Econometric Theory, **31**, 2015, 647–667.
doi:10.1017/S0266466614000449

THE ASYMPTOTIC PROPERTIES OF THE SYSTEM GMM ESTIMATOR IN DYNAMIC PANEL DATA MODELS WHEN BOTH N AND T ARE LARGE

KAZUHIKO HAYAKAWA
Hiroshima University

In this paper, we derive the asymptotic properties of the system generalized method of moments (GMM) estimator in dynamic panel data models with individual and time effects when both N and T, the dimensions of cross-section and time series, are large. Specifically, we show that the two-step system GMM estimator is consistent when a suboptimal weighting matrix where off-diagonal blocks are set to zero is used. Such consistency results theoretically support the use of the system GMM estimator in large N and T contexts even though it was originally developed for large N and small T panels. Simulation results indicate that the large N and large T asymptotic results approximate the finite sample behavior reasonably well unless persistency of data is strong and/or the variance ratio of individual effects to the disturbances is large.

1. INTRODUCTION

In recent decades, with the growing availability of comprehensive statistical databases, the use of dynamic panel models has steadily increased. The advantages are clear: dynamic panel models not only allow us to consider the dynamics of economic activity but also control for unobservable heterogeneity. To estimate dynamic panel data models, several estimators have been proposed. These include the instrumental variables (IV) estimator (Anderson and Hsiao, 1981), the least squares dummy variable (LSDV) estimator (Nickell, 1981), the first-difference (FD-) GMM estimator (Holtz-Eakin, Newey, and Rosen, 1988; Arellano and Bond, 1991), the level and the FOD-GMM estimator[1] (Arellano and Bover, 1995), the system GMM estimator (Arellano and Bover, 1995; Blundell

This paper is a revision of the second chapter of my Ph.D. thesis submitted to Hitotsubashi University. I am deeply grateful to my advisers Taku Yamamoto and Eiji Kurozumi for their guidance and suggestions. I also acknowledge Guido Kuersteiner (co-editor), two anonymous referees, Katsuto Tanaka, In Choi, Ryo Okui, Hiroaki Chigira, and the participants of the European Meeting of Econometric Society in Milan, a research seminar at HKUST, and the Kansai Econometrics Conference for their helpful comments. Part of this paper was written while the author is visiting University of Cambridge as a JSPS Postdoctoral Fellow for Research Abroad. I acknowledge the financial support from the JSPS Fellowship and the Grant-in-Aid for Scientific Research (KAKENHI 20830056, 22730178) provided by the JSPS. However, all remaining errors are my own. 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.

and Bond, 1998), the limited information maximum likelihood (LIML)-type estimator (Alonso-Borrego and Arellano, 1999), and the maximum likelihood (ML) estimator (Bhargava and Sargan, 1983; Hsiao, Pesaran, and Tahmiscioglu, 2002). Among these estimators, the system GMM estimator is the most widely used in empirical analysis. For example, Blundell and Bond (2000), Bond, Hoeffler, and Temple (2001), Dollar and Kraay (2002), and Beck, Levine, and Loayza (2000) have all used the system GMM estimator. There are two main reasons for the system GMM estimator's popularity. The first is that, as shown by Blundell and Bond (1998), the FD-GMM estimator suffers from the weak instrument problem when the persistency of data is strong, while the system GMM estimator does not. The second is that in comparison to the FD- and level GMM estimators, the system GMM estimator is more efficient. These two useful properties render the system GMM estimator popular in empirical studies. However, with regard to the first advantage, Bun and Windmeijer (2010) show that the system GMM estimator suffers from the weak instrument problem if the variance ratio of individual effects to the disturbance is large.[2] This implies that the weak instrument problem in the GMM estimation of dynamic panel data models is not addressed. Associated with this issue, several studies propose alternative estimators or investigate the issue in detail. In line with the former approach, Hahn, Hausman, and Kuersteiner (2007), Han and Phillips (2010), and Han, Phillips, and Sul (2014) propose variants of IV/GMM estimators, while Hayakawa and Pesaran (2012) extend the ML approach of Hsiao et al. (2002) to allow for cross-sectional heteroskedasticity. In line with the latter approach, Hahn et al. (2007), Kruiniger (2009), Hayakawa (2009a), and Hayakawa and Nagata (2013) investigate the weak instrument problem in detail using the near unit root asymptotics. Hahn et al. (2007) and Kruiniger (2009) demonstrate that the FD- and system GMM estimators are inconsistent in general and follow nonstandard distribution under near unit root asymptotics. Hayakawa (2009a) and Hayakawa and Nagata (2013) show that the weak instrument problem of the FD-GMM estimator is not always a problem if initial conditions do not follow the stationary distribution. In fact, they demonstrate that if the process is not mean stationary, there are cases where the instruments can be strong even if persistency is strong.

While the above studies assume small T and large N, where T and N denote the time-series and the cross-sectional size, respectively, motivated by the increasing availability of panel data in which T and N are both large, many studies consider using large N and large T asymptotics.[3] Early contributions include Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003). They derived the large N and large T asymptotic properties of typical estimators for dynamic panel data models such as the LSDV, the FOD-GMM, the LIML-type, the FD-GMM, and the random effect ML estimators. More recently, Hayakawa (2009b) proposes an IV estimator for the AR(p) model that is efficient when both N and T are large.

Several studies also try to incorporate cross-sectional dependence in dynamic panel data models beyond the standard dynamic panel data model assuming cross-sectional independence. These include Phillips and Sul (2003, 2007), Sarafidis and Robertson (2009), Sarafidis, Yamagata and Robertson (2009), Robertson, Sarafidis, and Symons (2010), Sarafidis and Yamagata (2010), Moon and Weidner (2010), and Hayakawa (2012). These utilize a factor structure to model the cross-sectional dependence. In contrast, Elhorst (2004, 2010), Su and Yang (2007), and Yu, Jong, and Lee (2008) employ a spatial approach to model the cross-section dependence.

The present paper contributes to the existing literature by complementing the paper by Alvarez and Arellano (2003). As discussed above, while some potential problems exist, and alternative estimation procedures have been suggested for models possibly with a cross-sectional dependence, the most popular estimator in empirical studies is still the system GMM estimator. However, Alvarez and Arellano (2003) studied only the FOD- and FD-GMM estimators. Therefore, it is of interest to investigate the asymptotic properties of the system GMM estimator under large N and T asymptotics. Moreover, although Alvarez and Arellano (2003) consider the panel AR(1) model with individual effects only, we consider a panel AR(1) model with both individual and time effects since time effects are often included in empirical studies.[4]

To derive the properties of the system GMM estimator, we need to consider the level GMM estimator since the system GMM estimator is obtained from the moment conditions of the FD-(FOD-) and the level GMM estimators. Specifically, we consider two types of models in levels. The first is the original level model (see (1)), and the second is a model transformed by the generalized least squares (GLS) principle (see (10)). The reason we consider these two models is as follows. In cross-section models, it is known that as the sample size and number of instruments get larger, the two-stage least squares (2SLS) estimator tends to the limit of the OLS estimator (Kunitomo, 1980; Morimune, 1983; Bekker, 1994). Alvarez and Arellano (2003) demonstrate that a similar result also holds for the GMM estimation of dynamic panel data models. They show that the GMM estimator for models in forward orthogonal deviations is consistent under large N and T asymptotics since the OLS estimator for that model is the LSDV estimator, which is consistent when N and T are large.[5] They also show that the GMM estimator for models in first differences with a nonoptimal weighting matrix is inconsistent since the OLS estimator for that model is inconsistent when N and T are large. In this paper, we show that the same result holds true for the level GMM estimator; that is, the GMM estimator for models transformed by the GLS principle is consistent since the OLS estimator for that model is the random effect GLS estimator that is consistent when N and T are large. Moreover, we show that the GMM estimator for original level models is inconsistent since the OLS estimator for that model is inconsistent when N and T are large. We then combine this result with that of the FOD-GMM estimator to derive the asymptotic properties of the system GMM estimator. Specifically, we demonstrate that the system GMM estimator with a suboptimal weighting matrix, wherein off-diagonal blocks are set to zero, is consistent under large N and T asymptotics. Further, we derive the asymptotic distributions for the consistent GMM estimators. As a result, we find that the FOD-, level, and system GMM estimators using all the available instruments have the same asymptotic variance, although their biases are different. Simulation results show that the large N and large T asymptotic results approximate the finite sample behavior reasonably well when persistency of data is not strong and/or the variance ratio of individual effects to the disturbances is not large. When persistency is strong, inferences based on the system GMM estimators are very inaccurate and deteriorate as the variance ratio of individual effects to the disturbances increases.

The remainder of this paper is organized as follows. In the next section, we provide the model and estimators. The main results are reported in Section 3. In Section 4, the simulation results are provided to assess the theoretical implications obtained in Section 3. Finally, Section 5 concludes the paper.

With regard to the notation, let us denote $T_1 = T-1$ and $\underset{N,T\to\infty}{\longrightarrow}$ a convergence when N and T jointly tend to infinity. $\underset{N,T\to\infty}{\overset{p}{\longrightarrow}}$ and $\underset{N,T\to\infty}{\overset{d}{\longrightarrow}}$ denote convergence in probability and in distribution, respectively, when both N and T jointly tend to infinity. $\underset{N(T)\to\infty}{\overset{p}{\longrightarrow}}$ denotes convergence in probability when N tends to infinity and T is allowed to be fixed or tend to infinity.

