Discussion Paper Series

No.129

The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both *N* and *T* are Large

Kazuhiko Hayakawa

January 2006

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences

A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T are Large

Kazuhiko Hayakawa*†

Graduate School of Economics

Hitotsubashi University

January 11, 2006

Abstract

This paper complements Alvarez and Arellano (2003) by showing the asymptotic properties of the system GMM estimator for AR(1) panel data models when both N and T tend to infinity. We show that the system GMM estimator with the instruments which Blundell and Bond (1998) used will be inconsistent when both N and T are large. We also show that the system GMM estimator with all available instruments, including redundant ones, will be consistent if $\sigma_\eta^2/\sigma_v^2 = 1 - \alpha$ holds.

*E-mail : ed051009@srv.cc.hit-u.ac.jp

†I am deeply grateful to Taku Yamamoto and Katsuto Tanaka for helpful comments. I also gratefully acknowledge the financial support from the JSPS Fellowship. All remaining errors are mine.

1 Introduction

With the growing availability of comprehensive statistical databases, the use of dynamic panel models has increased steadily in recent decades. The advantages are clear: dynamic panel models not only allow us to take the dynamics of economic activity into account, they also make it possible to control for unobservable heterogeneity. To estimate dynamic panel data models, several estimators have been proposed. These include the instrumental variables estimator (Anderson and Hsiao, 1981), the within groups estimator (Nickell, 1981), the first difference GMM estimator (Arellano and Bond, 1991), the level and the FOD-GMM estimator[1] (Arellano and Bover, 1995), the system GMM estimator (Blundell and Bond, 1998), the LIML-type estimator (Alonso-Borrego and Arellano, 1999), and the random effect maximum likelihood (RML) estimator (Blundell and Smith, 1991; Alvarez and Arellano, 2003). 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), Beck, Levine and Loayza (2000) and others have all used the system GMM estimator.

Motivated by the increasing availability of micropanels in which T, the time series dimension, is not negligible relative to N, the cross-sectional size, a seminal paper by Alvarez and Arellano (2003) discussed the asymptotic properties of some of the estimators mentioned above when both N and T are large. Included in their discussion were the FOD-GMM, the LIML, the first difference GMM, and the RML estimators. However, oddly enough, Alvarez and Arellano did not show the asymptotic properties of the system GMM estimator even though this estimator is widely used in empirical analyses. One possible reason for this omission may be the technical difficulties involved. In fact, to derive the asymptotic properties of the system GMM estimator, we need to show the asymtotic properties of the level GMM estimator. However in deriving the asymptotic properties of the level GMM estimator, we need to derive the explicit expression of the inverse matrix of the population

[1]The FOD-GMM estimator refers to the GMM estimator where individual effects in the model are removed by the forward orthogonal deviation transformation, and the instruments in levels are used in estimation. In Alvarez and Arellano (2003), the FOD-GMM estimator is simply called the GMM estimator.

moment matrix of the instruments. Although deriving the explicit expression of the inverse matrix is somewhat trivial in the case of the first difference GMM estimator, it seems nontrivial in the case of the level GMM estimator. The purpose of the present paper thus is to derive the explicit expression of the inverse matrix and show the asymptotic properties of the level and the system GMM estimators, thereby complementing the work of Alvarez and Arellano (2003).

We find that the level GMM estimator with all instruments becomes inconsistent when both N and T are large, and that the level GMM estimator which Blundell and Bond (1998) used to construct the system GMM estimator will be consistent when N is large regardless of whether T is fixed or tends to infinity. Combining the results of the first difference and the level GMM estimators, we provide the results for the system GMM estimator. We consider three cases. The first is the case where all available instruments, including the redundant ones, are used. In this case, the system GMM estimator will be inconsistent unless $1-\alpha = \sigma_\eta^2/\sigma_v^2$ holds, where α is the parameter of the lagged dependent variable and σ_η^2 and σ_v^2 are the variances of individual effects and disturbances respectively. The second case we consider is where one instrument is used for each period in the first differenced and level models. In this case, the GMM estimator will be consistent when N is large regardless of whether T is fixed or tends to infinity. The third case, finally, is to examine the instruments used by Blundell and Bond (1998). In this case, the GMM estimator will be inconsistent when both N and T are large.

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

2 The Model and Estimators

We consider an AR(1) panel data model given by

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it} \quad i = 1, ..., N \quad and \quad t = 2, ..., T \tag{1}$$

where α is the parameter of interest with $|\alpha| < 1$ and v_{it} has mean zero given $\eta_i, y_{i1}, ..., y_{i,t-1}$. By letting $y_t = (y_{t,1}, ..., y_{t,N})'$, $y_{t-1} = (y_{t-1,1}, ..., y_{t-1,N})$, $u_t =$

$(u_{t,1},...,u_{t,N})$ and $u_{it} = \eta_i + v_{it}$, (1) can be expressed in vector form as

$$y_t = \alpha y_{t-1} + u_t \tag{2}$$

By stacking by time we obtain

$$y = \alpha y_{-1} + u \tag{3}$$

where $y = (y_2', ..., y_T')'$ and $y_{-1} = (y_1', ..., y_{T-1}')'$.

We impose the following assumptions.

Assumption 1. *$\{v_{it}\}$ $(t = 2, ..., T; i = 1, ..., N)$ are i.i.d across time and individuals and independent of η_i and y_{i1} with $E(v_{it}) = 0$, $var(v_{it}) = \sigma_v^2$, and finite moments up to fourth order.*

Assumption 2. *η_i are i.i.d across individuals with $E(\eta_i) = 0$, $var(\eta_i) = \sigma_\eta^2$, and finite moments up to fourth order.*

Assumption 3. *The initial observations satisfy*

$$y_{i1} = \frac{\eta_i}{1-\alpha} + w_{i1} \qquad for \quad i = 1, ..., N \tag{4}$$

where w_{i1} is $w_{i1} = \sum_{j=0}^{\infty} \alpha^j v_{i,1-j}$ and independent of η_i.

These assumptions are the same as those in Alvarez and Arellano (2003).

The GMM estimators we consider in this paper are the first differencing GMM estimator by Arellano and Bond (1991), the level GMM estimator by Arellano and Bover (1995), and the system GMM estimator by Blundell and Bond (1998). For simplicity, we consider the inefficient one-step GMM estimator. We now define these estimators.

By first differencing model (2), we have

$$\Delta y_t = \alpha \Delta y_{t-1} + \Delta v_t \quad t = 3, ..., T \tag{5}$$

We consider two types of instrumental variables, $Z_t^{d1} = (y_1, ..., y_{t-2})$ and $Z_t^{d2} = y_{t-2}$. Let Z^{d1} and Z^{d2} denote block diagonal matrices whose $(t-2)$th blocks are Z_t^{d1} and Z_t^{d2} respectively. Then the first differencing GMM estimators are

$$\hat{\alpha}_{d1} = \frac{\Delta y_{-1}' P^{d1} \Delta y}{\Delta y_{-1}' P^{d1} \Delta y_{-1}} = \frac{\sum_{t=3}^{T} \Delta y_{t-1}' P_t^{d1} \Delta y_t}{\sum_{t=3}^{T} \Delta y_{t-1}' P_t^{d1} \Delta y_{t-1}} \tag{6}$$

$$\hat{\alpha}_{d2} = \frac{\Delta y'_{-1} P^{d2} \Delta y}{\Delta y'_{-1} P^{d2} \Delta y_{-1}} = \frac{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d2} \Delta y_t}{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d2} \Delta y_{t-1}} \tag{7}$$

where $P_t^{d1} = Z_t^{d1}(Z_t^{d1\prime} Z_t^{d1})^{-1} Z_t^{d1\prime}$ and $P_t^{d2} = Z_t^{d2}(Z_t^{d2\prime} Z_t^{d2})^{-1} Z_t^{d2\prime}$.

Next, we consider the level estimator. As instruments for the level model, we use $Z_t^{l1} = (\Delta y_2, ..., \Delta y_{t-1})$ and $Z_t^{l2} = \Delta y_{t-1}$, and let Z^{l1} and Z^{l2} denote block diagonal matrices whose blocks are Z_t^{l1} and Z_t^{l2} respectively. Then the level GMM estimators are

$$\hat{\alpha}_{l1} = \frac{y'_{-1} P^{l1} y}{y'_{-1} P^{l1} y_{-1}} = \frac{\sum_{t=3}^{T} y'_{t-1} P_t^{l1} y_t}{\sum_{t=3}^{T} y'_{t-1} P_t^{l1} y_{t-1}} \tag{8}$$

$$\hat{\alpha}_{l2} = \frac{y'_{-1} P^{l2} y}{y'_{-1} P^{l2} y_{-1}} = \frac{\sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_t}{\sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_{t-1}} \tag{9}$$

where $P_t^{l1} = Z_t^{l1}(Z_t^{l1\prime} Z_t^{l1})^{-1} Z_t^{l1\prime}$ and $P_t^{l2} = Z_t^{l2}(Z_t^{l2\prime} Z_t^{l2})^{-1} Z_t^{l2\prime}$.

