series context, Hayakawa (2006) derives the finite-sample bias.

REFERENCES

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

Ahn, S.C. & P. Schmidt (1997) Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics* 76, 309–321.

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

Alvarez, J. & M. Arellano (2004) Robust Likelihood Estimation of Dynamic Panel Data Models. CEMFI Working paper no. 0421.

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

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

Arellano, M. (2003a) Modelling Optimal Instrumental Variables for Dynamic Panel Data Models. CEMFI Working paper no. 0310.

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

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

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

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

Bun, M.J.G. & J.F. Kiviet (2006) The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 127, 409–444.

Bun, M.J.G. & F. Windmeijer (2007) The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. University of Bristol Economics Working paper no. 07/595.

Chowdhury, G. (1987) A note on correcting biases in dynamic panel models. *Applied Economics* 19, 31–37.

Hahn, J., J. Hausman, & G. Kuersteiner (2007) Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics* 127, 574–617.

Hahn, J. & 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.

Han, C. & P.C.B. Phillips (2009) GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory*, forthcoming.

Hayakawa, K. (2006) A note on bias in first-differenced AR(1) models. *Economics Bulletin* 3, 1–10.

Hayakawa, K. (2007a) Consistent OLS estimation of AR(1) dynamic panel data models with short time series. *Applied Economics Letters* 14, 1141–1145.

Hayakawa, K. (2007b) Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters* 95, 32–38.

Hayakawa, K. (2008) On the Effect of Nonstationary Initial Conditions in Dynamic Panel Data Models. Hi-Stat DP series, no. 245, Hitotsubashi University.

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

Lee, Y. (2007) A General Approach to Bias Correction in Dynamic Panels under Time Series Misspecification. Working paper, University of Michigan.

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

Moon, H.R. & P.C.B. Phillips (2000) Estimation of autoregressive roots near unity using panel data. *Econometric Theory* 16, 927–997.
Okui, R. (2008) The Optimal Choice of Moments in Dynamic Panel Data Models. Working paper, HKUST.
Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica*, 67, 1057–1111.
Ramalho, J. (2005) Feasible bias-corrected OLS, within-groups, and first-differences estimators for typical micro and macro AR(1) panel data models. *Empirical Economics* 30, 735–748.
So, B.S. & D.W. Shin (1999) Recursive mean adjustment in time series inferences. *Statistics and Probability Letters* 46, 65–73.
Whittle, P. (1951) *Hypothesis Testing in Time-Series Analysis*. Almqvist and Wiksell.
Wise, J. (1955) The autocorrelation function and the spectral density function. *Biometrika* 42, 151–159.

APPENDIX: Proof of Proposition 1

LEMMA A. *Let Assumptions 1–3 hold. Then,* $\mathbf{h}_{it}$ *and* $\mathbf{z}_{it(\ell)}$ *can be written as*

(i)

$$\mathbf{h}_{it} = \mathrm{E}(\mathbf{x}_{it}^{*}|\mathbf{y}_{i}^{t-1}) = c_t \left[\mathbf{I}_p - \frac{\boldsymbol{\Pi}(\mathbf{I}_p - \boldsymbol{\Pi}^{T-t})(\mathbf{I}_p - \boldsymbol{\Pi})^{-1}}{T-t}\right][\mathbf{w}_{i,t-1} + \boldsymbol{g}_{it}]$$

$$= c_t \left[\mathbf{I}_p - O\left(\frac{1}{T-t}\right)\right]\left[\mathbf{w}_{i,t-1} + O_p\left(\frac{1}{\sqrt{t}}\right)\right], \qquad \textbf{(A.1)}$$