The proofs of theorems and all simulation results are provided in supplementary material to this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

2. THE MODEL AND ESTIMATORS

We consider an AR(1) panel data model with both individual and time effects:

$$\begin{aligned} Y_{it} &= \alpha Y_{i,t-1} + \gamma_i + \lambda_t + \varepsilon_{it} \qquad (i = 1,\ldots,N+1;\ t = 1,\ldots,T) \\ &= \alpha X_{it} + \gamma_i + \lambda_t + \varepsilon_{it}, \end{aligned}$$

where $X_{it} = Y_{i,t-1}$. α is the parameter of interest with $|\alpha| < 1$, and γ_i and λ_t represent unobserved individual and time effects, respectively. To remove the time effects λ_t, first, we take an orthogonal deviation over cross-sectional units[6]:

$$\begin{aligned} y_{it} &= \alpha x_{it} + \eta_i + v_{it} \qquad (i = 1,\ldots,N;\ t = 1,\ldots,T) \\ &= \alpha x_{it} + u_{it}, \end{aligned} \tag{1}$$

where $y_{it} = a_i[Y_{it} - (Y_{i+1,t} + \cdots + Y_{N+1,t})/(N+1-i)]$, $x_{it} = a_i[X_{it} - (X_{i+1,t} + \cdots + X_{N+1,t})/(N+1-i)]$, $\eta_i = a_i[\gamma_i - (\gamma_{i+1} + \cdots + \gamma_{N+1})/(N+1-i)]$, $v_{it} = a_i[\varepsilon_{it} - (\varepsilon_{i+1,t} + \cdots + \varepsilon_{N+1,t})/(N+1-i)]$, and $u_{it} = \eta_i + v_{it}$, with $a_i^2 = (N+1-i)/(N+2-i)$.

We impose the following assumptions.

Assumption 1. $\{\varepsilon_{it}\}$ $(i = 1,\ldots,N+1;\ t = 1,\ldots,T)$ are $i.i.d.$ across time and individuals and independent of γ_i and Y_{i0} with $E(\varepsilon_{it}) = 0$, $var(\varepsilon_{it}) = \sigma_\varepsilon^2$, and finite moments up to the eighth order.

Assumption 2. $\{\gamma_i\}$ $(i = 1, \ldots, N)$ are $i.i.d.$ across individuals with $E(\gamma_i) = 0$, $var(\gamma_i) = \sigma_\gamma^2$, and finite moments up to the eighth order.

Assumption 3. The initial observations satisfy

$$Y_{i0} = \frac{\gamma_i}{1-\alpha} + \lambda_0 + e_{i0}, \qquad (i = 1, \ldots N+1),$$

where $e_{i0} = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,-j}$ and are independent of γ_i.

Note that no assumption is required for time effects $\{\lambda_t\}_{t=0}^{T}$ since we remove λ_t from both the model and instruments, implying that the time effects are completely removed from the estimators. Furthermore, since $E(v_{it}^2) = \sigma_v^2 = \sigma_\varepsilon^2$ and $E(v_{it} v_{jt}) = 0$, $(i \neq j)$ under Assumption 1, the errors are cross-sectionally uncorrelated.[7] Thus, model (1), where time effects are removed by orthogonal deviations over cross-sectional units, is essentially the same as the one considered in Alvarez and Arellano (2003) where time effects are not included. Additionally, note that $var(\eta_i) = \sigma_\eta^2 = \sigma_\gamma^2$.

Assumptions 1 and 2 are stronger than those in Alvarez and Arellano (2003) wherein finite fourth-order moments are assumed. Finite eighth-order moments are used to derive the asymptotic properties of the two-step level and system GMM estimators. For the FOD-GMM estimator, finite fourth-order moments suffice. Assumption 3, which implies mean-stationarity, is required to ensure the consistency of the level and the system GMM estimators under large N and fixed T asymptotics.

Under Assumption 3, Y_{it} can be written as

$$Y_{it} = \frac{\gamma_i}{1-\alpha} + \lambda_0 + \sum_{j=0}^{t-1} \alpha^j \lambda_{t-j} + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{i,t-j}.$$

Hence, we have

$$y_{it} = \mu_i + w_{it}, \qquad \textbf{(2)}$$

where $\mu_i = \eta_i/(1-\alpha)$ and $w_{it} = \sum_{j=0}^{\infty} \alpha^j v_{i,t-j}$.

We now define the GMM estimators. Those considered in this paper are the FOD-, level, and system GMM estimators.

GMM estimator for models in forward orthogonal deviations. Let us consider a model in forward orthogonal deviations:

$$y_{it}^* = \alpha x_{it}^* + v_{it}^* \qquad (i = 1, \ldots, N;\ t = 1, \ldots, T-1), \qquad \textbf{(3)}$$

where y_{it}^*, x_{it}^*, and v_{it}^* are defined as $y_{it}^* = c_t[y_{it} - (y_{i,t+1} + \cdots + y_{iT})/(T-t)]$, $x_{it}^* = c_t[x_{it} - (x_{i,t+1} + \cdots + x_{iT})/(T-t)]$, and $v_{it}^* = c_t[v_{it} - (v_{i,t+1} + \cdots + v_{iT})/(T-t)]$, respectively, with $c_t^2 = (T-t)/(T-t+1)$. Note that since

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

$E(v_{it}^{*2}) = \sigma_v^2$ and $E(v_{it}^* v_{is}^*) = 0,\ (t \neq s)$, v_{it}^* is homoskedastic and serially uncorrelated. Moreover, note that a pooled OLS estimator of (3) is the LSDV estimator.

The moment condition for model (3) is given by[8]

$$E[y_{is} v_{it}^*] = 0, \qquad (s = 0, \ldots, t-1;\ t = 1, \ldots, T-1). \tag{4}$$

Note that instead of $Y_{i0}, \ldots, Y_{i,T-2}$, which are often used in empirical studies, we use transformed instruments $y_{i0}, \ldots, y_{i,T-2}$ without time effects. This simplifies the theoretical derivations.[9] The moment conditions (4) can be written in a matrix form as follows:

$$E[\underline{\mathbf{Z}}_i^{l\prime} \underline{\mathbf{v}}_i^*] = \mathbf{0},$$

where $\underline{\mathbf{v}}_i^* = (v_{i1}^*, \ldots, v_{i,T-1}^*)'$ and $\underline{\mathbf{Z}}_i^l = \text{diag}(\mathbf{z}_{i1}^{l\prime}, \mathbf{z}_{i2}^{l\prime}, \cdots, \mathbf{z}_{i,T-1}^{l\prime})$ with $\mathbf{z}_{it}^l = (y_{i0}, \ldots, y_{i,t-1})'$.

Then, the FOD-GMM estimator is defined as

$$\begin{aligned}
\widehat{\alpha}_{F2} &= \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i^{*\prime} \underline{\mathbf{Z}}_i^l\right)\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{l\prime} \underline{\mathbf{Z}}_i^l\right)^{-1} \left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{l\prime} \underline{\mathbf{x}}_i^*\right)\right]^{-1} \\
&\quad \times \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i^{*\prime} \underline{\mathbf{Z}}_i^l\right)\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{l\prime} \underline{\mathbf{Z}}_i^l\right)^{-1} \left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{l\prime} \underline{\mathbf{y}}_i^*\right)\right] \\
&= \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{x}_t^*\right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{y}_t^*\right),
\end{aligned}$$

where $\underline{\mathbf{x}}_i^* = (x_{i1}^*, \ldots, x_{i,T-1}^*)'$, $\underline{\mathbf{y}}_i^* = (y_{i1}^*, \ldots, y_{i,T-1}^*)'$, $\mathbf{x}_t^* = (x_{1t}^*, \ldots, x_{Nt}^*)'$, $\mathbf{y}_t^* = (y_{1t}^*, \ldots, y_{Nt}^*)'$, $\mathbf{z}_{it}^l = (y_{i0}, \ldots, y_{i,t-2})'$, $\mathbf{Z}_t^l = (\mathbf{z}_{1t}^l, \ldots, \mathbf{z}_{Nt}^l)'$, and $\mathbf{M}_t^l = \mathbf{Z}_t^l (\mathbf{Z}_t^{l\prime} \mathbf{Z}_t^l)^{-1} \mathbf{Z}_t^{l\prime}$.

Note that $\widehat{\alpha}_{F2}$ is optimal under Assumption 1 when T is fixed and N is large, and computable when $T \leq N$. In addition, note that the number of moment conditions in (4) is $m_{l2} = T(T-1)/2 = O(T^2)$.

In the following discussion of the level and system GMM estimators, we first assume that σ_η^2 and σ_v^2 are known and then proceed to the case where estimated variances are used.