We consider three types of system estimators by choosing different instruments. The model of the system estimator can be expressed as

$$\begin{bmatrix} \Delta y_{it} \\ y_{it} \end{bmatrix} = \alpha \begin{bmatrix} \Delta y_{i,t-1} \\ y_{i,t-1} \end{bmatrix} + \begin{bmatrix} \Delta v_{it} \\ u_{it} \end{bmatrix} \tag{10}$$

Let $Z_t^{all} = diag(Z_t^{d1}, Z_t^{l1})$, $Z_t^{min} = diag(Z_t^{d2}, Z_t^{l2})$, and $Z_t^{bb} = diag(Z_t^{d1}, Z_t^{l2})$. Then the system GMM estimator with the instruments Z_t^{all}, Z_t^{min}, and Z_t^{bb} are

$$\hat{\alpha}_{all} = \frac{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d1} \Delta y_t + \sum_{t=3}^{T} y'_{t-1} P_t^{l1} y_t}{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d1} \Delta y_{t-1} + \sum_{t=3}^{T} y'_{t-1} P_t^{l1} y_{t-1}} \tag{11}$$

$$\hat{\alpha}_{min} = \frac{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d2} \Delta y_t + \sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_t}{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d2} \Delta y_{t-1} + \sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_{t-1}} \tag{12}$$

$$\hat{\alpha}_{bb} = \frac{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d1} \Delta y_t + \sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_t}{\sum_{t=3}^{T} \Delta y'_{t-1} P_t^{d1} \Delta y_{t-1} + \sum_{t=3}^{T} y'_{t-1} P_t^{l2} y_{t-1}} \tag{13}$$

$\hat{\alpha}_{all}$ exploits all available instruments in the first differencing and the level estimators. Some instruments in the level estimators are redundant since they are linear transformations of the instruments used in the first differencing estimator. $\hat{\alpha}_{min}$ uses the minimum number of instruments in the sense that it uses one instrument for each period. $\hat{\alpha}_{bb}$ is the system estimator proposed by Blundell and Bond (1998) in which the redundant instruments in the level model are excluded.

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. To derive the main results, we provide some lemmas. Some of these are reported in Alvarez and Arellano (2003).

Lemma 1. *Let Assumptions 1, 2, and 3 hold. Then, as both N and T tend to infinity, provided that $T/N \to c$, $0 \leq c \leq 1$,* [2]

$$\frac{1}{N(T-2)}\Delta y'_{-1}P^{d1}\Delta v \quad \to^p \quad -\sigma_v^2\frac{c}{2} \tag{14}$$

$$\frac{1}{N(T-2)}\Delta y'_{-1}P^{d1}\Delta y_{-1} \quad \to^p \quad \sigma_v^2\left(\frac{c}{2}+\frac{1-\alpha}{1+\alpha}\right) \tag{15}$$

$$\frac{1}{N(T-2)}y'_{-1}P^{l1}u \quad \to^p \quad \frac{1}{2}c\sigma_\eta^2\left(\frac{1}{1-\alpha}\right) \tag{16}$$

$$\frac{1}{N(T-2)}y'_{-1}P^{l1}y_{-1} \quad \to^p \quad \frac{1}{2}c\sigma_\eta^2\left(\frac{1}{1-\alpha}\right)^2+\frac{\sigma_v^2}{1-\alpha^2} \tag{17}$$

Lemma 2. *Let Assumptions 1, 2, and 3 hold. Then, as N tends to infinity, regardless of whether T is fixed or tends to infinity,*

$$\frac{1}{N(T-2)}\Delta y'_{-1}P^{d2}\Delta v \quad \to^p \quad 0 \tag{18}$$

$$\frac{1}{N(T-2)}\Delta y'_{-1}P^{d2}\Delta y_{-1} \quad \to^p \quad \sigma_v^2\left[\left(\frac{1+\alpha}{1-\alpha}\right)+k\left(\frac{1+\alpha}{1-\alpha}\right)^2\right]^{-1} \tag{19}$$

$$\frac{1}{N(T-2)}y'_{-1}P^{l2}u \quad \to^p \quad 0 \tag{20}$$

$$\frac{1}{N(T-2)}y'_{-1}P^{l2}y_{-1} \quad \to^p \quad \frac{\sigma_v^2}{2}\left(\frac{1}{1+\alpha}\right) \tag{21}$$

By utilizing the above lemmas, the following results are readily obtained.

Theorem 1. *Let Assumptions 1, 2, and 3 hold. Then, as both N and T tend to infinity, provided that $T/N \to c$, $0 \leq c \leq 1$,*

$$\hat{\alpha}_{d1} \quad \to^p \quad \alpha-\frac{c}{2\left(\frac{1-\alpha}{1+\alpha}\right)+c} \tag{22}$$

$$\hat{\alpha}_{l1} \quad \to^p \quad \alpha+\frac{\frac{c}{2}k\left(\frac{1}{1-\alpha}\right)}{\frac{c}{2}k\left(\frac{1}{1-\alpha}\right)^2+\frac{1}{1-\alpha^2}} \tag{23}$$

where $k=\sigma_\eta^2/\sigma_v^2$.

[2] We impose this condition since in most of the micro panels N is larger than T.

Theorem 2. *Let Assumptions 1, 2, and 3 hold. Then as $N \to \infty$ regardless of whether $T \to \infty$ or is fixed,*

$$\hat{\alpha}_{d2} \quad \to^p \quad \alpha \tag{24}$$

$$\hat{\alpha}_{l2} \quad \to^p \quad \alpha \tag{25}$$

Remark 1 We find that $\hat{\alpha}_{d1}$ is inconsistent and the direction of the bias is downward. This result has already been shown by Alvarez and Arellano (2003). On the other hand, $\hat{\alpha}_{d2}$ is consistent with large N regardless of whether $T \to \infty$ or is fixed.

Remark 2 We find that $\hat{\alpha}_{l1}$ is inconsistent when both N and T are large, and the direction of the bias is upward. On the other hand, $\hat{\alpha}_{l2}$ is consistent when N is large regardless of whether $T \to \infty$ or is fixed.

Now we consider the system GMM estimators. They can be derived by utilizing the results obtained above.

Theorem 3. *Let Assumptions 1, 2, and 3 hold. Then, as both N and T tend to infinity, provided that $T/N \to c$, $0 \leq c \leq 1$,*

$$\hat{\alpha}_{all} \to^p \alpha + \frac{\frac{1}{2}c\left[-1 + \frac{1}{1-\alpha}k\right]}{\left[\frac{1+(1-\alpha)^2}{1-\alpha^2} + \frac{1}{2}c\right] + \frac{c}{2}k\left(\frac{1}{1-\alpha}\right)^2} \tag{26}$$

$$\hat{\alpha}_{bb} \to^p \alpha - \frac{c}{c + \frac{3-2\alpha}{1+\alpha}} \tag{27}$$

Theorem 4. *Let Assumptions 1, 2, and 3 hold. Then, as $N \to \infty$ regardless of whether $T \to \infty$ or is fixed,*

$$\hat{\alpha}_{min} \to^p \alpha \tag{28}$$

Remark 3 $\hat{\alpha}_{all}$ will be consistent only when $\sigma_\eta^2/\sigma_v^2 = 1 - \alpha$ holds. However, in other cases with $\sigma_\eta^2/\sigma_v^2 \neq 1 - \alpha$, $\hat{\alpha}_{all}$ is inconsistent. $\hat{\alpha}_{min}$ will be consistent when N is large regardless of whether $T \to \infty$ or is fixed.

Remark 4 In the case of $\hat{\alpha}_{all}$, since $\hat{\alpha}_{d1}$ has negative asymptotic bias and $\hat{\alpha}_{l1}$ has positive asymptotic bias, the biases cancel each other out in the system GMM estimator. However, in the case of $\hat{\alpha}_{bb}$, since $\hat{\alpha}_{d1}$ has negative asymptotic bias and $\hat{\alpha}_{l2}$ is consistent, the "balance" between $\hat{\alpha}_{d1}$ and $\hat{\alpha}_{l2}$ breaks down. Hence, unlike $\hat{\alpha}_{all}$, $\hat{\alpha}_{bb}$ is always inconsistent unless $c = 0$.

4 Monte Carlo Experiments

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

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + v_{it} \tag{29}$$

where $\eta_i \sim iidN(0, \sigma_\eta^2)$, $y_{i,1} \sim iidN(\eta_i/(1-\alpha), \sigma_v^2/(1-\alpha^2))$, and $v_{it} \sim iidN(0, \sigma_v^2)$. Here we consider $N = 50, 100$, $T = 10, 25, 50$ and $\sigma_\eta^2 = 0.2, 1, 10$. σ_v^2 is set to 1. The number of replications is 1000 for all cases. For each estimator, we compute the median (median), the interquartile range (iqr), and the median absolute error (mae).

Tables 1 and 2 show the results for the first differencing and the level GMM estimators for the cases of $N = 50$ and $N = 100$, respectively. Tables 3 and 4 report the results for the system GMM estimator for the cases of $N = 50$ and $N = 100$, respectively. The theoretical asymptotic biases calculated in the previous section are tabulated in Table 5.

We begin by considering the first differencing and level GMM estimators. Since the asymptotic properties of $\hat{\alpha}_{d1}$ are discussed by Alvarez and Arellano (2003), here we focus on the other estimators, $\hat{\alpha}_{d2}$, $\hat{\alpha}_{l1}$, and $\hat{\alpha}_{l2}$. Although $\hat{\alpha}_{d2}$ is consistent when N is large regardless of whether T is fixed or tends to infinity, its finite sample bias is substantial even in the case of $\alpha = 0.2, 0.5$, where the effects of weak instruments may be small. Especially as σ_η^2/σ_v^2 gets larger, the bias gets larger.