(ii)

$$\mathbf{z}_{it(\ell)} = c_t \left[\mathbf{w}_{i,t-1} + \widetilde{\boldsymbol{g}}_{it}\right] \qquad \textbf{(A.2)}$$

$$= c_t \left[\mathbf{w}_{i,t-1} + O_p\left(\frac{1}{\sqrt{t}}\right)\right], \qquad \textbf{(A.3)}$$

where

$$\boldsymbol{g}_{it} = \boldsymbol{\iota}_p \left[\frac{\mu_i\left(1 + \phi\boldsymbol{\kappa}_p'\mathbf{R}^{-1}\boldsymbol{\kappa}_p\right) - \phi\left[(1 - \boldsymbol{\alpha}'\boldsymbol{\iota}_p)(v_{i,t-1} + \cdots + v_{i,1}) + \boldsymbol{\kappa}_p'\mathbf{R}^{-1}\mathbf{r}_i\right]}{1 + \phi\left\{(1 - \boldsymbol{\alpha}'\boldsymbol{\iota}_p)^2(t-1) + \boldsymbol{\kappa}_p'\mathbf{R}^{-1}\boldsymbol{\kappa}_p\right\}}\right],$$

$$\widetilde{\boldsymbol{g}}_{it} = \mu_i \boldsymbol{G}_p\boldsymbol{\iota}_p - (\mathbf{I}_p - \boldsymbol{G}_p)\left\{\frac{\left(\boldsymbol{\Phi}_1 v_{i,t-1-\ell} + \cdots + \boldsymbol{\Phi}_{t-1-\ell} v_{i1}\right)\mathbf{d}_1 + \boldsymbol{\Phi}_{t-\ell}\mathbf{w}_{i0} + \mathbf{q}_i}{t-\ell}\right\}, \qquad \textbf{(A.4)}$$

$$\boldsymbol{\Phi}_j = \boldsymbol{\Pi}^0 + \boldsymbol{\Pi}^1 + \cdots + \boldsymbol{\Pi}^{j-1} = (\mathbf{I}_p - \boldsymbol{\Pi})^{-1}(\mathbf{I}_p - \boldsymbol{\Pi}^j),$$

$$\mathbf{q}_i = \left[\sum_{s=1}^{p-1} y_{i,-s}, \sum_{s=2}^{p-1} y_{i,-s}, \ldots, y_{i,-p+1}, 0\right]',$$

$$\boldsymbol{G}_p = \mathrm{diag}\left[\frac{p-1}{t-\ell+p-1}, \frac{p-2}{t-\ell+p-2}, \ldots, 0\right],$$

and $\boldsymbol{\kappa}_p$, $\mathbf{R}$, *and* $\mathbf{r}_i$ *are defined later.*

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

Proof of Lemma A.

(i) First, note that under the assumption that $\mathrm{E}(\mu_i|\mathbf{y}_i^{t-1})$ coincides with the linear projection, we have

$$\mathrm{E}(\mathbf{x}_{it}^*|\mathbf{y}_i^{t-1}) = c_t \left[\mathbf{I}_p - \frac{\mathbf{\Pi}(\mathbf{I}_p - \mathbf{\Pi}^{T-t})(\mathbf{I}_p - \mathbf{\Pi})^{-1}}{T-t}\right] \left[\mathbf{x}_{it} - \boldsymbol{\iota}_p \frac{\phi\left(\boldsymbol{\iota}_{t+p-1}' \mathbf{V}_{t-1}^{-1} \mathbf{y}_i^{t-1}\right)}{1+\phi\left(\boldsymbol{\iota}_{t+p-1}' \mathbf{V}_{t-1}^{-1} \boldsymbol{\iota}_{t+p-1}\right)}\right],$$

where $\phi = \sigma_\mu^2/\sigma_v^2$, $\mathbf{V}_{t-1} = \sigma_v^{-2}\mathrm{E}\left[\left(\mathbf{y}_i^{t-1} - \mu_i \boldsymbol{\iota}_{t+p-1}\right)\left(\mathbf{y}_i^{t-1} - \mu_i \boldsymbol{\iota}_{t+p-1}\right)'\right]$, $\mu_i = \eta_i/(1 - \boldsymbol{\alpha}'\boldsymbol{\iota}_p)$, and $\sigma_\mu^2 = \mathrm{var}(\mu_i)$. In what follows, we derive formulas of $\boldsymbol{\iota}_{t+p-1}' \mathbf{V}_{t-1}^{-1} \mathbf{y}_i^{t-1}$ and $\boldsymbol{\iota}_{t+p-1}' \mathbf{V}_{t-1}^{-1} \boldsymbol{\iota}_{t+p-1}$.