GMM estimators for models in levels. Let us consider a model in levels (1). Under Assumptions 1 and 3, we have the following moment conditions:

$$E[\Delta y_{is} u_{it}] = 0, \qquad (s = 1, \ldots, t-1;\ t = 2, \ldots, T),$$

where all available instruments are used, or

$$E[\Delta y_{i,t-1} u_{it}] = 0, \qquad (t = 2, \ldots, T),$$

where only a subset of instruments are used. These can be written in a matrix form:

$$E[\underline{\mathbf{Z}}_i^{d2\prime}\underline{\mathbf{u}}_i] = \mathbf{0}, \tag{5}$$

$$E[\underline{\mathbf{Z}}_i^{d1\prime}\underline{\mathbf{u}}_i] = \mathbf{0}, \tag{6}$$

where $\underline{\mathbf{u}}_i = (u_{i2}, \ldots, u_{iT})'$, $\underline{\mathbf{Z}}_i^{d2} = \text{diag}\left(\mathbf{z}_{i2}^{d2\prime}, \mathbf{z}_{i3}^{d2\prime}, \cdots, \mathbf{z}_{iT}^{d2\prime}\right)$ with $\mathbf{z}_{it}^{d2} = (\Delta y_{i1}, \ldots, \Delta y_{i,t-1})'$, and $\underline{\mathbf{Z}}_i^{d1} = \text{diag}(\Delta y_{i1}, \Delta y_{i2}, \cdots, \Delta y_{i,T-1})$. Note that the numbers of moment conditions in (5) and (6) are $m_{d2} = T(T-1)/2 = O(T^2)$ and $m_{d1} = T - 1 = O(T)$, respectively. The infeasible level GMM estimator based on the moment condition (5), which is optimal under Assumptions 1–3 when N is large and T is fixed, is given by[10]

$$\widetilde{\alpha}_{L2}^* = \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i' \underline{\mathbf{Z}}_i^{d2}\right)\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d2\prime}\dot{\mathbf{\Omega}}_{T_1}\underline{\mathbf{Z}}_i^{d2}\right)^{-1}\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d2\prime}\underline{\mathbf{x}}_i\right)\right]^{-1} \times \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i' \underline{\mathbf{Z}}_i^{d2}\right)\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d2\prime}\dot{\mathbf{\Omega}}_{T_1}\underline{\mathbf{Z}}_i^{d2}\right)^{-1}\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d2\prime}\underline{\mathbf{y}}_i\right)\right], \tag{7}$$

where $\underline{\mathbf{x}}_i = (x_{i2}, \ldots, x_{iT})'$, $\underline{\mathbf{y}}_i = (y_{i2}, \ldots, y_{iT})'$, and $\dot{\mathbf{\Omega}}_{T_1} = \sigma_v^2 \mathbf{I}_{T_1} + \sigma_\eta^2 \boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}'$, with $\boldsymbol{\iota}_{T_1}$ being a $T_1 \times 1$ vector of ones. Similarly, we can define $\widetilde{\alpha}_{L1}^*$ by replacing $\underline{\mathbf{Z}}_i^{d2}$ with $\underline{\mathbf{Z}}_i^{d1}$. Although $\widetilde{\alpha}_{L2}^*$ is easily computed when T is small, this is not the case for a large T since we need to invert an $m_{d2} \times m_{d2}$ optimal weighting matrix that becomes very large when T is large. For example, $m_{l2} = 105$ when $T = 15$, $m_{l2} = 300$ when $T = 25$, and $m_{l2} = 1225$ when $T = 50$.[11] Furthermore, if $m_{l2} > N$, we cannot compute $\widetilde{\alpha}_{L2}^*$, because of the singularity of $\sum_{i=1}^N \mathbf{Z}_i^{d2\prime}\dot{\mathbf{\Omega}}_{T_1}\mathbf{Z}_i^{d2}$. Although we may use a generalized inverse in such a case, Satchachai and Schmidt (2008) show that this does not solve the problem.

A simple remedy for this problem is to use nonoptimal weighting matrices $\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d2\prime}\underline{\mathbf{Z}}_i^{d2}\right)^{-1}$ and $\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{d1\prime}\underline{\mathbf{Z}}_i^{d1}\right)^{-1}$, which yield

$$\widehat{\alpha}_{L2}^{non} = \left(\sum_{t=2}^T \mathbf{x}_t' \mathbf{M}_t^{d2} \mathbf{x}_t\right)^{-1}\left(\sum_{t=2}^T \mathbf{x}_t' \mathbf{M}_t^{d2} \mathbf{y}_t\right), \tag{8}$$

$$\widehat{\alpha}_{L1}^{non} = \left(\sum_{t=2}^T \mathbf{x}_t' \mathbf{M}_t^{d1} \mathbf{x}_t\right)^{-1}\left(\sum_{t=2}^T \mathbf{x}_t' \mathbf{M}_t^{d1} \mathbf{y}_t\right), \tag{9}$$

where $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})'$, $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\mathbf{z}_{it}^{d2} = (\Delta y_{i1}, \ldots, \Delta y_{i,t-1})'$, $\mathbf{Z}_t^{d2} = (\mathbf{z}_{1t}^{d2}, \ldots, \mathbf{z}_{Nt}^{d2})'$, $\mathbf{M}_t^{d2} = \mathbf{Z}_t^{d2}(\mathbf{Z}_t^{d2\prime}\mathbf{Z}_t^{d2})^{-1}\mathbf{Z}_t^{d2\prime}$, $z_{it}^{d1} = \Delta y_{i,t-1}$, $\mathbf{Z}_t^{d1} = (z_{1t}^{d1}, \ldots, z_{N,t}^{d1})'$, and $\mathbf{M}_t^{d1} = \mathbf{Z}_t^{d1}(\mathbf{Z}_t^{d1\prime}\mathbf{Z}_t^{d1})^{-1}\mathbf{Z}_t^{d1\prime}$. Compared with (7), (8) and (9) are computationally attractive since they do not require inversions of large dimensional matrices and are computable when $T \le N$. Note that these nonoptimal level GMM estimators are discussed in Bun and Kiviet (2006).

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

In the cross-sectional model, it is known that as both the sample size and the number of instruments become larger jointly, the 2SLS estimator tends to the limit of OLS estimator. As noted in Section 1, Alvarez and Arellano (2003) demonstrate that the same applies to the dynamic panel data models in forward orthogonal deviations or in first-differences. This paper shows that the same also applies to models in levels. Since the OLS estimator of the model $y_{it} = \alpha x_{it} + u_{it}$ is inconsistent when N and T are large, we can expect that $\widehat{\alpha}_{L2}^{non}$ would be inconsistent. This is demonstrated in Theorem 3. However, since the random effect GLS estimator of the model $y_{it} = \alpha x_{it} + u_{it}$ is consistent when N and T are large, we can expect that the corresponding GMM estimator would be consistent when N and T are large.

For this purpose, we introduce a GLS transformation that simplifies the optimal weighting matrix, which is similar to the forward orthogonal deviation. Since $\underline{\mathbf{u}}_i$ is serially correlated owing to the individual effects and has a covariance matrix $\sigma_v^{-2} E(\underline{\mathbf{u}}_i \underline{\mathbf{u}}_i') = \mathbf{\Omega}_{T_1} = \mathbf{I}_{T_1} + r \boldsymbol{\imath}_{T_1} \boldsymbol{\imath}_{T_1}'$ where $r = \sigma_\eta^2/\sigma_v^2$, we apply the GLS principle. Specifically, let $\mathbf{\Omega}_{T_1}^{-1/2}$ be the upper triangular Cholesky factorization of $\mathbf{\Omega}_{T_1}^{-1}$, that is, $\mathbf{\Omega}_{T_1}^{-1} = \mathbf{\Omega}_{T_1}^{-1/2\prime} \mathbf{\Omega}_{T_1}^{-1/2}$, and define $\underline{\mathbf{y}}_i^+ = \mathbf{\Omega}_{T_1}^{-1/2} \underline{\mathbf{y}}_i = (y_{i2}^+, \ldots, y_{iT}^+)'$, $\underline{\mathbf{x}}_i^+ = \mathbf{\Omega}_{T_1}^{-1/2} \underline{\mathbf{x}}_i = (x_{i2}^+, \ldots, x_{iT}^+)'$, and $\underline{\mathbf{u}}_i^+ = \mathbf{\Omega}_{T_1}^{-1/2} \underline{\mathbf{u}}_i = (u_{i2}^+, \ldots, u_{iT}^+)'$. Consequently, the model to be estimated is given as follows:

$$\underline{\mathbf{y}}_i^+ = \alpha \underline{\mathbf{x}}_i^+ + \underline{\mathbf{u}}_i^+ \qquad (i = 1, \ldots, N). \tag{10}$$

Note that a pooled OLS estimator of (10) is the random effect GLS estimator. Since the inverse of the optimal weighting matrix under large N and fixed T asymptotics is given by

$$E\left(\underline{\mathbf{Z}}_i^{d\prime} \underline{\mathbf{u}}_i^+ \underline{\mathbf{u}}_i^{+\prime} \underline{\mathbf{Z}}_i^d\right) = E\left(\underline{\mathbf{Z}}_i^{d\prime} \mathbf{\Omega}_{T_1}^{-1/2} E\left[\underline{\mathbf{u}}_i \underline{\mathbf{u}}_i' | \underline{\mathbf{Z}}_i^d\right] \mathbf{\Omega}_{T_1}^{-1/2\prime} \underline{\mathbf{Z}}_i^d\right) = \sigma_v^2 E\left(\underline{\mathbf{Z}}_i^{d\prime} \underline{\mathbf{Z}}_i^d\right),$$

computationally convenient, infeasible, optimal level GMM estimators are given by

$$\widehat{\alpha}_{L2}^* = \left(\sum_{t=2}^{T} \mathbf{x}_t^{+\prime} \mathbf{M}_t^{d2} \mathbf{x}_t^+\right)^{-1} \left(\sum_{t=2}^{T} \mathbf{x}_t^{+\prime} \mathbf{M}_t^{d2} \mathbf{y}_t^+\right), \tag{11}$$

$$\widehat{\alpha}_{L1}^* = \left(\sum_{t=2}^{T} \mathbf{x}_t^{+\prime} \mathbf{M}_t^{d1} \mathbf{x}_t^+\right)^{-1} \left(\sum_{t=2}^{T} \mathbf{x}_t^{+\prime} \mathbf{M}_t^{d1} \mathbf{y}_t^+\right), \tag{12}$$

where $\mathbf{x}_t^+ = (x_{1t}^+, \ldots, x_{Nt}^+)'$ and $\mathbf{y}_t^+ = (y_{1t}^+, \ldots, y_{Nt}^+)'$. Hayakawa (2010) shows that $\widetilde{\alpha}_{L2}^*$ and $\widehat{\alpha}_{L2}^*$ are numerically equivalent and that the order of magnitude of the finite sample bias of $\widehat{\alpha}_{L2}^*$ is smaller than that of $\widehat{\alpha}_{L2}^{non}$.[12]

In practice, these two estimators are easily computed by using a built-in procedure of Cholesky factorization contained in, say, GAUSS or MATLAB. However, to derive the asymptotic properties of the level GMM estimators (11) and (12), we need the explicit form of $\mathbf{\Omega}_{T_1}^{-1/2}$.[13] The explicit expression of $\mathbf{\Omega}_{T_1}^{-1/2}$ is given in the supplementary material (see also Hayakawa, 2010).