Next, we turn to the level GMM estimators. Table 1 shows that the bias of $\hat{\alpha}_{l1}$ is substantial when T and σ_η^2/σ_v^2 are large. For example, in the case of $\alpha = 0.2$, $T = 50$ and $\sigma_\eta^2/\sigma_v^2 = 10$, although the true value is $\alpha = 0.2$, the median is 0.908. This value obtained by simulation is close to the theoretical asymptotic bias reported in Table 5. Also, although $\hat{\alpha}_{l2}$ is consistent when N is large, its finite sample bias crucially

depends upon the magnitude of σ_η^2/σ_v^2. In the case of $\alpha = 0.2$ and $\sigma_\eta^2/\sigma_v^2 = 0.2$, the finite sample biases of $\hat{\alpha}_{l2}$ at around 0.21 are not so large. However, when $\alpha = 0.2$ and $\sigma_\eta^2/\sigma_v^2 = 10$, the biases of $\hat{\alpha}_{l2}$ are around 0.52 and the finite sample bias is substantially large.

Finally, we consider the system GMM estimator. The theoretical results indicate that if $1 - \alpha = \sigma_\eta^2/\sigma_v^2$ holds, $\hat{\alpha}_{all}$ is consistent. The simulation confirms this theoretical result. In the case of $\alpha = 0.8$ and $\sigma_\eta^2/\sigma_v^2 = 0.2$, $\hat{\alpha}_{all}$ will be consistent when both N and T are large. Looking at Table 4, the biases of $\hat{\alpha}_{all}$ are very small irrespective of T and N. However, in the other cases where $\hat{\alpha}_{all}$ is inconsistent, the biases of $\hat{\alpha}_{all}$ are very large. The magnitude of the asymptotic biases obtained by simulation is very close to the theoretical values. For example, in the case of $\alpha = 0.2$, $T = N = 50$ and $\sigma_\eta^2/\sigma_v^2 = 10$, the simulation value is 0.776, while the theoretical value is 0.774. Although $\hat{\alpha}_{min}$ is consistent when N is large regardless of whether T is fixed or tends to infinity, its finite sample biases are quite large especially when σ_η^2/σ_v^2 is large.[3] With regards to $\hat{\alpha}_{bb}$, the magnitude of the biases is substantially large in almost all the cases. Comparing the theoretical value for the asymptotic bias and the simulation result for the case of $\alpha = 0.2$, $N = T = 50$, and $\sigma_\eta^2/\sigma_v^2 = 0.2$, we find that both are very close: the theoretical value is -0.116, while the simulation value is -0.111. However, in the case of $\sigma_\eta^2/\sigma_v^2 = 10$, the simulation value is 0.061 but theoretical value is -0.116. This difference springs from the large finite sample bias of the consistent estimator $\hat{\alpha}_{l2}$.

These simulation results have several implications. The first is that the instruments which Blundell and Bond (1998) used does not provide desirable results when T is large. Using all instruments, including redundant ones, would be preferable since $\hat{\alpha}_{all}$ is consistent if $1 - \alpha = \sigma_\eta^2/\sigma_v^2$ holds. In contrast, $\hat{\alpha}_{bb}$ is not consistent. Second, although we can reduce the number of instruments and obtain consistency, there remain large finite sample biases when σ_η^2/σ_v^2 is large. Thus, when the value of T is not negligible relative to N, and σ_η^2/σ_v^2 is large, the system GMM estimator does not work well. In such cases where the system GMM estimator breaks down, it is advisable to use the FOD-GMM estimator.

[3]Bun and Kiviet (2006) and Hayakawa (2005) have shown that the finite sample bias of the system GMM estimator is heavily affected by the magnitude of σ_η^2/σ_v^2.

Alvarez and Arellano (2003) have shown that the FOD-GMM estimator is consistent when both N and T are large, and it is robust to large σ_η^2/σ_v^2. However, Okui (2005) and Hayakawa (2006) have shown that inference based on the FOD-GMM estimator is inaccurate. To overcome this problem, Hayakawa (2006) proposed a new form of instruments with which both the asymptotic bias and the variance of the GMM estimator can be reduced simultaneously and which allow accurate inferences. Hence, when T is large and σ_η^2/σ_v^2 may be large, we propose to use the GMM estimator by Hayakawa (2006) instead of the system GMM estimator.

5 Conclusion

In this paper, we considered the asymptotic properties of the system GMM estimators when both N and T are large. We showed that if we use the all available instruments, including redundant ones in the level model, the system GMM estimator will be inconsistent except for the case when $1-\alpha = \sigma_\eta^2/\sigma_v^2$ holds. If we reduce the number of instruments so that we use one instruments for each period, the system GMM estimator is consistent, although its finite sample bias becomes very large when σ_η^2/σ_v^2 is large. We also showed that the original system GMM estimator by Blundell and Bond (1998) will be inconsistent unless $c = 0$. Thus, the system GMM estimator is not recommendable when T is not negligible relative to N and σ_η^2/σ_v^2 may be large. In this case, one possible solution is to use the GMM estimator proposed by Hayakawa (2006), since its asymptotic bias and variance become small simultaneously and inference is accurate even if there is large heterogeneity.

References

[1] Alonso-Borrego, C. and M. Arellano (1999) " Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business and Economic Statistics*, 17, 36-49.

[2] Alvarez, J. and M. Arellano (2003) " The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121-1159.

[3] Anderson, T.W. and C. Hsiao (1981) " Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598-606.

[4] Arellano, M. and S.R. Bond (1991) " Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.

[5] Arellano, M. and O. Bover (1995) " Another Look at the Instrumental Variable Estimation of Error-Component Models," *Journal of Econometrics*, 68, 29-45.

[6] Beck, T., R. Levine, and N. Loayza (2000) " Finance and the Sources of Growth," *Journal of Financial Economics*, 58, 261 - 300.

[7] Blundell, R. and S. Bond (1998) " Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-43.

[8] Blundell, R. and S. Bond (2000) " GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 19, 321-340.

[9] Blundell, R. and R. Smith (1991) " Initial Conditions and Efficient Estimation in Dynamic Panel Data Models," *Annales d'Economie et de Statistique*, 20/21, 109-123.

[10] Bun, M. and J. Kiviet (2006) " The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, forthcoming.

[11] Bond, S., A. Hoeffler and J. Temple (2001) " GMM Estimation of Empirical Growth Models," University of Oxford, Nuffield College, Economics Group, Working Paper No. 2001-W21.

[12] Dollar, D. and A. Kraay (2002) " Growth is Good for the Poor," *Journal of Economic Growth*, 7, 195 - 225.

[13] Hayakawa, K. (2005) " Small Sample Bias Properties of the System GMM Estimators in Dynamic Panel Data Models," Hi-Stat Discussion Paper, No.82, Hitotsubashi University.

[14] Hayakawa, K. (2006) " Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present," Hi-Stat Discussion Paper, No.130, Hitotsubashi University.

[15] Hamilton, J.D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

[16] Okui, R. (2005): "The Optimal Choice of Moments in Dynamic Panel Data Models," unpublished manuscript.

[17] Tanaka, K. (1996): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory,* New York, Wiley.

A Appendix

In this appendix, we give the proofs of the lemmas and theorems given in the main context. To begin with, note that, under Assumptions 1, 2, and 3, y_{it} can be expressed as

$$y_{it} = \left(\frac{1}{1-\alpha}\right)\eta_i + w_{it} \tag{30}$$

where $w_{it} = \sum_{j=0}^{\infty} \alpha^j v_{i,t-j}$. Note that $\Delta y_{it} = \Delta w_{it}$ holds. We give some lemmas which are useful to prove the main results.

Lemma 3. *Under Assumptions 1, 2, and 3, we have*

$$\begin{aligned}
\left[E(z_{it}^{l1} z_{it}^{l1'})\right]^{-1} &= \sigma_v^{-2} \begin{bmatrix} \frac{2}{1+\alpha} & -\left(\frac{1-\alpha}{1+\alpha}\right) & \cdots & -\alpha^{t-4}\left(\frac{1-\alpha}{1+\alpha}\right) \\ -\left(\frac{1-\alpha}{1+\alpha}\right) & \frac{2}{1+\alpha} & & \\ \vdots & & \ddots & \\ -\alpha^{t-4}\left(\frac{1-\alpha}{1+\alpha}\right) & & & \frac{2}{1+\alpha} \end{bmatrix}^{-1} \\
&= \Upsilon_{11} - \frac{\Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) E(w_{i,t-1} z_{it}^{l1'}) \Upsilon_{11}}{E(w_{i,t-1}^2) + E(w_{i,t-1} z_{it}^{l1'}) \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1})}
\end{aligned} \tag{31}$$

$$E(w_{i,t-1} z_{it}^{l1'}) \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) = \frac{\sigma_v^2 (t-2)}{(1+\alpha)^2} \tag{32}$$

where $z_{it}^{l1} = (\Delta w_{i,2}, ..., \Delta w_{i,t-1})'$, $E(w_{i,t-1} z_{it}^{l1'}) = \sigma_v^2 (\alpha^{t-3}, \cdots, 1)/(1+\alpha)$, $E(w_{i,t-1}^2) = \sigma_v^2/(1-\alpha^2)$, *and*