Following Whittle (1951) and Wise (1955), let us define the $(t+p-1) \times (t+p-1)$ matrix $\mathbf{U}^k$ as follows:

$$\mathbf{U}^k = \begin{bmatrix} \mathbf{O}_{(t+p-k-1)\times k} & \mathbf{I}_{t+p-k-1} \\ \mathbf{O}_{k\times k} & \mathbf{O}_{k\times(t+p-k-1)} \end{bmatrix}, \qquad (k = 1, \ldots, p).$$

Then, $\mathbf{y}_i^{t-1}$ can be written as

$$\mathbf{y}_i^{t-1} = \alpha_1 \mathbf{U}\mathbf{y}_i^{t-1} + \alpha_2 \mathbf{U}^2 \mathbf{y}_i^{t-1} + \cdots + \alpha_p \mathbf{U}^p \mathbf{y}_i^{t-1} + \eta_i \begin{bmatrix} \boldsymbol{\iota}_{t-1} \\ \mathbf{O}_{p\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_i^{t-1} \\ \mathbf{O}_{p\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{O}_{(t-1)\times 1} \\ \mathbf{r}_i \end{bmatrix},$$

where $\mathbf{v}_i^{t-1} = (v_{i,t-1}, \ldots, v_{i,1})'$ and

$$\mathbf{r}_i = \left[y_{i0} - \sum_{k=1}^{p-1} \alpha_k y_{i,-k}, y_{i,-1} - \sum_{k=2}^{p-1} \alpha_{k-1} y_{i,-k}, \ldots, y_{i,-p+2} - \alpha_1 y_{i,-p+1}, y_{i,-p+1}\right]'.$$

Because y_{it} is stationary and its conditional mean given η_i is $\mu_i = \eta_i/(1 - \boldsymbol{\alpha}'\boldsymbol{\iota}_p)$,

$$\begin{aligned}\left(\mathbf{I}_{t+p-1} - \mathbf{\Delta}\right)\widetilde{\mathbf{y}}_i^{t-1} &= \eta_i \begin{bmatrix} \boldsymbol{\iota}_{t-1} \\ \mathbf{O}_{p\times 1} \end{bmatrix} - \mu_i (\mathbf{I}_{t+p-1} - \mathbf{\Delta})\boldsymbol{\iota}_{t+p-1} + \begin{bmatrix} \mathbf{v}_i^{t-1} \\ \mathbf{O}_{p\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{O}_{p\times 1} \\ \mathbf{r}_i \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_i^{t-1} \\ \mathbf{O}_{p\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{O}_{(t-1)\times 1} \\ \mathbf{r}_i - \boldsymbol{\kappa}_p \mu_i \end{bmatrix} \equiv \widetilde{\mathbf{v}}_i^{t-1} + \widetilde{\mathbf{r}}_i,\end{aligned}$$

where $\widetilde{\mathbf{y}}_i^{t-1} = \mathbf{y}_i^{t-1} - \mu_i \boldsymbol{\iota}_{t+p-1}$, $\mathbf{\Delta} = \left(\alpha_1 \mathbf{U} + \alpha_2 \mathbf{U}^2 + \cdots + \alpha_p \mathbf{U}^p\right)$,

$$\left(\mathbf{I}_{t+p-1} - \mathbf{\Delta}\right)\boldsymbol{\iota}_{t+p-1} = \begin{bmatrix} \boldsymbol{\iota}_{t-1}(1 - \boldsymbol{\alpha}'\boldsymbol{\iota}_p) \\ \boldsymbol{\kappa}_p \end{bmatrix}, \quad \text{and} \tag{A.5}$$

$$\boldsymbol{\kappa}_p = \left[1 - \sum_{k=1}^{p-1} \alpha_k, 1 - \sum_{k=1}^{p-2} \alpha_k, \ldots, 1 - \alpha_1, 1\right]'.$$