Thus far, we assumed that r is known. However, in empirical studies, r is usually unknown, and hence, we have to estimate it. Here, we assume that a consistent estimate of r, $\widehat{r} = \widehat{\sigma}_\eta^2/\widehat{\sigma}_\upsilon^2$, is available where $\widehat{\sigma}_\upsilon^2 = \frac{1}{NT_1}\sum_{i=1}^N \sum_{t=1}^{T_1}(y_{it}^* - \widehat{\alpha}x_{it}^*)^2$ and $\widehat{\sigma}_\eta^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^{T}(y_{it} - \widehat{\alpha}x_{it})^2 - \widehat{\sigma}_\upsilon^2$ with $\widehat{\alpha}$ being a consistent estimate of α. Then, using $\widehat{\mathbf{\Omega}}_{T_1}^{-1/2}$ with r replaced by $\widehat{r}$, the two-step level GMM estimators are defined as

$$\widehat{\alpha}_{L2} = \left(\sum_{t=2}^T \widehat{\mathbf{x}}_t^{+\prime}\mathbf{M}_t^{d2}\widehat{\mathbf{x}}_t^+\right)^{-1}\left(\sum_{t=2}^T \widehat{\mathbf{x}}_t^{+\prime}\mathbf{M}_t^{d2}\widehat{\mathbf{y}}_t^+\right),$$

$$\widehat{\alpha}_{L1} = \left(\sum_{t=2}^T \widehat{\mathbf{x}}_t^{+\prime}\mathbf{M}_t^{d1}\widehat{\mathbf{x}}_t^+\right)^{-1}\left(\sum_{t=2}^T \widehat{\mathbf{x}}_t^{+\prime}\mathbf{M}_t^{d1}\widehat{\mathbf{y}}_t^+\right),$$

where $\widehat{\mathbf{x}}_t^+ = (\widehat{x}_{1t}^+, \ldots, \widehat{x}_{Nt}^+)'$ and $\widehat{\mathbf{y}}_t^+ = (\widehat{y}_{1t}^+, \ldots, \widehat{y}_{Nt}^+)'$ are obtained from $\widehat{\underline{\mathbf{x}}}_i = \widehat{\mathbf{\Omega}}_{T_1}^{-1/2}\underline{\mathbf{x}}_i$ and $\widehat{\underline{\mathbf{y}}}_i = \widehat{\mathbf{\Omega}}_{T_1}^{-1/2}\underline{\mathbf{y}}_i$ for $i = 1, \ldots, N$.

System GMM estimators. Finally, we consider the system GMM estimator. The model of the system GMM estimators can be expressed as

$$\begin{bmatrix} \underline{\mathbf{y}}_i^* \\ \underline{\mathbf{y}}_i^+ \end{bmatrix} = \alpha \begin{bmatrix} \underline{\mathbf{x}}_i^* \\ \underline{\mathbf{x}}_i^+ \end{bmatrix} + \begin{bmatrix} \underline{\mathbf{v}}_i^* \\ \underline{\mathbf{u}}_i^+ \end{bmatrix},$$

which we rewrite as

$$\underline{\mathbf{y}}_i^\dagger = \alpha \underline{\mathbf{x}}_i^\dagger + \underline{\mathbf{u}}_i^\dagger. \tag{13}$$

The moment condition for system (13), which is proposed by Blundell and Bond (1998), is

$$E(\underline{\mathbf{Z}}_i^{s1\prime}\underline{\mathbf{u}}_i^\dagger) = \mathbf{0}, \tag{14}$$

where $\underline{\mathbf{Z}}_i^{s1} = \text{diag}(\underline{\mathbf{Z}}_i^l, \underline{\mathbf{Z}}_i^{d1})$. The optimal system GMM estimator is defined as[14]

$$\widehat{\alpha}_{SYS}^* = \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i^{\dagger\prime}\underline{\mathbf{Z}}_i^{s1}\right)\widehat{\Psi}^{sys}\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{s1\prime}\underline{\mathbf{x}}_i^\dagger\right)\right]^{-1} \times \left[\left(\sum_{i=1}^N \underline{\mathbf{x}}_i^{\dagger\prime}\underline{\mathbf{Z}}_i^{s1}\right)\widehat{\Psi}^{sys}\left(\sum_{i=1}^N \underline{\mathbf{Z}}_i^{s1\prime}\underline{\mathbf{y}}_i^{\dagger\prime}\right)\right],$$

where $\widehat{\Psi}^{sys}$ is a consistent estimate of

$$\Psi^{sys} = \begin{bmatrix} E(\underline{\mathbf{Z}}_i^{l\prime}\underline{\mathbf{v}}_i^*\underline{\mathbf{v}}_i^{*\prime}\underline{\mathbf{Z}}_i^l) & E(\underline{\mathbf{Z}}_i^{l\prime}\underline{\mathbf{v}}_i^*\underline{\mathbf{u}}_i^{+\prime}\underline{\mathbf{Z}}_i^{d1}) \\ E(\underline{\mathbf{Z}}_i^{d1\prime}\underline{\mathbf{u}}_i^+\underline{\mathbf{v}}_i^{*\prime}\underline{\mathbf{Z}}_i^l) & E(\underline{\mathbf{Z}}_i^{d1\prime}\underline{\mathbf{u}}_i^+\underline{\mathbf{u}}_i^{+\prime}\underline{\mathbf{Z}}_i^{d1}) \end{bmatrix}^{-1}.$$

Since $\widehat{\alpha}_{SYS}^{*}$ is not computationally attractive for the same reason as $\widetilde{\alpha}_{L2}^{*}$, we may use the GLS principle in (13) to obtain a computationally useful estimator as we did in obtaining $\widehat{\alpha}_{L2}$ and $\widehat{\alpha}_{L1}$. However, we cannot use the GLS principle in the system, because the covariance matrix of $\underline{\mathbf{u}}_i^{\dagger}$ is singular. This can be shown as follows. Let $\mathbf{F}_T$ be a transformation matrix that induces forward orthogonal deviations and $\underline{\dot{\mathbf{u}}}_i = (u_{i1}, \ldots, u_{iT})'$. Then, using $\underline{\mathbf{v}}_i^{*} = \mathbf{F}_T \underline{\dot{\mathbf{u}}}_i$ and $\underline{\mathbf{u}}_i^{+} = \mathbf{\Omega}_{T_1}^{-1/2} \underline{\mathbf{u}}_i = \mathbf{\Omega}_{T_1}^{-1/2} \mathbf{J}_T \underline{\dot{\mathbf{u}}}_i$ with $\mathbf{J}_T = (\mathbf{0},\ \mathbf{I}_{T_1})$, we have $\underline{\mathbf{u}}_i^{\dagger} = \mathbf{H}\dot{\mathbf{u}}_i$ with $\mathbf{H} = \left[\mathbf{F}_T'\ (\mathbf{\Omega}_{T_1}^{-1/2}\mathbf{J}_T)'\right]'$ being a $2T_1 \times T$ matrix. Therefore, it follows that $E(\underline{\mathbf{u}}_i^{\dagger}\underline{\mathbf{u}}_i^{\dagger\prime}) = \mathbf{H}\mathbf{\Omega}_T\mathbf{H}'$. However, this covariance matrix is rank deficient since $\text{rank}(\mathbf{H}\mathbf{\Omega}_T\mathbf{H}') \leq \min\left(\text{rank}(\mathbf{H}), \text{rank}(\mathbf{\Omega}_T\mathbf{H}')\right) \leq T < 2T_1$, where $2T_1$ is the number of columns of $\mathbf{H}\mathbf{\Omega}_T\mathbf{H}'$. Thus, when $T > 2$, $\underline{\mathbf{u}}_i^{\dagger}$ has a singular covariance matrix.

Therefore, to achieve computational attractiveness and simplify the theoretical derivation, we consider the system GMM estimator using a suboptimal weighting matrix where off-diagonal blocks of the optimal weighting matrix are set to zeros, that is, $\widehat{\mathbf{\Psi}}^{s1} = \text{diag}\left(\sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{l\prime}\underline{\mathbf{Z}}_i^{l}, \sum_{i=1}^{N} \underline{\mathbf{Z}}_i^{d1\prime}\underline{\mathbf{Z}}_i^{d1}\right)$. Then, the suboptimal system GMM estimator is given by

$$\widehat{\alpha}_{S1}^{*} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime}\mathbf{M}_t^{l}\mathbf{x}_t^{*} + \sum_{t=2}^{T} \mathbf{x}_t^{+\prime}\mathbf{M}_t^{d1}\mathbf{x}_t^{+}\right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime}\mathbf{M}_t^{l}\mathbf{y}_t^{*} + \sum_{t=2}^{T} \mathbf{x}_t^{+\prime}\mathbf{M}_t^{d1}\mathbf{y}_t^{+}\right). \quad \textbf{(15)}$$

Note that setting the off-diagonal blocks of the weighting matrix to zero leads to computational attractiveness at the cost of efficiency.