$$\Upsilon_{11} = \sigma_v^{-2}(\alpha^2 - \alpha + 1) I_{t-2} + \sigma_v^{-2}(1-\alpha) \begin{bmatrix} \alpha & \iota_{t-3}' \\ \iota_{t-3} & \Lambda \end{bmatrix} \tag{33}$$

$$\Lambda = \begin{bmatrix} 1 & 2-\alpha & 2-\alpha & 2-\alpha & \cdots & \cdots & 2-\alpha \\ 2-\alpha & 2-\alpha & 3-2\alpha & 3-2\alpha & \cdots & \cdots & 3-2\alpha \\ 2-\alpha & 3-2\alpha & 3-2\alpha & 4-3\alpha & \cdots & \cdots & 4-3\alpha \\ 2-\alpha & 3-2\alpha & 4-3\alpha & 4-3\alpha & & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & & \ddots & (t-3)-(t-4)\alpha \\ 2-\alpha & 3-2\alpha & 4-3\alpha & \cdots & \cdots & (t-3)-(t-4)\alpha & (t-3)-(t-4)\alpha \end{bmatrix}$$

$$\iota_{t-3} = (1, ..., 1)'$$

Proof of (31) and (32) To derive the explicit expression of $[E(z_{it}^{l1} z_{it}^{l1\prime})]^{-1}$, let us define $\tilde{z}_{it}$ as follows:

$$\begin{aligned}\tilde{z}_{it} &= \begin{bmatrix} z_{it}^{l1} \\ -w_{i,t-1} \end{bmatrix} = \begin{bmatrix} \Delta w_{i,2} \\ \Delta w_{i,3} \\ \vdots \\ \Delta w_{i,t-1} \\ -w_{i,t-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & & & O \\ 0 & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & 1 & 0 \\ & & & & -1 & 1 \\ O & & & & 0 & -1 \end{bmatrix} \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,t-1} \end{bmatrix} \\ &= \tilde{D} w_i^{(t-1)} \end{aligned} \tag{34}$$

where $w_i^{(t-1)} = (w_{i,1}, ..., w_{i,t-1})'$. Then it follows that

$$\begin{aligned}\left[E(\tilde{z}_{it}\tilde{z}_{it}')\right]^{-1} &= \left[\tilde{D}E(w_i^{(t-1)} w_i^{(t-1)\prime})\tilde{D}'\right]^{-1} = \left(\tilde{D}'\right)^{-1} \left[E(w_i^{(t-1)} w_i^{(t-1)\prime})\right]^{-1} \left(\tilde{D}\right)^{-1} \\ &= \begin{bmatrix} E(z_{it}^{l1} z_{it}^{l1\prime}) & -E(z_{it}^{l1} w_{i,t-1}) \\ -E(z_{it}^{l1\prime} w_{i,t-1}) & E(w_{i,t-1}^2) \end{bmatrix}^{-1} = \begin{bmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{bmatrix}\end{aligned} \tag{35}$$

where Υ_{11}, Υ_{12}, Υ_{21}, and Υ_{22} are partitioned conformably, i.e. Υ_{11} is a $(t-2)\times(t-2)$ matrix, Υ_{12} and Υ_{21}' are $(t-2)\times 1$ vectors, and Υ_{22} is a scalar. From the partitioned inverse formula, we have

$$\Upsilon_{11} = \left[E(z_{it}^{l1} z_{it}^{l1\prime}) - E(w_{i,t-1} z_{it}^{l1})[E(w_{i,t-1}^2)]^{-1} E(z_{it}^{l1} w_{i,t-1})\right]^{-1} \tag{36}$$

After some algebra, we obtain

$$\left[E(z_{it}^{l1} z_{it}^{l1\prime})\right]^{-1} = \left[\Upsilon_{11}^{-1} + E(w_{i,t-1} z_{i,t}^{l1})[E(w_{i,t-1}^2)]^{-1} E(z_{it}^{l1\prime} w_{i,t-1})\right]^{-1} \tag{37}$$

Using the fact that $(A + BCB')^{-1} = A^{-1} - A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1}$, (37) can be expressed as

$$\left[E(z_{it}^{l1} z_{it}^{l1\prime})\right]^{-1} = \Upsilon_{11} - \frac{\Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) E(w_{i,t-1} z_{it}^{l1\prime}) \Upsilon_{11}}{E(w_{i,t-1}^2) + E(w_{i,t-1} z_{it}^{l1\prime}) \Upsilon_{11} E(w_{i,t-1} z_{it}^{l1})} \tag{38}$$

If Υ_{11} can be expressed explicitly, the explicit form of $[E(z_{it}^{l1} z_{it}^{l1\prime})]^{-1}$ is readily obtained since it is straightforward to obtain the expectations in (38). To this end, we need to calculate $\tilde{D}^{-1}$ and $[E(w_i^{(t-1)} w_i^{(t-1)\prime})]^{-1}$ in (35). To begin with, from Tanaka (1996) we have

$$\tilde{D}^{-1} = \begin{bmatrix} -1 & \cdots & -1 \\ & \ddots & \vdots \\ O & & -1 \end{bmatrix} \tag{39}$$

Next, it follows that[4]

$$\left[E(w_i^{(t-1)}w_i^{(t-1)'})\right]^{-1} = \sigma_v^{-2}L'L \tag{40}$$

where

$$L = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix} \tag{41}$$

Thus, using these results, we can calculate Υ_{11} as in (33) and, hence, the explicit formula of $[E(z_{it}^{l1}z_{it}^{l1'})]^{-1}$ is obtained. With regards to (32), after some manipulation, we get

$$\begin{aligned} E(w_{i,t-1}z_{it}^{l1'})\Upsilon_{11}E(z_{it}^{l1}w_{i,t-1}) &= \begin{bmatrix} E(w_{i,t-1}z_{it}^{l1'}) & 0 \end{bmatrix} \begin{bmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{bmatrix} \begin{bmatrix} E(z_{it}^{l1}w_{i,t-1}) \\ 0 \end{bmatrix} \\ &= \frac{\sigma_v^2(t-2)}{(1+\alpha)^2} \end{aligned} \tag{42}$$

Lemma 4. *Under Assumptions 1, 2, and 3, as $T \to \infty$ regardless of whether $N \to \infty$ or is fixed, we have*

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} w_{t-2}'P_t^{d1}w_{t-2} \to^p \frac{\sigma_v^2}{1-\alpha^2} \tag{43}$$

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} w_{t-1}'P_t^{l1}w_{t-1} \to^p \frac{\sigma_v^2}{1-\alpha^2} \tag{44}$$

and as $N \to \infty$ regardless of whether $T \to \infty$ or is fixed, we have

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} w_{t-1}'P_t^{d2}w_{t-1} \to^p \frac{\sigma_v^2}{1-\alpha^2}\left[1 + k\left(\frac{1+\alpha}{1-\alpha}\right)\right]^{-1} \tag{45}$$

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} w_{t-1}'P_t^{l2}w_{t-1} \to^p \frac{\sigma_v^2}{2}\left(\frac{1}{1+\alpha}\right) \tag{46}$$

Proof of Lemma 4

Proof of (43) See Lemma C2 in Alvarez and Arellano (2003).

[4]See Hamilton (1994, p.120).

Proof of (44) Let ε_t denote the $N \times 1$ vector of errors of the population linear projection of w_{t-1} on Z_t^{l1}:

$$w_{t-1} = Z_t^{l1}\delta_t + \varepsilon_t \tag{47}$$

where $\delta_t = [E(z_{it}^{l1} z_{it}^{l1\prime})]^{-1} E(z_{it}^{l1} w_{i,t-1})$. Using Lemma 3, δ_t can be expressed as

$$\delta_t = \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) \left(\frac{E(w_{i,t-1}^2)}{E(w_{i,t-1}^2) + E(w_{i,t-1} z_{it}^{l1\prime}) \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1})} \right) \tag{48}$$

Also, note that

$$\begin{aligned}
\begin{bmatrix} \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) \\ \Upsilon_{21} E(z_{it}^{l1} w_{i,t-1}) \end{bmatrix} &= \begin{bmatrix} \Upsilon_{11} & \Upsilon_{12} \\ \Upsilon_{21} & \Upsilon_{22} \end{bmatrix} \begin{bmatrix} E(z_{it}^{l1} w_{i,t-1}) \\ 0 \end{bmatrix} \\
&= \left(\tilde{D}'\right)^{-1} \left[E(w_i^{(t-1)} w_i^{(t-1)\prime}) \right]^{-1} \left(\tilde{D}\right)^{-1} \begin{bmatrix} E(z_{it}^{l1} w_{i,t-1}) \\ 0 \end{bmatrix} \\
&= \left(\frac{1}{1+\alpha}\right) \begin{bmatrix} 1 & & O \\ \vdots & \ddots & \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ (1-\alpha) \\ \vdots \\ (1-\alpha) \\ -\alpha \end{bmatrix}
\end{aligned} \tag{49}$$