Then it follows that

$$\mathbf{V}_{t-1}^{-1} = (\mathbf{I}_{t+p-1} - \mathbf{\Delta}')\begin{bmatrix} \mathbf{I}_{t-1} & O \\ O & \mathbf{R}^{-1} \end{bmatrix}(\mathbf{I}_{t+p-1} - \mathbf{\Delta}),$$

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

where $\mathbf{R} = \sigma_v^{-2}\mathrm{E}[(\mathbf{r}_i - \boldsymbol{\kappa}_p\mu_i)(\mathbf{r}_i - \boldsymbol{\kappa}_p\mu_i)']$.

Therefore, using (A.5) we have

$$\boldsymbol{\iota}_{t+p-1}'\mathbf{V}_{t-1}^{-1}\boldsymbol{\iota}_{t+p-1} = (1-\boldsymbol{\alpha}'\boldsymbol{\iota}_p)^2(t-1) + \boldsymbol{\kappa}_p'\mathbf{R}^{-1}\boldsymbol{\kappa}_p,$$

$$\boldsymbol{\iota}_{t+p-1}'\mathbf{V}_{t-1}^{-1}\mathbf{y}_i^{t-1} = (1-\boldsymbol{\alpha}'\boldsymbol{\iota}_p)[\eta_i(t-1) + v_{i,t-1} + \cdots + v_{i,1}] + \boldsymbol{\kappa}_p'\mathbf{R}^{-1}\mathbf{r}_i.$$

Expression (A.1) is easily obtained because $v_{i,t-1}, \ldots, v_{i1}$ in the numerator of $\mathbf{g}_{it}$ are independent random variables whereas the denominator of $\mathbf{g}_{it}$ is $O(t)$.

(ii) Note that $\mathbf{z}_{it(\ell)}$ can be written as follows:

$$\mathbf{z}_{it(\ell)} = c_t\left\{\mathbf{x}_{it} - (\mathbf{I}_p - \mathbf{G}_p)\frac{\mathbf{x}_{i,t-\ell} + \cdots + \mathbf{x}_{i1} + \mathbf{q}_i}{t-\ell}\right\}.$$

Then, using $\mathbf{x}_{it} = \mu_i\boldsymbol{\iota}_p + \mathbf{w}_{i,t-1}$, we get

$$\mathbf{x}_{i,t-\ell} + \cdots + \mathbf{x}_{i1} = (t-\ell)\mu_i\boldsymbol{\iota}_p + (\boldsymbol{\Phi}_1 v_{i,t-1-\ell} + \cdots + \boldsymbol{\Phi}_{t-1-\ell}v_{i1})\mathbf{d}_1 + \boldsymbol{\Phi}_{t-\ell}\mathbf{w}_{i0}.$$

Thus, we obtain (A.2). To prove (A.3), we have to show that $\widetilde{\mathbf{g}}_{it}$ is $O_p(1/\sqrt{t})$. However, because p is fixed and $v_{i,t-\ell-1}, \ldots, v_{i1}$ are independent random variables, the second term in (A.4) is $O_p(1/\sqrt{t})$. Then, the result follows from the fact that the first term is $O(1/t)$ because $\|\mathrm{Var}(\mu_i\mathbf{G}_p\boldsymbol{\iota}_p)\| = \sigma_\mu^2\sum_{j=0}^{p-1}[j/(t-\ell+j)]^2 = O(1/t^2)$. ■

LEMMA B. *Let Assumptions 1–3 hold. Then,* $\|\mathrm{E}(\mathbf{g}_{it}\mathbf{w}_{i,t-1}')\|$ *and* $\|\mathrm{E}(\widetilde{\mathbf{g}}_{it}\mathbf{w}_{i,t-1}')\|$ *are* $O(1/t)$.