Next, we consider an alternative system GMM estimator that uses all available moment conditions given by

$$E(\underline{\mathbf{Z}}_i^{s2\prime}\underline{\mathbf{u}}_i^{\dagger}) = \mathbf{0}, \quad \textbf{(16)}$$

where $\underline{\mathbf{Z}}_i^{s2} = \text{diag}(\underline{\mathbf{Z}}_i^{l}, \underline{\mathbf{Z}}_i^{d2})$. Note that the moment conditions included in (16) but not in (14) are redundant since they can be obtained by a linear combination of (14).

The suboptimal system GMM estimator obtained from the moment condition (16) is defined as

$$\widehat{\alpha}_{S2}^{*} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime}\mathbf{M}_t^{l}\mathbf{x}_t^{*} + \sum_{t=2}^{T} \mathbf{x}_t^{+\prime}\mathbf{M}_t^{d2}\mathbf{x}_t^{+}\right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime}\mathbf{M}_t^{l}\mathbf{y}_t^{*} + \sum_{t=2}^{T} \mathbf{x}_t^{+\prime}\mathbf{M}_t^{d2}\mathbf{y}_t^{+}\right). \quad \textbf{(17)}$$

Intuitively, using redundant moment conditions does not improve efficiency, since they do not provide new information. In fact, Breusch, Qian, Schmidt, and Wyhowski (1999) formally demonstrate that this is the case for the GMM with the optimal weighting matrix. However, the setup here is different in that the optimal weighting matrix is not used. In the supplementary material, in a general framework, we show that, depending on the structure of redundancy and the weighting matrix associated with the redundant moment conditions, using redundant moment conditions may improve or worsen efficiency compared with the GMM without redundant moment conditions. However, for the specific panel AR(1) model considered in this paper, we demonstrate that using redundant moment conditions as in (16) improves efficiency. However, it should be noted that whether such an efficiency gain holds in a more general model with exogenous variables is inconclusive and may require extensive simulation studies.

Thus far, we assumed that the variances of the individual effects and disturbances are known. The feasible system GMM estimators using estimated variances are defined as

$$\widehat{\alpha}_{S2} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{x}_t^* + \sum_{t=2}^{T} \widehat{\mathbf{x}}_t^{+\prime} \mathbf{M}_t^{d2} \widehat{\mathbf{x}}_t^+ \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{y}_t^* + \sum_{t=2}^{T} \widehat{\mathbf{x}}_t^{+\prime} \mathbf{M}_t^{d2} \widehat{\mathbf{y}}_t^+ \right),$$

$$\widehat{\alpha}_{S1} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{x}_t^* + \sum_{t=2}^{T} \widehat{\mathbf{x}}_t^{+\prime} \mathbf{M}_t^{d1} \widehat{\mathbf{x}}_t^+ \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{y}_t^* + \sum_{t=2}^{T} \widehat{\mathbf{x}}_t^{+\prime} \mathbf{M}_t^{d1} \widehat{\mathbf{y}}_t^+ \right).$$

Moreover, for comparison purposes, we consider the following system GMM estimators:

$$\widehat{\alpha}_{S2}^{non} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{x}_t^* + \sum_{t=2}^{T} \mathbf{x}_t^{\prime} \mathbf{M}_t^{d2} \mathbf{x}_t \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{y}_t^* + \sum_{t=2}^{T} \mathbf{x}_t^{\prime} \mathbf{M}_t^{d2} \mathbf{y}_t \right),$$

$$\widehat{\alpha}_{S1}^{non} = \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{x}_t^* + \sum_{t=2}^{T} \mathbf{x}_t^{\prime} \mathbf{M}_t^{d1} \mathbf{x}_t \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{x}_t^{*\prime} \mathbf{M}_t^l \mathbf{y}_t^* + \sum_{t=2}^{T} \mathbf{x}_t^{\prime} \mathbf{M}_t^{d1} \mathbf{y}_t \right).$$

$\widehat{\alpha}_{S2}^{non}$ and $\widehat{\alpha}_{S1}^{non}$ are the system GMM estimators that use the optimal weighting matrix only for models in the forward orthogonal deviation, while the nonoptimal weighting matrix is used for equation in levels. Since these estimators can be used to obtain the first-step estimates in the two-step system GMM estimators, it is interesting to investigate the properties of these estimators.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

In this section, we derive the asymptotic properties of the GMM estimators defined in the previous section when both N and T are large. The following Theorem 1 is the result for the FOD and level GMM estimators. Note that (a) and (d) are derived in Alvarez and Arellano (2003).

THEOREM 1. *Let Assumptions 1–3 hold. Then, as N and T tend to infinity, we have*

(a) $\widehat{\alpha}_{F2} \xrightarrow[N,T\to\infty]{p} \alpha, \qquad \text{if} \quad (\log T)^2/N \to 0,$

(b) $\widehat{\alpha}_{L2} \xrightarrow[N,T\to\infty]{p} \alpha, \qquad \text{if} \quad (\log T)^2/N \to 0,$

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

(c) $\widehat{\alpha}_{L1} \underset{N,T\to\infty}{\overset{p}{\longrightarrow}} \alpha.$

(d) $\sqrt{NT_1}\left[\widehat{\alpha}_{F2} - \alpha + \frac{(1+\alpha)}{N}\right] \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(0, 1-\alpha^2\right)$ if $(\log T)^2/N \to 0$ *and* $T/N \to c$ $(0 \leq c < \infty)$,

(e) $\sqrt{NT_1}\left[\widehat{\alpha}_{L2} - \alpha + \frac{(1+\alpha)}{N}\left(\frac{\alpha}{r+1}\right)\right] \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(0, 1-\alpha^2\right)$ if $(\log T)^2/N \to 0$ *and* $T/N \to c$ $(0 \leq c < \infty)$,

(f) $\sqrt{NT_1}\left(\widehat{\alpha}_{L1} - \alpha\right) \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}(0, 2(1+\alpha)).$

Remark 1. We find that all the GMM estimators are consistent. Further, we find that the asymptotic variances of $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$ are identical although asymptotic biases are different.

Remark 2. As shown in Hahn and Kuersteiner (2002), if we assume that v_{it} is normally distributed, $\mathcal{N}(0, 1-\alpha^2)$ is the minimal asymptotic distribution. Thus, $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$ are asymptotically efficient if v_{it} is normally distributed.

Remark 3. The intuition behind the consistency of $\widehat{\alpha}_{L2}$ is that as T gets larger, the GMM estimator becomes close to the OLS estimator of (10), that is, the random effect GLS estimator, which is consistent when T is large. Alternatively, we may say that the "endogeneity bias" goes to zero as T becomes larger (Alvarez and Arellano, 2003, p. 1129). This is similar to the relationship between $\widehat{\alpha}_{F2}$ and the LSDV estimator discussed in Alvarez and Arellano (2003).

Remark 4. Note that $\widehat{\alpha}_{L1}$ has no asymptotic bias but has a larger variance than $\widehat{\alpha}_{L2}$ because $\widehat{\alpha}_{L1}$ uses a smaller number of instruments than $\widehat{\alpha}_{L2}$.

The following Theorem 2 is the result for the system GMM estimators.

THEOREM 2. *Let Assumptions 1–3 hold. Then, as N and T tend to infinity, provided* $(\log T)^2/N \to 0$,

(a) $\widehat{\alpha}_{S2} \underset{N,T\to\infty}{\overset{p}{\longrightarrow}} \alpha,$

(b) $\widehat{\alpha}_{S1} \underset{N,T\to\infty}{\overset{p}{\longrightarrow}} \alpha.$

If we further assume that $T/N \to c$ $(0 \leq c < \infty)$, *we have*

(c) $\sqrt{NT_1}\left[\widehat{\alpha}_{S2} - \alpha + \frac{1}{N}\frac{(1+\alpha)(r+1+\alpha)}{2(r+1)}\right] \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(0, 1-\alpha^2\right),$

(d) $\sqrt{NT_1}\left[\widehat{\alpha}_{S1} - \alpha + \frac{1}{N}\frac{2(1+\alpha)}{(3-\alpha)}\right] \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(0, (1-\alpha^2)\frac{2(5-3\alpha)}{(3-\alpha)^2}\right).$

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

Remark 5. The intuition behind the consistency of $\widehat{\alpha}_{S2}$ is similar to those for $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$ as given in Alvarez and Arellano (2003) and Remark 3. Namely, since the OLS estimator of system (13), which is a combination of the large-T consistent LSDV and random effect GLS estimators, is consistent when $N, T \to \infty$, it is not surprising that $\widehat{\alpha}_{S2}$ is also consistent when N and T are large.

Remark 6. If v_{it} is normally distributed, $\widehat{\alpha}_{S2}$ is asymptotically efficient. Intuitively, this is because $\widehat{\alpha}_{S2}$ is a weighted sum of two efficient GMM estimators, $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$. Moreover, since $\widehat{\alpha}_{S2}$ has the same asymptotic variance as $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$, it implies that the efficiency gain by exploiting stationary initial conditions vanishes as T increases. This result seems natural since as T increases, the effects of initial conditions become weak. Further, note that, as is the case for $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}_{L2}$, $\widehat{\alpha}_{S2}$ and $\widehat{\alpha}_{S1}$ have a bias of order $O(1/N)$.

Remark 7. In the supplementary material, by comparing the theoretical values of asymptotic variances under large N and fixed T, we show that using redundant moment conditions improves efficiency, i.e., $var(\widehat{\alpha}^*_{S2}) < var(\widehat{\alpha}^*_{S1})$. Theorems 2($c$) and ($d$) show that a similar result is also obtained when N and T are large. Since $1 < 2(5-3\alpha)/(3-\alpha)^2 < 9/8$ for $|\alpha| < 1$, the variance of $\widehat{\alpha}_{S1}$ is at most approximately 1.125 times that of $\widehat{\alpha}_{S2}$.