Hence, from (47), (48), and (49), it follows that

$$\begin{aligned}
\varepsilon_{it} &= w_{i,t-1} - z_{it}^{l1\prime} \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1}) \left(\frac{E(w_{i,t-1}^2)}{E(w_{i,t-1}^2) + E(w_{i,t-1} z_{it}^{l1\prime}) \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1})} \right) \\
&= w_{i,t-1} - \frac{w_{i,t-1} - w_{i1} + (1-\alpha)\{(t-3) w_{i,t-1} - (w_{i,2} + \cdots + w_{i,t-2})\}}{1+\alpha} \\
&\qquad \times \left(\frac{E(w_{i,t-1}^2)}{E(w_{i,t-1}^2) + E(w_{i,t-1} z_{it}^{l1\prime}) \Upsilon_{11} E(z_{it}^{l1} w_{i,t-1})} \right) \\
&= \frac{\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left(\frac{1}{1+\alpha}\right)[(1-\alpha)(w_{i,1} + \cdots + w_{i,t-1}) + \alpha(w_{i,t-1} + w_{i,1})]}{\left(\frac{\sigma_v^2}{1-\alpha^2}\right) + \frac{\sigma_v^2}{(1+\alpha)^2}(t-2)} \\
&= \frac{(1+\alpha)^{-1}(v_{i,t-1} + \cdots + v_{i,2}) + w_{i,1}}{1 + \left(\frac{1-\alpha}{1+\alpha}\right)(t-2)}
\end{aligned} \tag{50}$$

The last equality is due to the fact that $(1-\alpha)(w_{i,t-1} + \cdots + w_{i,1}) = (v_{i,t-1} + \cdots + v_{i,2} + w_{i,1}) - \alpha w_{i,t-1}$. Since (50) is a linear combination of $(t-1)$ independent random variables, we have

$$E(\varepsilon_{it}^2) = \frac{(1+\alpha)^{-2}(t-2)\sigma_v^2 + \frac{\sigma_v^2}{1-\alpha^2}}{\left[1 + \left(\frac{1-\alpha}{1+\alpha}\right)(t-2)\right]^2} = O\left(\frac{1}{t}\right) \tag{51}$$

Now we consider the decomposition:

$$w'_{t-1}P_t^{l1}w_{t-1} = w'_{t-1}w_{t-1} - w'_{t-1}(I_N - P_t^{l1})w_{t-1} \quad (52)$$

$$= w'_{t-1}w_{t-1} - \varepsilon'_t(I_N - P_t^{l1})\varepsilon_t \quad (53)$$

The second equality is due to the fact that $(I_N - P_t^{l1})w_{t-1} = (I_N - P_t^{l1})(Z_t^{l1}\delta_t + \varepsilon_t)$. Then we have

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} E(w'_{t-1}P_t^{l1}w_{t-1}) = E(w_{i,t-1}^2) - \frac{1}{N(T-2)}\sum_{t=3}^{T} E[\varepsilon'_t(I_N - P_t^{l1})\varepsilon_t] \quad (54)$$

Since the maximum eigenvalue of $(I_N - P_t^{l1})$ is equal to 1,

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} E(\varepsilon'_t(I_N - P_t^{l1})\varepsilon_t) \leq \frac{1}{N(T-2)}\sum_{t=3}^{T} E(\varepsilon'_t\varepsilon_t)$$

$$= \frac{1}{(T-2)}\sum_{t=3}^{T} E(\varepsilon_{it}^2) = \frac{1}{T-2}O(\log T) \to 0$$

Hence, as $T \to \infty$, it follows that

$$\frac{1}{N(T-2)}\sum_{t=3}^{T} E(w'_{t-1}P_t^{l1}w_{t-1}) \to E(w_{i,t-1}^2) = \frac{\sigma_v^2}{1-\alpha^2} \quad (55)$$

With regards to the proofs that the variances of $(N(T-2))^{-1}\sum_{t=3}^{T} w'_{t-1}w_{t-1}$ and $(N(T-2))^{-1}\sum_{t=3}^{T} \varepsilon'_t\varepsilon_t$ tend to zero, see Alvarez and Arellano (2003).

Proof of (45) and (46) See Hayakawa (2006).

Lemma 5. *Let κ_3^v and κ_4^v denote the third and fourth order cumulants of v_{it} and let κ_3^η and κ_4^η denote the third and fourth order cumulants of η_i. Also, let d_t, d_s, l_t and l_s denote the diagonal elements of P_t^d, P_s^d, P_t^l and P_s^l respectively so that, for $P_t^d = P_t^{d1}$ and $P_t^l = P_t^{l1}$, $d'_t d_s, l'_t l_s \leq (s-2)$, and for $P_t^d = P_t^{d2}$ and $P_t^l = P_t^{l2}$, $d'_t d_s, l'_t l_s \leq 1$. Then under Assumptions 1, 2, and 3, we have*

(i)

$$var(v'_{t-1}P_t^d v_{t-1}) = E(d'_t d_t)\kappa_4^v + 2\sigma_v^4 tr(P_t^d) \leq \begin{cases} (\kappa_4^v + 2\sigma_v^4)(t-2) & if \quad P_t^d = P_t^{d1} \\ (\kappa_4^v + 2\sigma_v^4) & if \quad P_t^d = P_t^{d2} \end{cases} \quad (56)$$

For $t > s$,

$$cov(v'_{t-1}P_t^d v_{t-1}, v'_{s-1}P_s^d v_{s-1}) = 0 \quad (57)$$

(ii)

$$E(v'_{t-1}P_t^d v_{t-1}v'_{t-1}P_t^d w_{t-2}) = \kappa_3^v E(d'_t P_t^d w_{-2}) \tag{58}$$

$$|E(d'_t P_t^d w_{t-2})| \leq \begin{cases} \sqrt{N(t-2)}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)^{1/2} & if \quad P_t^d = P_t^{d1} \\ \sqrt{N}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)^{1/2} & if \quad P_t^d = P_t^{d2} \end{cases} \tag{59}$$

(iii)

$$var(w'_{t-1}P_t^d v_{t-1}) = \alpha^2\sigma_v^2 E(w'_{t-2}P_t^d w_{t-2}) + 2\alpha\kappa_3^v E(d'_t P_t^d w_{t-2}) + E(d'_t d_t)\kappa_4^v + 2\sigma_v^4 tr(P_t^d) \tag{60}$$

For $t > s$,

$$cov(w'_{t-1}P_t^d v_{t-1}, w'_{s-1}P_s^d v_{s-1}) = 0 \tag{61}$$

(iv)

$$var(\eta' P_t^l \eta) = E(l'_t l_t)\kappa_4^\eta + 2\sigma_\eta^4 tr(P_t^l) \leq \begin{cases} (\kappa_4^\eta + 2\sigma_\eta^4)(t-2) & if \quad P_t^l = P_t^{l1} \\ (\kappa_4^\eta + 2\sigma_\eta^4) & if \quad P_t^l = P_t^{l2} \end{cases} \tag{62}$$

For $t > s$,

$$cov(\eta' P_t^l \eta, \eta' P_s^l \eta) = E(l'_t l_s)\kappa_4^\eta + 2\sigma_\eta^4 E[tr(P_t^l P_s^l)] \leq \begin{cases} (\kappa_4^\eta + 2\sigma_\eta^4)(s-2) & if \quad P_t^l = P_t^{l1} \\ (\kappa_4^\eta + 2\sigma_\eta^4) & if \quad P_t^l = P_t^{l2} \end{cases} \tag{63}$$

(v)

$$var(\eta' P_t^l w_{t-1}) = \sigma_\eta^2 E(w'_{t-1}P_t^l w_{t-1}) \tag{64}$$

For $t > s$

$$|cov(\eta' P_t^l w_{t-1}, \eta' P_s^l w_{s-1})| \leq N\left(\frac{\sigma_v^2\sigma_\eta^2}{1-\alpha^2}\right) \tag{65}$$

Proof of Lemma 5

proof of (i) For $t > s$,

$$\begin{aligned} cov(v'_{t-1}P^d_t v_{t-1}, v'_{s-1}P^d_s v_{s-1}) &= E(v'_{t-1}P^d_t v_{t-1}v'_{s-1}P^d_s v_{s-1}) - E(v'_{t-1}P^d_t v_{t-1})E(v'_{s-1}P^d_s v_{s-1}) \\ &= E[tr\{P^d_t E_{t-1}(v_{t-1}v'_{t-1})\}v'_{s-1}P^d_s v_{s-1}] - \sigma_v^4 tr(P^d_t)tr(P^d_s) \\ &= \sigma_v^2 tr(P^d_t)E[tr\{P^d_s E_{s-1}(v_{s-1}v'_{s-1})\}] - \sigma_v^4 tr(P^d_t)tr(P^d_s) \\ &= 0 \end{aligned}$$

where $E_t(.)$ denotes an expectation conditional on η_i and $\{v_{i,t-j}\}_{j=1}^{\infty}$. For $t = s$, see Alvarez and Arellano (2003).

proof of (ii) With regards to the proof of (58), see Alvarez and Arellano (2003). To prove (59), we consider the case of $P^d_t = P^{d1}_t$. From the Cauchy-Schwarz inequality,

$$(d'_t P^{d1}_t w_{t-2})^2 \leq (d'_t P^{d1}_t d_t)(w'_{t-2}P^{d1}_t w_{t-2}) \leq (d'_t d_t)(w'_{t-2}P^{d1}_t w_{t-2}) \leq (t-2)w'_{t-2}w_{t-2} \quad (66)$$

Hence,

$$E(d'_t P^{d1}_t w_{t-2})^2 \leq (t-2)E(w'_{t-2}w_{t-2}) = (t-2)N\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \quad (67)$$