Proof of Lemma B. First, note that $\mathrm{E}(\mu_i\mathbf{w}_{i,t-1}) = \mathbf{O}_{p\times 1}$. Next, because p is fixed, we have

$$\|\mathrm{E}[(v_{i,t-1} + \cdots + v_{i,1})\mathbf{w}_{i,t-1}']\| = \sigma_v^2\|\mathbf{d}_1' \times [(\mathbf{I}_p - \boldsymbol{\Pi})^{-1}(\mathbf{I}_p - \boldsymbol{\Pi}^{t-p})]'\| = O(1),$$

$$\left\|\mathrm{E}\left[\boldsymbol{\kappa}_p'\mathbf{R}^{-1}\mathbf{r}_i\mathbf{w}_{i,t-1}'\right]\right\| = O(1),$$

$$\|\mathrm{E}[(\boldsymbol{\Phi}_1 v_{i,t-1-\ell} + \cdots + \boldsymbol{\Phi}_{t-1-\ell}v_{i1})\mathbf{d}_1\mathbf{w}_{i,t-1}']\| = \sigma_v^2\left\|\sum_{j=1}^{t-2}\boldsymbol{\Phi}_j\mathbf{d}_1\mathbf{d}_1'(\boldsymbol{\Pi}')^j\right\| = O(1),$$

$$\|\mathrm{E}(\boldsymbol{\Phi}_{t-\ell}\mathbf{w}_{i,0} + \mathbf{q}_i)\mathbf{w}_{i,t-1}'\| = O(1).$$

Then, the result follows from the fact that the denominators of $\mathbf{g}_{it}$ and $\widetilde{\mathbf{g}}_{it}$ are $O(t)$. ■

Next, we derive the asymptotic properties of the IV estimators. Note that IV estimators $\widehat{\boldsymbol{\alpha}}_{IV}^{OPT}$ and $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$ can be written as

$$\sqrt{NT}(\widehat{\boldsymbol{\alpha}}_{IV} - \boldsymbol{\alpha}) = \widehat{\mathbf{A}}^{-1}\sqrt{NT}\,\widehat{\mathbf{b}},$$

where $\widehat{\mathbf{A}}$ denotes $\widehat{\mathbf{A}}_{IV}^{OPT}$, $\widehat{\mathbf{A}}_{IV}^{RMA\ell}$ and $\widehat{\mathbf{b}}$ denotes $\widehat{\mathbf{b}}_{IV}^{OPT}$, $\widehat{\mathbf{b}}_{IV}^{RMA\ell}$. The asymptotic behavior of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{b}}$ is given in the following lemma.

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

LEMMA C. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity,*

(i) $\widehat{\mathbf{A}}_{IV}^{OPT}, \widehat{\mathbf{A}}_{IV}^{RMA\ell} \xrightarrow{p} \mathrm{E}(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}')$,

(ii) $\widehat{\mathbf{b}}_{IV}^{OPT}, \widehat{\mathbf{b}}_{IV}^{RMA\ell} \xrightarrow{p} 0$.
If we further assume that Assumption 4 holds, then as both N and T tend to infinity,

(iii) $\sqrt{NT}\,\widehat{\mathbf{b}}_{IV}^{OPT}, \sqrt{NT}\,\widehat{\mathbf{b}}_{IV}^{RMA\ell} \xrightarrow{d} \mathcal{N}\left[0, \sigma_v^2 \mathrm{E}\left(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}'\right)\right]$.

Proof of Lemma C. To derive the results, we use the following decomposition:

$$\mathbf{x}_{it}^* = \mathbf{\Psi}_t \mathbf{w}_{i,t-1} - c_t \widetilde{\mathbf{v}}_{itT},$$
$$\mathbf{\Psi}_t = c_t \left(\mathbf{I}_p - \frac{1}{T-t} \mathbf{\Pi} \mathbf{\Phi}_{T-t} \right),$$
$$\widetilde{\mathbf{v}}_{itT} = \frac{(\mathbf{\Phi}_{T-t} v_{it} + \mathbf{\Phi}_2 v_{i,T-2} + \cdots + \mathbf{\Phi}_1 v_{i,T-1})\mathbf{d}_1}{T-t}.$$