Next, we derive the asymptotic properties of the GMM estimators with nonoptimal weighting matrices.

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

$$\text{(a)}\quad \widehat{\alpha}^{non}_{L2} \xrightarrow[N,T\to\infty]{p} \alpha + \frac{\frac{cr}{2}\left(\frac{1}{1-\alpha}\right)}{\frac{cr}{2}\left(\frac{1}{1-\alpha}\right)^2 + \frac{1}{1-\alpha^2}} \qquad \text{if } T/N \to c\ (0 \le c < \infty),$$

$$\text{(b)}\quad \widehat{\alpha}^{non}_{L1} \xrightarrow[N,T\to\infty]{p} \alpha,$$

$$\text{(c)}\quad \sqrt{NT_1}\left[\widehat{\alpha}^{non}_{L1} - \alpha - \frac{2r(1+\alpha)}{N(1-\alpha)}\right] \xrightarrow[N,T\to\infty]{d} \mathcal{N}(0, 2(1+\alpha)) \qquad \text{if } T/N \to c\ (0 \le c < \infty).$$

Remark 8. From (a), we find that the level GMM estimator $\widehat{\alpha}^{non}_{L2}$ is inconsistent when both N and T are large. This result is analogous to the case of the first-difference-GMM estimator as shown by Alvarez and Arellano (2003). However, from (b), we find that if we reduce the number of moment conditions from $O(T^2)$ to $O(T)$, the level GMM estimator is consistent. This result can be conjectured from Bun and Kiviet (2006).

Remark 9. Surprisingly, the asymptotic variances of $\widehat{\alpha}_{L1}$ and $\widehat{\alpha}^{non}_{L1}$ are identical. However, there is a significant difference in their biases. The bias of $\widehat{\alpha}^{non}_{L1}$ increases as r grows.

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

The last result is the asymptotic properties of the system GMM estimators using a nonoptimal weighting matrix for equation in levels.

THEOREM 4. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity,*

(a) $\widehat{\alpha}_{S2}^{non} \xrightarrow[N,T\to\infty]{p} \alpha + \dfrac{cr}{\frac{4}{1+\alpha} + \frac{cr}{1-\alpha}} \qquad$ if $T/N \to c$ $(0 \le c < \infty)$,

(b) $\widehat{\alpha}_{S1}^{non} \xrightarrow[N,T\to\infty]{p} \alpha$,

(c) $$\sqrt{NT_1}\left[\widehat{\alpha}_{S1}^{non} - \alpha + \frac{2(1+\alpha)}{(3-\alpha)}\frac{1-r}{N}\right] \xrightarrow[N,T\to\infty]{d} \mathcal{N}\left(0, (1-\alpha^2)\frac{2(5-3\alpha)}{(3-\alpha)^2}\right)$$
if $T/N \to c$ $(0 \le c < \infty)$.

Remark 10. Since only $\widehat{\alpha}_{S1}^{non}$ is consistent, it can be used to obtain the first-step estimates to compute the two-step GMM estimators.

Remark 11. From (c), we find that the asymptotic bias in $\widehat{\alpha}_{S1}^{non}$ disappears when $r = 1$, implying that the bias of $\widehat{\alpha}_{S1}^{non}$ can be small when $r = 1$. This result is consistent with the literature; see, for example, Blundell and Bond (1998), Bun and Kiviet (2006), and Hayakawa (2007). Moreover, note that the bias becomes larger as r becomes larger.

4. MONTE CARLO EXPERIMENTS

In this section, we confirm the theoretical implications through Monte Carlo experiments. We consider the following AR(1) model:

$$Y_{it} = \alpha Y_{i,t-1} + \gamma_i + \lambda_t + \varepsilon_{it} \qquad (i = 1, \ldots, N; t = 2, \ldots, T), \tag{18}$$

where $\gamma_i \sim iid\mathcal{N}(0, \sigma_\gamma^2)$, $\lambda_t \sim iid\mathcal{N}(0, \sigma_\lambda^2)$, and $\varepsilon_{it} \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$. Initial values Y_{i1} are generated from the stationary distribution. We set $\alpha = 0.3, 0.6, 0.9$, $N = 200$, $T = 5, 10, 20$, $\sigma_\gamma^2 = 0.2, 1, 5$, $\sigma_\lambda^2 = 1$, and $\sigma_\varepsilon^2 = 1$. The number of replications is 1000 for all cases. For the FOD and the level GMM estimators, we compute $\widehat{\alpha}_{F2}$, $\widehat{\alpha}_{L2}^*$, $\widehat{\alpha}_{L2}$, $\widehat{\alpha}_{L2}^{non}$, $\widehat{\alpha}_{L1}^*$, $\widehat{\alpha}_{L1}$, and $\widehat{\alpha}_{L1}^{non}$. We use $\widehat{\alpha}_{L1}^{non}$ to obtain a consistent estimate of r. Similarly, for the system GMM estimator, we compute $\widehat{\alpha}_{S2}^*$, $\widehat{\alpha}_{S2}$, $\widehat{\alpha}_{S1}^*$, $\widehat{\alpha}_{S1}$, $\widehat{\alpha}_{S2}^{non}$, $\widehat{\alpha}_{S1}^{non}$, $\widehat{\alpha}_{SYS}^*$, and $\widehat{\alpha}_{SYS}$. We use $\widehat{\alpha}_{S1}^{non}$ to obtain a consistent estimate of r. For each estimator listed above, we compute the mean, the standard deviation, the root mean squared error (RMSE), and the empirical sizes of the Wald test at the 5% significant level on the basis of fixed T and large N asymptotics, and large T and large N asymptotics. In Table 1, we report the means, standard deviations, and the empirical sizes based on fixed T and large N asymptotics. Complete results are provided in Tables A1 to A6 in the supplementary material.

We summarize the simulation results. From Tables 1 and A1, we find that $\widehat{\alpha}_{F2}$ is negatively biased and its degree is substantial when T is small, and that

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

α and r are large. Further, it is observed that as T becomes larger, the bias of $\widehat{\alpha}_{F2}$ decreases. This result conforms with the theoretical result that $\widehat{\alpha}_{F2}$ is consistent when T is large. With regard to the level GMM estimators, we find that infeasible optimal level GMM estimators, $\widehat{\alpha}^*_{L2}$, and $\widehat{\alpha}^*_{L1}$, perform quite well for most values of T, α, and r. Before we consider the two-step level GMM estimators,

TABLE 1. Simulation results

			$\alpha=0.3$			$\alpha=0.6$			$\alpha=0.9$		
Estimator	T	N	$r=0.2$	$r=1$	$r=5$	$r=0.2$	$r=1$	$r=5$	$r=0.2$	$r=1$	$r=5$
						Mean					
$F2$	5	200	0.290	0.283	0.278	0.578	0.562	0.534	0.699	0.490	0.399
$S2$	5	200	0.296	0.299	0.312	0.591	0.601	0.644	0.889	0.933	0.973
$S2^{non}$	5	200	0.298	0.309	0.359	0.595	0.615	0.696	0.902	0.952	0.986
$S1$	5	200	0.295	0.294	0.299	0.590	0.595	0.616	0.878	0.914	0.956
$S1^{non}$	5	200	0.297	0.305	0.342	0.595	0.613	0.677	0.898	0.946	0.984
SYS	5	200	0.300	0.303	0.310	0.595	0.603	0.626	0.882	0.914	0.957
$F2$	10	200	0.290	0.289	0.288	0.583	0.575	0.569	0.794	0.724	0.695
$S2$	10	200	0.296	0.297	0.301	0.595	0.598	0.617	0.899	0.942	0.983
$S2^{non}$	10	200	0.300	0.321	0.411	0.603	0.636	0.749	0.916	0.960	0.989
$S1$	10	200	0.294	0.294	0.295	0.590	0.587	0.588	0.870	0.898	0.959
$S1^{non}$	10	200	0.295	0.303	0.338	0.594	0.610	0.676	0.896	0.943	0.982
SYS	10	200	0.298	0.299	0.299	0.598	0.597	0.596	0.880	0.899	0.958
$F2$	20	200	0.292	0.291	0.290	0.587	0.584	0.583	0.848	0.831	0.826
$S2$	20	200	0.296	0.297	0.299	0.596	0.596	0.603	0.896	0.927	0.979
$S2^{non}$	20	200	0.305	0.338	0.465	0.609	0.657	0.786	0.923	0.965	0.991
$S1$	20	200	0.295	0.294	0.294	0.591	0.588	0.590	0.866	0.872	0.932
$S1^{non}$	20	200	0.296	0.301	0.328	0.594	0.604	0.652	0.890	0.929	0.976
SYS	20	200	0.295	0.294	0.295	0.592	0.589	0.590	0.867	0.872	0.932
						Standard deviation					
$F2$	5	200	0.069	0.090	0.110	0.100	0.148	0.186	0.306	0.414	0.439
$S2$	5	200	0.056	0.072	0.098	0.065	0.088	0.125	0.086	0.097	0.078
$S2^{non}$	5	200	0.055	0.073	0.115	0.063	0.082	0.117	0.067	0.061	0.042
$S1$	5	200	0.062	0.079	0.103	0.077	0.106	0.147	0.115	0.131	0.122
$S1^{non}$	5	200	0.060	0.072	0.102	0.072	0.087	0.113	0.082	0.072	0.049
SYS	5	200	0.056	0.064	0.073	0.065	0.079	0.111	0.093	0.116	0.107
$F2$	10	200	0.033	0.038	0.042	0.041	0.049	0.055	0.086	0.117	0.125
$S2$	10	200	0.030	0.035	0.042	0.034	0.041	0.056	0.040	0.040	0.025
$S2^{non}$	10	200	0.030	0.038	0.068	0.033	0.043	0.059	0.027	0.022	0.014
$S1$	10	200	0.034	0.039	0.044	0.040	0.048	0.057	0.066	0.082	0.062
$S1^{non}$	10	200	0.033	0.037	0.046	0.038	0.041	0.054	0.043	0.035	0.021
SYS	10	200	0.032	0.033	0.036	0.034	0.037	0.047	0.051	0.072	0.060