Because $E(d'_t P^{d1}_t w_{t-2})^2 = var(d'_t P^{d1}_t w_{t-2}) + [E(d'_t P^{d1}_t w_{t-2})]^2$, it follows that

$$[E(d'_t P^{d1}_t w_{t-2})]^2 \leq (t-2)N\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \quad (68)$$

The case of $P^d_t = P^{d2}_t$ is proven in a similar way.

proof of (iii) By using $w_{t-1} = \alpha w_{t-2} + v_{t-1}$, we get

$$\begin{aligned} cov(w'_{t-1}P^d_t v_{t-1}, w'_{s-1}P^d_s v_{s-1}) &= \alpha^2 E(w'_{t-2}P^d_t v_{t-1}v'_{s-1}P^d_s w_{s-2}) + cov(v'_{t-1}P^d_t v_{t-1}, v'_{s-1}P^d_s v_{s-1}) \\ &\quad +\alpha E(w'_{t-2}P^d_t v_{t-1}v'_{s-1}P^d_s v_{s-1} + w'_{s-2}P^d_s v_{s-1}v'_{t-1}P^d_t v_{t-1}) \end{aligned}$$

In the case of $t = s$, from (56) and (58),

$$\begin{aligned} var(w'_{t-1}P^d_t v_{t-1}) &= \alpha^2\sigma_v^2 E(w'_{t-2}P^d_t w_{t-2}) + 2\alpha E(v'_{t-1}P^d_t v_{t-1}v'_{t-1}P^d_t w_{t-2}) + var(v'_{t-1}P^d_t v_{t-1}) \\ &= \alpha^2\sigma_v^2 E(w'_{t-2}P^d_t w_{t-2}) + 2\alpha\kappa_3^v E(d'_t P^d_t w_{t-2}) + E(d'_t d_s)\kappa_4^v + 2\sigma_v^4 tr(P^d_t P^d_s) \end{aligned}$$

For $t > s$, by using (57), the result is readily obtained.

proof of (iv) The proofs are analogous to (i).

proof of (v) Since $|cov(\eta' P_t^l w_{t-1}, \eta' P_s^l w_{s-1})| \leq \sqrt{var(\eta' P_t^l w_{t-1})}\sqrt{var(\eta' P_s^l w_{s-1})}$, we have

$$\begin{aligned}
|cov(\eta' P_t^l w_{t-1}, \eta' P_s^l w_{s-1})| &\leq \sqrt{var(\eta' P_t^l w_{t-1})}\sqrt{var(\eta' P_s^l w_{s-1})} && (69)\\
&= \sigma_\eta^2 \sqrt{E(w'_{t-1} P_t^l w_{t-1})}\sqrt{E(w'_{s-1} P_s^l w_{s-1})} && (70)\\
&\leq \sigma_\eta^2 \sqrt{E(w'_{t-1} w_{t-1})}\sqrt{E(w'_{s-1} w_{s-1})} && (71)\\
&= N\left(\frac{\sigma_v^2 \sigma_\eta^2}{1-\alpha^2}\right) && (72)
\end{aligned}$$

Proof of (14) and (18)

We shall use the decomposition as follows,

$$\begin{aligned}
\frac{\Delta y'_{-1} P^d \Delta v}{N(T-2)} &= \frac{1}{N(T-2)} \sum_{t=3}^{T} \Delta y'_{t-1} P_t^d \Delta v_t \\
&= -A^{(1)} + A^{(2)} - A^{(3)} + A^{(4)}
\end{aligned}$$

where

$$A^{(1)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-1} P_t^d v_{t-1} \qquad A^{(2)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-1} P_t^d v_t \tag{73}$$

$$A^{(3)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-2} P_t^d v_t \qquad A^{(4)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-2} P_t^d v_{t-1} \tag{74}$$

Note that only $A^{(1)}$ has nonzero mean which is given by $\sigma_v^2 \sum_{t=3}^{T} tr(P_t^d)/N(T-2)$. If $P_t^d = P_t^{d1}$, $\lim_{N,T\to\infty} E(A^{(1)}) = \sigma_v^2 c/2$ and if $P_t^d = P_t^{d2}$, $\lim_{N\to\infty} E(A^{(1)}) = 0$. To prove that these are the probability limits, we show that the variances of $A^{(1)}$, $A^{(2)}$, $A^{(3)}$, and $A^{(4)}$ tend to zero. From (60) and (61), we have

$$\begin{aligned}
var(A^{(1)}) &= \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} var(w'_{t-1} P_t^d v_{t-1}) \\
&\quad + \frac{2}{N^2(T-2)^2} \sum_{s} \sum_{t>s} cov(w'_{t-1} P_t^d v_{t-1}, w'_{s-1} P_s^d v_{s-1}) \\
&= \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} var(w'_{t-1} P_t^d v_{t-1}) \\
&= \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} \Big[\alpha^2 \sigma_v^2 E(w'_{t-2} P_t^d w_{t-2}) + 2\alpha \kappa_3^v E(d'_t P_t^d w_{t-2})
\end{aligned}$$

$$+E(d_t'd_t)\kappa_4^v + 2\sigma_v^2 tr(P_t^d)\Big]$$

In view of Lemmas 4 and 5, the first, third and fourth terms converge to zero. The second term would be

$$\begin{aligned}\left|\frac{2\alpha}{N^2(T-2)^2}\sum_{t=3}^{T}\kappa_3^v E(d_t'P_t^d w_{t-2})\right| &\leq \frac{|2\alpha|}{N^2(T-2)^2}\sum_{t=3}^{T}|\kappa_3^v|\left|E(d_t'P_t^d w_{t-2})\right| \\ &\leq \frac{|2\alpha||\kappa_3^v|}{N^2(T-2)^2}\sum_{t=3}^{T}\sqrt{N(t-2)}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)^{1/2} \\ &< \left(\frac{\sigma_v^2}{1-\alpha^2}\right)^{1/2}\frac{|2\alpha||\kappa_3^v|}{N^{3/2}(T-2)^2}\sum_{t=3}^{T}(t-2)\to 0\end{aligned}$$

Hence, $var(A^{(1)})$ converges to zero. Next, we consider the variance of $A^{(2)}$.

$$\begin{aligned}var(A^{(2)}) &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T}var(w_{t-1}'P_t^d v_t) + \frac{2}{N^2(T-2)^2}\sum_{s}\sum_{t>s}cov(w_{t-1}'P_t^d v_t, w_{s-1}'P_s^d v_s) \\ &= \frac{\sigma_v^2}{N^2(T-2)^2}\sum_{t=3}^{T}E(w_{t-1}'P_t^d w_{t-1}) \\ &= \frac{1}{N^2(T-2)^2}\left(\frac{\sigma_v^4}{1-\alpha^2}\right)\sum_{t=3}^{T}tr(P_t^d)\to 0\end{aligned}$$

The second equality is due to the fact that $var(w_{t-1}'P_t^d v_t) = E(w_{t-1}'P_t^d v_t v_t'P_t^d w_{t-1}) = E(w_{t-1}'P_t^d E_t(v_t v_t')P_t^d w_{t-1}) = \sigma_v^2 E(w_{t-1}'P_t^d w_{t-1})$ and for $t > s$ $cov(w_{t-1}'P_t^d v_t, w_{s-1}'P_s^d v_s) = E(w_{t-1}'P_t^d E_t(v_t)w_{s-1}'P_s^d v_s) = 0$. The last convergence holds because $\sum_{t=3}^{T}tr(P_t^d) = O(T^2)$ for $P_t^d = P_t^{d1}$, and $\sum_{t=3}^{T}tr(P_t^d) = O(T)$ for $P_t^d = P_t^{d2}$. With regards to $var(A^{(3)})$, from Lemma 4, it follows that

$$\begin{aligned}var(A^{(3)}) &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T}var(w_{t-2}'P_t^d v_t) + \frac{2}{N^2(T-2)^2}\sum_{s}\sum_{t>s}cov(w_{t-2}'P_t^d v_t, w_{s-2}'P_s^d v_s) \\ &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T}E(w_{t-2}'P_t^d w_{t-2})\to 0\end{aligned}$$

$var(A^{(4)})\to 0$ is proven in a similar way to $var(A^{(3)})$.