(a) First, we consider $\widehat{\mathbf{A}}_{IV}^{OPT}$. Using Lemmas A and B and the preceding decomposition, we have

$$\begin{aligned}
\mathrm{E}\left(\widehat{\mathbf{A}}_{IV}^{OPT}\right) &= \frac{1}{T} \sum_{t=1}^{T-1} \mathrm{E}\left(\mathbf{h}_{it} \mathbf{x}_{it}^{*'}\right) \\
&= \frac{1}{T} \sum_{t=1}^{T-1} \left[\mathbf{I}_p - O\left(\frac{1}{T-t}\right) \right] \left[\mathrm{E}\left(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}'\right) + O\left(\frac{1}{t}\right) \right] \\
&\to \mathrm{E}\left(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}'\right).
\end{aligned}$$

The last convergence comes from $T^{-1}\sum_{t=1}^{T-1} O(1/(T-t)) = O(\log T/T) \to 0$. The variance of $\widehat{\mathbf{A}}_{IV}^{OPT}$ is shown to tend to zero as follows:

$$\left\| \mathrm{var}\left(\mathrm{vec}\left\{ \widehat{\mathbf{A}}_{IV}^{OPT} \right\} \right) \right\| = \frac{1}{NT^2} \left\| \mathrm{var}\left(\sum_{t=1}^{T-1} \mathrm{vec}\left\{ \mathbf{h}_{it} \mathbf{x}_{it}^{*'} \right\} \right) \right\| = O\left(\frac{1}{N}\right) \to 0.$$

For $\widehat{\boldsymbol{\alpha}}_{IV}^{RMA\ell}$, we have

$$\begin{aligned}
\mathrm{E}\left(\widehat{\mathbf{A}}_{IV}^{RMA\ell}\right) &= \frac{1}{T} \sum_{t=1}^{T-1} \left[\mathrm{E}(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}') \left\{ \mathbf{I}_p + O\left(\frac{1}{T-t}\right) \right\} + O\left(\frac{1}{t}\right) \right] \\
&\to \mathrm{E}(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}').
\end{aligned}$$

The variance of $\widehat{\mathbf{A}}_{IV}^{RMA\ell}$ is shown to tend to zero in a similar way to $\widehat{\mathbf{A}}_{IV}^{OPT}$.

(b), (c) Note that, using Lemma A(i), we have

$$\begin{aligned}
\sqrt{NT}\widehat{\mathbf{b}}_{IV}^{OPT} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T-1} c_t \left[\mathbf{I}_p - O\left(\frac{1}{T-t}\right) \right] (\mathbf{w}_{i,t-1} + \mathbf{g}_{it}) v_{it}^* \\
&= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T-1} c_t \mathbf{w}_{i,t-1} v_{it}^* + o_p(1)
\end{aligned}$$

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \mathbf{w}_{i,t-1} v_{it} + o_p(1).$$

Because the generalized Lindeberg–Feller condition for joint asymptotic normality is satisfied under Assumptions 1–4, we can apply the central limit theorem of Phillips and Moon (1999, Thm. 2). (A detailed proof can be found in Hahn and Kuersteiner, 2002, Lem, 5; and Lee, 2007, Lems. 3 and 4.) Thus, we have

$$\sqrt{NT}\,\widehat{\mathbf{b}}_{IV}^{OPT} \xrightarrow{d} \mathcal{N}\left[0, \sigma_v^2 \mathrm{E}\left(\mathbf{w}_{i,t-1}\mathbf{w}_{i,t-1}'\right)\right].$$

The result for $\sqrt{NT}\,\widehat{\mathbf{b}}_{IV}^{RMA\ell}$ is obtained in a similar way. For the proof of (ii), it is straightforward from (iii) to show that $\widehat{\mathbf{b}}_{IV}^{RMA\ell}, \widehat{\mathbf{b}}_{IV}^{OPT} \xrightarrow{p} 0$. ■

Proof of Proposition 1. Using Lemma C, the results are easily obtained. ■

https://doi.org/10.1017/S0266466609090707 Published online by Cambridge University Press