Table continues on overleaf

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

TABLE 1. *continued*

$F2$	20	200	0.020	0.021	0.021	0.021	0.023	0.023	0.030	0.036	0.038
$S2$	20	200	0.019	0.020	0.021	0.019	0.022	0.024	0.023	0.028	0.017
$S2^{non}$	20	200	0.019	0.025	0.051	0.018	0.027	0.039	0.015	0.012	0.006
$S1$	20	200	0.021	0.022	0.022	0.022	0.024	0.024	0.031	0.040	0.048
$S1^{non}$	20	200	0.020	0.021	0.025	0.021	0.023	0.028	0.023	0.021	0.013
SYS	20	200	0.021	0.022	0.022	0.022	0.024	0.024	0.031	0.040	0.048
Empirical size (standard errors obtained under large N and fixed T asymptotics)											
$F2$	5	200	0.052	0.053	0.055	0.067	0.070	0.077	0.151	0.217	0.222
$S2$	5	200	0.060	0.065	0.092	0.072	0.101	0.199	0.087	0.326	0.715
$S2^{non}$	5	200	0.053	0.067	0.148	0.060	0.092	0.296	0.068	0.341	0.749
$S1$	5	200	0.058	0.063	0.067	0.068	0.089	0.130	0.090	0.215	0.566
$S1^{non}$	5	200	0.058	0.058	0.111	0.059	0.068	0.227	0.047	0.210	0.655
SYS	5	200	0.078	0.093	0.132	0.097	0.140	0.298	0.199	0.434	0.774
$F2$	10	200	0.056	0.062	0.071	0.081	0.081	0.094	0.258	0.407	0.460
$S2$	10	200	0.061	0.060	0.071	0.075	0.070	0.148	0.161	0.579	0.935
$S2^{non}$	10	200	0.051	0.098	0.453	0.069	0.180	0.704	0.179	0.776	0.985
$S1$	10	200	0.061	0.062	0.063	0.066	0.057	0.075	0.145	0.289	0.681
$S1^{non}$	10	200	0.054	0.057	0.174	0.059	0.059	0.373	0.053	0.355	0.918
SYS	10	200	0.159	0.173	0.245	0.191	0.214	0.402	0.417	0.733	0.942
$F2$	20	200	0.071	0.068	0.069	0.096	0.114	0.115	0.428	0.545	0.562
$S2$	20	200	0.068	0.060	0.062	0.072	0.080	0.092	0.192	0.539	0.974
$S2^{non}$	20	200	0.070	0.374	0.951	0.082	0.627	0.992	0.450	0.985	1.000
$S1$	20	200	0.068	0.064	0.061	0.078	0.088	0.071	0.243	0.230	0.509
$S1^{non}$	20	200	0.060	0.054	0.244	0.060	0.057	0.543	0.068	0.352	0.988
SYS	20	200	0.748	0.768	0.796	0.801	0.828	0.864	0.926	0.944	0.985

Note: $F2$, $S2$, $S1$, and SYS denote $\widehat{\alpha}_{F2}$, $\widehat{\alpha}_{S2}$, $\widehat{\alpha}_{S1}$, and $\widehat{\alpha}_{SYS}$, respectively. $S2^{non}$ and $S1^{non}$ denote $\widehat{\alpha}_{S2}^{non}$ and $\widehat{\alpha}_{S1}^{non}$, respectively. $r = \sigma_\eta^2/\sigma_v^2$.

$\widehat{\alpha}_{L2}$ and $\widehat{\alpha}_{L1}$, we consider the nonoptimal level GMM estimators, $\widehat{\alpha}_{L2}^{non}$ and $\widehat{\alpha}_{L1}^{non}$, since the latter, $\widehat{\alpha}_{L1}^{non}$, is used to compute the two-step GMM estimators. On examining the results, we observe that the biases of $\widehat{\alpha}_{L2}^{non}$ and $\widehat{\alpha}_{L1}^{non}$ increase as the variance ratio r becomes large. This result has already been obtained in the literature. Further, it coincides with the theoretical implication of Theorem 3 that the bias becomes larger as r grows. The large bias of $\widehat{\alpha}_{L1}^{non}$ results in a poor estimate of r, which is used to compute $\widehat{\alpha}_{L2}$ and $\widehat{\alpha}_{L1}$. Consequently, although two-step level GMM estimators are less biased than the (inconsistent) nonoptimal level GMM estimators, they are more biased than the infeasible optimal level GMM estimators for most values of T, α, and r.

Next, we consider the system GMM estimators. We find that when $T = 5$ and $\alpha = 0.9$, although values of $\widehat{\alpha}_{F2}$ are heavily biased on account of the well-known weak instrument problem, the system GMM estimators are not, unless r is large. We also observe that the biases of the infeasible system GMM estimators,

$\widehat{\alpha}^*_{S2}$ and $\widehat{\alpha}^*_{S1}$, are small except for certain cases of $\alpha = 0.9$. In the case of $\widehat{\alpha}^*_{S1}$, a large bias results from the fact that $\widehat{\alpha}^*_{S1}$ is a linear combination of $\widehat{\alpha}_{F2}$ and $\widehat{\alpha}^*_{L1}$ whose biases are both negative for a large r. For instance, see the values of $\widehat{\alpha}_{F2}$, $\widehat{\alpha}^*_{L2}$, and $\widehat{\alpha}^*_{S1}$ in the case of $T = 20$. Prior to considering the two-step system GMM estimators, we consider $\widehat{\alpha}^{non}_{S2}$ and $\widehat{\alpha}^{non}_{S1}$, where the latter is used to compute $\widehat{\alpha}_{S2}$ and $\widehat{\alpha}_{S1}$. On examining the results, we find that the performance of these estimators is heavily affected by the variance ratio r. This is because these system GMM estimators are linear combinations of the FOD-GMM estimator and nonoptimal level GMM estimator whose bias heavily depends upon r.[15] With regard to the two-step system GMM estimators, we find that although $\widehat{\alpha}_{S1}$ and $\widehat{\alpha}_{S2}$ are very similar when $\alpha = 0.3, 0.6$, $\widehat{\alpha}_{S1}$ is less biased than $\widehat{\alpha}_{S2}$ when $\alpha = 0.9$.

With regard to dispersion, we find from the Tables 1 and A2 that when $T = 5$, $\widehat{\alpha}_{SYS}$ is more efficient than other estimators. However, as T becomes larger, the difference in efficiency between $\widehat{\alpha}_{F2}$, $\widehat{\alpha}_{L2}$, $\widehat{\alpha}_{S2}$, and $\widehat{\alpha}_{SYS}$ becomes very small. This is consistent with the theoretical result that $\widehat{\alpha}_{F2}$, $\widehat{\alpha}_{L2}$, and $\widehat{\alpha}_{S2}$ have the same asymptotic variance. Moreover, in the supplementary material, we showed that $\widehat{\alpha}_{S2}$, which uses redundant moment conditions, is more efficient than $\widehat{\alpha}_{S1}$. From the simulation results, we find that similar results are also observed when $T = 5, 10, 20$.

In terms of RMSE, which is reported in Table A3 in the supplementary material, although $\widehat{\alpha}_{SYS}$ performs best when $T = 5$, the results of $\widehat{\alpha}_{S2}$, $\widehat{\alpha}_{S1}$, and $\widehat{\alpha}_{SYS}$ are comparable when $T = 10, 20$ and α is not large. Further, note that when $T = 5$, the degree of improvement of the system GMM estimators over $\widehat{\alpha}_{F2}$ becomes larger as α grows.

With regard to the inference, we find from Tables 1, A4, and A5 that inference based on large N and fixed T asymptotics is more accurate than that based on large N and large T asymptotics even when T is large. From Tables 1 and A4, we find that the size distortion of $\widehat{\alpha}_{SYS}$ becomes substantial as T and/or r grow larger. The unreported simulation result indicates that this is owing to the underestimated standard error.[16] With regard to $\widehat{\alpha}_{S2}$ and $\widehat{\alpha}_{S1}$, although their sizes are close to the nominal level when $\alpha = 0.3, 0.6$, the size distortion becomes substantial when $\alpha = 0.9$. This is because of the bias in estimates and standard errors.

Finally, to assess the accuracy of large N and large T approximations, we tabulate theoretical values of estimators in Table A6 for the case of $T = 20$.[17] For simplicity, we exclude two-step estimators. In Table A6, we provide theoretical values of each estimator. Note that all the consistent estimates are in the form of $\widehat{\alpha} \approx \alpha + \mu/q$ where μ are μ_{F2}, $\bar{\mu}_{L2}$, $\bar{\mu}_{L1}$, and μ^{non}_{L1}, which are given in Lemmas A4, A7, (C.33), and (C.34) in the supplementary material, and q is the probability limit of $\frac{1}{NT}\sum_t \mathbf{x}^{*\prime}_t \mathbf{M}^l_t \mathbf{x}^*_t$, $\frac{1}{NT}\sum_t \mathbf{x}^{+\prime}_t \mathbf{M}^d_t \mathbf{x}^+_t$, and $\frac{1}{NT}\sum_t \mathbf{x}'_t \mathbf{M}^d_t \mathbf{x}_t$. For inconsistent estimators, theoretical values are based on their probability limits. In Table A6(a), approximated μ and theoretical asymptotic value q are used. Note that theoretical values in Table A6(a) are identical to those obtained from asymptotic distributions and probability limits given in Theorems 1–3.