Proof of (15) and (19)

We shall use the decomposition as follows:

$$\begin{aligned}\frac{\Delta y_{-1}'P^d\Delta y_{-1}}{N(T-2)} &= \frac{1}{N(T-2)}\sum_{t=3}^{T}\Delta y_{t-1}'P_t^d\Delta y_{t-1} \\ &= (1-\alpha)^2 B^{(1)} - 2(1-\alpha)B^{(2)} + B^{(3)}\end{aligned}$$

where

$$B^{(1)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-2} P_t^d w_{t-2} \qquad B^{(2)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-2} P_t^d v_{t-1} \tag{75}$$

$$B^{(3)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} v'_{t-1} P_t^d v_{t-1} \tag{76}$$

For $B^{(1)}$ and $B^{(2)}$, see Lemma 4 and $A^{(3)}$ respectively. The expectation of $B^{(3)}$ is given by

$$E(B^{(3)}) = \frac{\sigma_v^2}{N(T-2)} \sum_{t=3}^{T} tr(P_t^d) \tag{77}$$

If $P_t^d = P_t^{d1}$, $\lim_{N,T\to\infty} E(B^{(3)}) = \sigma_v^2 c/2$ and if $P_t^d = P_t^{d2}$, $\lim_{N\to\infty} E(B^{(3)}) = 0$. The variance of $B^{(3)}$ is shown to converge to zero as follows,

$$\begin{aligned} var(B^{(3)}) &= \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} var(v'_{t-1} P_t^d v_{t-1}) \\ &\qquad + \frac{2}{N^2(T-2)^2} \sum_{s} \sum_{t>s} cov(v'_{t-1} P_t^d v_{t-1}, v'_{s-1} P_s^d v_{s-1}) \\ &= \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} var(v'_{t-1} P_t^d v_{t-1}) \\ &\leq \frac{1}{N^2(T-2)^2} \sum_{t=3}^{T} (\kappa_4^v + 2\sigma_v^4) tr(P_t^d) \to 0 \end{aligned}$$

Proof of (16) and (20)

We decompose as follows:

$$\begin{aligned} \frac{y'_{-1} P^l u}{N(T-2)} &= \frac{1}{N(T-2)} \sum_{t=3}^{T} y'_{t-1} P_t^l u_t \\ &= \left(\frac{1}{1-\alpha}\right) C^{(1)} + \left(\frac{1}{1-\alpha}\right) C^{(2)} + C^{(3)} + C^{(4)} \end{aligned}$$

where

$$C^{(1)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} \eta' P_t^l \eta \qquad C^{(2)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} \eta' P_t^l v_t$$

$$C^{(3)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-1} P_t^l \eta \qquad C^{(4)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w'_{t-1} P_t^l v_t$$

Only $C^{(1)}$ has nonzero mean which is given by

$$E(C^{(1)}) = \frac{\sigma_\eta^2}{N(T-2)} \sum_{t=3}^{T} tr(P_t^l)$$

If $P_t^l = P_t^{l1}$, $\lim_{N,T\to\infty} E(C^{(1)}) = c\sigma_\eta^2/2$ and if $P_t^l = P_t^{l2}$, $\lim_{N\to\infty} E(C^{(1)}) = 0$. Using Lemmas 4 and 5, the variances of $C^{(1)}$, $C^{(2)}$, $C^{(3)}$, and $C^{(4)}$ are shown to tend to zero as follows:

$$\begin{aligned}
var(C^{(1)}) &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T} var(\eta' P_t^l \eta) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s} cov(\eta' P_t^l \eta, \eta' P_s^l \eta) \\
&= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T}\left[E(l_t' l_t)\kappa_4^\eta + 2\sigma_\eta^4 tr(P_t^l)\right] \\
&\quad + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s}\left[E(l_t' l_s)\kappa_4^\eta + 2\sigma_\eta^4 E(tr(P_t^l P_s^l))\right] \\
&\leq \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T}(\kappa_4^\eta + 2\sigma_\eta^4)(t-2) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s}(s-2)(\kappa_4^\eta + 2\sigma_\eta^2) \to 0
\end{aligned}$$

$$\begin{aligned}
var(C^{(2)}) &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T} var(v_t' P_t^l \eta) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s} cov(v_t' P_t^l \eta, v_s' P_s^l \eta) \\
&= \frac{\sigma_\eta^2}{N^2(T-2)^2}\sum_{t=3}^{T} E(v_t' P_t^l v_t) \\
&= \frac{\sigma_v^2\sigma_\eta^2}{N^2(T-2)^2}\sum_{t=3}^{T} tr(P_t) \to 0
\end{aligned}$$

$$\begin{aligned}
|var(C^{(3)})| &\leq \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T} var(\eta' P_t^l w_{t-1}) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s} |cov(\eta' P_t^l w_{t-1}, \eta' P_s^l w_{s-1})| \\
&= \frac{\sigma_\eta^2}{N^2(T-2)^2}\sum_{t=3}^{T} E(w_{t-1}' P_t^l w_{t-1}) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s} |cov(\eta' P_t^l w_{t-1}, \eta' P_s^l w_{s-1})| \\
&\leq \frac{\sigma_\eta^2}{N^2(T-2)^2}\sum_{t=3}^{T} E(w_{t-1}' P_t^l w_{t-1}) + \frac{2}{N(T-2)^2}\sum_s\sum_{t>s}\left(\frac{\sigma_v^2\sigma_\eta^2}{1-\alpha^2}\right) \\
&\to 0
\end{aligned}$$

$$\begin{aligned}
var(C^{(4)}) &= \frac{1}{N^2(T-2)^2}\sum_{t=3}^{T} var(w_{t-1}' P_t^l v_t) + \frac{2}{N^2(T-2)^2}\sum_s\sum_{t>s} cov(w_{t-1}' P_t^l v_t, w_{s-1}' P_s^l v_t) \\
&= \frac{\sigma_v^2}{N^2(T-2)^2}\sum_{t=3}^{T} E(w_{t-1}' P_t^l w_{t-1}) \to 0
\end{aligned}$$

Proof of (17) and (21)

We decompose as follows:

$$\frac{y_{-1}' P^l y_{-1}}{N(T-2)} = \frac{1}{N(T-2)}\sum_{t=3}^{T} y_{t-1}' P_t^l y_{t-1}$$

$$= \left(\frac{1}{1-\alpha}\right)^2 D^{(1)} + D^{(2)} + \left(\frac{2}{1-\alpha}\right) D^{(3)}$$

where

$$D^{(1)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} \eta' P_t^l \eta \qquad D^{(2)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w_{t-1}' P_t^l w_{t-1} \tag{78}$$

$$D^{(3)} = \frac{1}{N(T-2)} \sum_{t=3}^{T} w_{t-1}' P_t^l \eta \tag{79}$$

Notice that $D^{(1)} = C^{(1)}$ and $D^{(3)} = C^{(3)}$. With regards to $D^{(2)}$, see Lemma 4.

Proof of Theorems 1, 2, 3, 4 Theorems 1, 2, 3, and 4 are immediately shown by using Lemmas 1 and 2.

Table 1: Medians, Interquartile Ranges, and Median Absolute Errors of the First Differencing and Level GMM Estimators ($N = 50$)

		$\sigma_\eta^2/\sigma_v^2 = 0.2$				$\sigma_\eta^2/\sigma_v^2 = 1$				$\sigma_\eta^2/\sigma_v^2 = 10$			
T		$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$
$\alpha = 0.2$													
10	median	0.068	0.165	0.223	0.207	0.051	0.140	0.318	0.254	0.032	-0.032	0.698	0.510
	iqr	0.102	0.104	0.086	0.104	0.115	0.119	0.123	0.118	0.124	0.241	0.153	0.181
	mae	0.132	0.058	0.046	0.053	0.149	0.073	0.119	0.070	0.168	0.233	0.498	0.310
25	median	-0.079	0.164	0.260	0.213	-0.095	0.135	0.431	0.253	-0.101	-0.050	0.839	0.521
	iqr	0.053	0.063	0.054	0.065	0.054	0.072	0.086	0.064	0.050	0.137	0.062	0.115
	mae	0.279	0.041	0.060	0.033	0.295	0.066	0.231	0.055	0.301	0.250	0.639	0.321
50	median	-0.235	0.164	0.308	0.211	-0.242	0.132	0.551	0.255	-0.246	-0.061	0.908	0.526
	iqr	0.030	0.042	0.040	0.042	0.029	0.044	0.066	0.044	0.030	0.103	0.026	0.092
	mae	0.435	0.037	0.108	0.023	0.442	0.068	0.351	0.055	0.446	0.261	0.708	0.326
$\alpha = 0.5$													
10	median	0.234	0.412	0.537	0.518	0.197	0.329	0.647	0.588	0.173	-0.024	0.897	0.838
	iqr	0.134	0.149	0.089	0.115	0.146	0.199	0.109	0.127	0.152	0.355	0.075	0.111
	mae	0.266	0.095	0.056	0.058	0.303	0.173	0.147	0.094	0.327	0.524	0.397	0.338
25	median	0.043	0.412	0.576	0.522	0.022	0.319	0.738	0.594	0.012	-0.023	0.950	0.841
	iqr	0.064	0.078	0.052	0.071	0.065	0.106	0.060	0.072	0.064	0.201	0.024	0.065
	mae	0.457	0.089	0.076	0.038	0.478	0.181	0.238	0.094	0.488	0.523	0.450	0.341
50	median	-0.114	0.418	0.621	0.522	-0.127	0.317	0.810	0.590	-0.129	-0.025	0.971	0.843
	iqr	0.031	0.049	0.037	0.045	0.033	0.082	0.042	0.052	0.033	0.158	0.010	0.050
	mae	0.614	0.082	0.121	0.028	0.627	0.183	0.310	0.090	0.629	0.525	0.471	0.343
$\alpha = 0.8$													
10	median	0.203	0.492	0.856	0.843	0.112	0.191	0.933	0.916	0.088	-0.043	0.990	0.987
	iqr	0.196	0.265	0.066	0.116	0.208	0.423	0.054	0.083	0.210	0.449	0.023	0.036
	mae	0.597	0.308	0.059	0.061	0.688	0.609	0.133	0.117	0.712	0.843	0.190	0.187
25	median	0.042	0.480	0.880	0.843	0.020	0.209	0.954	0.919	0.015	-0.040	0.994	0.989
	iqr	0.070	0.163	0.033	0.064	0.076	0.214	0.020	0.052	0.072	0.251	0.007	0.019
	mae	0.758	0.320	0.080	0.047	0.780	0.591	0.154	0.119	0.785	0.840	0.194	0.189
50	median	-0.043	0.474	0.905	0.846	-0.051	0.192	0.970	0.919	-0.051	-0.038	0.996	0.987
	iqr	0.037	0.107	0.019	0.044	0.036	0.150	0.010	0.035	0.034	0.178	0.002	0.013
	mae	0.843	0.326	0.105	0.046	0.851	0.608	0.170	0.119	0.851	0.838	0.196	0.187

Note: 1000 replications; "iqr" is the interquartile range; "mae" denotes the median absolute error.