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

On comparing Tables A1 and A6(a), we observe that although simulated values are close to the theoretical ones when $\alpha = 0.3, 0.6$, it is not the case when $\alpha = 0.9$ and/or $r = 5$. To ascertain the source of the poor approximation, we see Table A6(b)–(d). In Table A6(b), the exact μ and theoretical asymptotic values q are used. However, the results are not so different from Table A6(a), implying that q does not approximate $\frac{1}{NT}\sum_t \mathbf{x}_t^{*\prime}\mathbf{M}_t^l\mathbf{x}_t^*$, $\frac{1}{NT}\sum_t \mathbf{x}_t^{+\prime}\mathbf{M}_t^d\mathbf{x}_t^+$, and $\frac{1}{NT}\sum_t \mathbf{x}_t^{\prime}\mathbf{M}_t^d\mathbf{x}_t$ well. Hence, in Table A6(c) and (d), we use a simulated q on the basis of 1000 replications. On comparing Table A6(c) that uses an approximate value of μ with Table A6(d) that uses an exact value of μ, it is observed that Table A6(d) provides more accurate values. Thus, poor approximation of the estimates arises from a poor approximation of both μ and q. Further, it is also observed that estimates of two-step estimators are not close to the theoretical values. This is because a two-step procedure causes additional finite sample bias in $\widehat{\alpha}_{L1}$.

On summarizing the simulation results, we present the following implications. In terms of bias, the large N and large T asymptotic results can explain the finite sample behavior of estimators reasonably well unless r is large. When r is large, the large N and large T asymptotic distributions are poor approximations to finite sample distributions. In terms of inference, using the standard errors under large N and fixed T asymptotics is preferable to using the standard errors under large N and large T asymptotics even when T is as large as $T = 20$. However, although the sizes are generally close to the nominal level when $\alpha = 0.3$ and 0.6, it is not the case when $\alpha = 0.9$. Hence, despite the fact that the system GMM estimator is known to address the weak instrument problem associated with α being close to one, it does not work well in terms of inference even when r is not large.

5. CONCLUSION

In this paper, we considered the asymptotic properties of system GMM estimators when both N and T are large. We first showed that the two-step level GMM estimator with an optimal weighting matrix is consistent under large N and T asymptotics, while this is not the case with the nonoptimal weighting matrix. Next, using this result, we derived the asymptotic properties of the two-step system GMM estimators. Consequently, we found that the system GMM estimator using the suboptimal weighting matrix is still consistent even when T is large. We also found that using redundant moment conditions could improve efficiency in both small and large T cases. This result implies that the system GMM estimator originally developed for large N and small T panel data is also usable for large N and large T panel data. The simulation studies revealed that the large N and large T asymptotic results approximate the finite sample behavior reasonably well if persistency of data is not strong and/or the variance ratio of individual effects to disturbances is not large. When persistency is strong, inferences are very inaccurate and deteriorate as the variance ratio of individual effects to disturbances gets larger.

Finally, we indicate certain possibilities for future research. Although we considered a panel AR(1) model, it is important to consider a more general model that contains additional regressors and investigate whether results similar to those of this paper are obtained. In particular, investigating whether using redundant moment conditions improves efficiency would be an interesting issue.

NOTES

1. The FOD-GMM estimator refers to the GMM estimator where individual effects in the model are removed by forward orthogonal deviation (FOD) transformation and instruments in levels are used in the estimation. In Alvarez and Arellano (2003), the FOD-GMM estimator is simply termed the GMM estimator.
2. Hayakawa (2007) demonstrates that the finite sample bias of the system GMM estimator becomes large when the variance ratio of individual effects to the disturbances is large.
3. Phillips and Moon (1999) develop a general asymptotic theory where both N and T tend to infinity.
4. In the literature, Hahn and Moon (2006) and Hsiao and Tahmiscioglu (2008) consider dynamic panel models with both individual and time effects.
5. Alvarez and Arellano (2003) provide an insightful interpretation that consistency follows because the "endogeneity bias" goes to zero as T gets larger.
6. This transformation is suggested by a referee.
7. If we remove time effects by taking a deviation from a cross-sectional average over all units, a cross-section dependence would be induced. In such a case, some corrections may be required for the results of this paper.
8. Although additional moment conditions suggested by Ahn and Schmidt (1995, 1997) may be used, we do not exploit them to simplify the derivation.
9. Unreported Monte Carlo simulation results show that behaviors of the GMM estimators using $Y_{i0},\dots,Y_{i,T-2}$ and the ones using $y_{i0},\dots,y_{i,T-2}$ are very similar.
10. A superscript $*$ indicates that the estimator is infeasible.
11. Note that if additional regressors are included in the model, further moment conditions are available.
12. Although Hayakawa (2010) derives the order of magnitude of finite sample bias, consistency and asymptotic distribution are not derived. Hence, in this paper, we derive consistency and asymptotic distribution for the AR(1) case.
13. Although an alternative expression of $\mathbf{\Omega}_{T_1}^{-1/2}$ is derived by Wansbeek and Kapteyn (1982, 1983), it cannot be used in this context, because it is not upper triangular.
14. Note that since we use a parameter-dependent transformation for equation in levels, the system GMM estimator is slightly different from the original one suggested by Blundell and Bond (1998). Additionally, let $\widehat{\alpha}_{SYS}$ denote a feasible version of $\widehat{\alpha}^*_{SYS}$.
15. Note that the same result is found in Bun and Kiviet (2006) that considered $\widehat{\alpha}_{S1}^{non}$.
16. See also Bond and Windmeijer (2005).
17. To save space, we provide the results with $T = 20$ only. Other results are available on request.

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.

Alonso-Borrego, C. & M. Arellano (1999) Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business and Economic Statistics* 17, 36–49.

https://doi.org/10.1017/S0266466614000449 Published online by Cambridge University Press

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

Arellano, M. (2003) *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.

Beck, T., R. Levine, & N. Loayza (2000) Finance and the sources of growth. *Journal of Financial Economics* 58, 261–300.

Bekker, P.A. (1994) Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.

Bhargava, A. & J.D. Sargan (1983) Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51, 1635–1659.

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

Blundell, R. & S. Bond (2000) GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19, 321–340.

Bond, S., A. Hoeffler, & J. Temple (2001) GMM Estimation of Empirical Growth Models. Working paper no. 2001-W21, University of Oxford, Nuffield College, Economics Group.

Bond, S. & F. Windmeijer (2005) Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews* 24, 1–37.

Breusch, T.S., H. Qian, P. Schmidt, & D. Wyhowski (1999) Redundancy of moment conditions. *Journal of Econometrics* 91, 89–111.

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* 132, 409–444.

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

Dollar, D. & A. Kraay (2002) Growth is good for the poor. *Journal of Economic Growth* 7, 195–225.

Elhorst, J.P. (2004) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37, 85–106.

Elhorst, J.P. (2010) Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40, 272–282.

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.

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

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

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

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

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

Hayakawa, K. (2009b) A simple efficient instrumental variable estimator in panel AR(p) models when both N and T are large. *Econometric Theory* 25, 873–890.

Hayakawa, K. (2010) The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models: Some additional results. *Journal of Econometrics* 159, 202–208.

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

Hayakawa, K. & S. Nagata (2013) On the Behavior of the GMM Estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions. Mimeo.

Hayakawa, K. & M.H. Pesaran (2012) Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models. Cambridge Working Papers in Economics No. 1224.

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

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

Hsiao, C. & A.K. Tahmiscioglu (2008) Estimation of dynamic panel data models with both individual and time specific effects. *Journal of Statistical Planning and Inference* 138, 2698–2721.

Kruiniger, H. (2009) GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory* 25, 1348–1391.

Kunitomo, N. (1980) Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association* 75, 693–700.

Moon, H.R. & M. Weidner (2010) Dynamic Linear Panel Regression Models with Interactive Fixed Effects. Mimeo.

Morimune, K. (1983) Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica* 51, 821–841.

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

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1111.

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

Phillips, P.C.B. & D. Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 127, 162–188.

Robertson, D., V. Sarafidis, & J. Symons (2010) IV Estimation of Panels with Factor Residuals. Mimeo.

Sarafidis, V. & D. Robertson (2009) On the impact of error cross-sectional dependence in short dynamic panel estimation. *Econometrics Journal* 12, 62–81.

Sarafidis, V. & T. Yamagata (2010) Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors under Cross- Section Dependence. Mimeo.

Sarafidis, V., T. Yamagata, & D. Robertson (2009) A test of cross section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics* 148, 149–161.

Satchachai, P. & P. Schmidt (2008) GMM with more moment conditions than observations. *Economics Letters* 99, 252–255.

Su, L. & Z. Yang (2007) QML Estimation of Dynamic Panel Data Models with Spatial Errors. Mimeo.

Wansbeek, T. & A. Kapteyn (1982) A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics* A11, 2105–2112.

Wansbeek, T. & A. Kapteyn (1983) A note on spectral decomposition and maximum likelihood estimation in models with balanced data. *Statistics and Probability Letters* 1, 213–215.

Yu, J., R. De Jong, & L.F. Lee (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118–134.