Table 2: Medians, Interquartile Ranges, and Median Absolute Errors of the First Differencing and Level GMM Estimators ($N = 100$)

		$\sigma_\eta^2/\sigma_v^2 = 0.2$				$\sigma_\eta^2/\sigma_v^2 = 1$				$\sigma_\eta^2/\sigma_v^2 = 10$			
T		$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{d2}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{l2}$
$\alpha = 0.2$													
10	median	0.129	0.184	0.212	0.203	0.116	0.166	0.264	0.225	0.105	0.030	0.574	0.406
	iqr	0.079	0.078	0.065	0.076	0.086	0.082	0.091	0.083	0.088	0.186	0.156	0.136
	mae	0.072	0.039	0.033	0.039	0.085	0.048	0.068	0.045	0.096	0.170	0.374	0.206
25	median	0.034	0.181	0.232	0.207	0.022	0.164	0.335	0.225	0.021	0.036	0.736	0.408
	iqr	0.040	0.041	0.036	0.042	0.040	0.047	0.060	0.044	0.043	0.104	0.084	0.084
	mae	0.166	0.025	0.032	0.021	0.178	0.037	0.135	0.029	0.179	0.164	0.536	0.208
50	median	-0.081	0.181	0.260	0.207	-0.087	0.164	0.427	0.228	-0.091	0.031	0.837	0.411
	iqr	0.026	0.030	0.026	0.030	0.026	0.034	0.059	0.031	0.025	0.072	0.045	0.061
	mae	0.281	0.021	0.060	0.016	0.287	0.036	0.227	0.028	0.291	0.169	0.637	0.211
$\alpha = 0.5$													
10	median	0.349	0.453	0.519	0.511	0.310	0.402	0.587	0.550	0.295	0.136	0.837	0.754
	iqr	0.094	0.093	0.066	0.082	0.112	0.137	0.090	0.087	0.112	0.325	0.092	0.111
	mae	0.151	0.060	0.036	0.041	0.190	0.102	0.088	0.058	0.205	0.364	0.337	0.254
25	median	0.198	0.456	0.542	0.512	0.178	0.397	0.657	0.552	0.171	0.103	0.910	0.765
	iqr	0.053	0.053	0.034	0.047	0.052	0.075	0.055	0.053	0.048	0.170	0.038	0.069
	mae	0.302	0.045	0.042	0.025	0.322	0.103	0.157	0.052	0.329	0.397	0.410	0.265
50	median	0.049	0.454	0.572	0.510	0.039	0.395	0.730	0.552	0.035	0.094	0.946	0.764
	iqr	0.030	0.036	0.026	0.034	0.031	0.051	0.041	0.035	0.030	0.126	0.017	0.048
	mae	0.451	0.046	0.072	0.018	0.461	0.105	0.230	0.052	0.465	0.406	0.446	0.264
$\alpha = 0.8$													
10	median	0.364	0.623	0.833	0.829	0.258	0.372	0.901	0.884	0.215	0.001	0.983	0.978
	iqr	0.161	0.174	0.056	0.084	0.177	0.323	0.055	0.078	0.190	0.434	0.027	0.036
	mae	0.436	0.177	0.039	0.048	0.542	0.428	0.101	0.085	0.585	0.799	0.183	0.178
25	median	0.177	0.607	0.853	0.827	0.145	0.351	0.930	0.889	0.135	0.007	0.989	0.978
	iqr	0.075	0.107	0.028	0.050	0.071	0.192	0.024	0.044	0.069	0.246	0.010	0.022
	mae	0.623	0.193	0.053	0.033	0.655	0.449	0.130	0.089	0.665	0.793	0.189	0.178
50	median	0.057	0.603	0.874	0.827	0.046	0.356	0.950	0.889	0.044	-0.005	0.993	0.978
	iqr	0.035	0.078	0.019	0.032	0.036	0.125	0.013	0.032	0.037	0.165	0.004	0.014
	mae	0.743	0.197	0.074	0.028	0.754	0.444	0.150	0.089	0.756	0.805	0.193	0.178

Note: 1000 replications; "iqr" is the interquartile range; "mae" denotes the median absolute error.

Table 3: Medians, Interquartile Ranges, and Median Absolute Errors of the System GMM Estimators ($N = 50$)

		$\sigma_\eta^2/\sigma_v^2 = 0.2$			$\sigma_\eta^2/\sigma_v^2 = 1$			$\sigma_\eta^2/\sigma_v^2 = 10$		
T		$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$
$\alpha = 0.2$										
10	median	0.157	0.186	0.124	0.213	0.211	0.141	0.564	0.457	0.310
	iqr	0.081	0.103	0.101	0.105	0.107	0.106	0.173	0.168	0.154
	mae	0.052	0.051	0.078	0.052	0.053	0.069	0.364	0.257	0.115
25	median	0.102	0.188	0.020	0.224	0.206	0.032	0.698	0.467	0.197
	iqr	0.052	0.061	0.055	0.076	0.058	0.053	0.094	0.108	0.093
	mae	0.098	0.031	0.180	0.041	0.029	0.168	0.498	0.267	0.047
50	median	0.037	0.185	-0.111	0.240	0.209	-0.094	0.776	0.472	0.061
	iqr	0.037	0.041	0.031	0.071	0.045	0.034	0.053	0.088	0.074
	mae	0.163	0.024	0.311	0.045	0.022	0.294	0.576	0.272	0.139
$\alpha = 0.5$										
10	median	0.453	0.488	0.389	0.552	0.531	0.434	0.853	0.809	0.709
	iqr	0.092	0.110	0.121	0.123	0.116	0.128	0.092	0.102	0.136
	mae	0.054	0.055	0.112	0.072	0.065	0.077	0.353	0.309	0.209
25	median	0.415	0.480	0.241	0.589	0.539	0.286	0.906	0.813	0.589
	iqr	0.055	0.066	0.060	0.078	0.069	0.067	0.037	0.065	0.099
	mae	0.085	0.035	0.259	0.089	0.045	0.214	0.406	0.313	0.093
50	median	0.374	0.480	0.087	0.624	0.535	0.129	0.930	0.815	0.445
	iqr	0.045	0.046	0.036	0.064	0.047	0.043	0.022	0.052	0.094
	mae	0.126	0.028	0.413	0.124	0.037	0.371	0.430	0.315	0.061
$\alpha = 0.8$										
10	median	0.793	0.799	0.677	0.902	0.891	0.809	0.985	0.983	0.966
	iqr	0.080	0.105	0.117	0.059	0.083	0.108	0.024	0.036	0.040
	mae	0.039	0.052	0.123	0.102	0.092	0.055	0.185	0.183	0.166
25	median	0.798	0.801	0.503	0.917	0.891	0.670	0.989	0.985	0.936
	iqr	0.049	0.064	0.075	0.031	0.051	0.087	0.007	0.019	0.033
	mae	0.024	0.032	0.297	0.117	0.091	0.130	0.189	0.185	0.136
50	median	0.796	0.798	0.332	0.929	0.892	0.518	0.992	0.983	0.884
	iqr	0.037	0.042	0.048	0.018	0.036	0.075	0.003	0.014	0.040
	mae	0.018	0.021	0.468	0.129	0.092	0.282	0.192	0.183	0.084

Note: 1000 replications; "iqr" is the interquartile range; "mae" denotes the median absolute error.

Table 4: Medians, Interquartile Ranges, and Median Absolute Errors of the System GMM Estimators ($N = 100$)

		$\sigma_\eta^2/\sigma_v^2 = 0.2$			$\sigma_\eta^2/\sigma_v^2 = 1$			$\sigma_\eta^2/\sigma_v^2 = 10$		
T		$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{min}$	$\hat{\alpha}_{bb}$
$\alpha = 0.2$										
10	median	0.175	0.193	0.161	0.208	0.202	0.165	0.457	0.373	0.274
	iqr	0.062	0.074	0.074	0.074	0.072	0.081	0.154	0.123	0.109
	mae	0.034	0.037	0.045	0.037	0.035	0.047	0.257	0.173	0.081
25	median	0.146	0.194	0.099	0.214	0.202	0.104	0.596	0.374	0.200
	iqr	0.033	0.042	0.039	0.050	0.042	0.041	0.100	0.075	0.062
	mae	0.054	0.022	0.101	0.025	0.021	0.096	0.396	0.174	0.031
50	median	0.103	0.192	0.014	0.226	0.204	0.020	0.695	0.376	0.112
	iqr	0.024	0.030	0.027	0.046	0.033	0.028	0.068	0.056	0.042
	mae	0.097	0.016	0.186	0.030	0.016	0.180	0.495	0.176	0.088
$\alpha = 0.5$										
10	median	0.475	0.487	0.435	0.530	0.520	0.460	0.793	0.735	0.649
	iqr	0.064	0.079	0.086	0.084	0.087	0.091	0.102	0.103	0.115
	mae	0.037	0.042	0.067	0.048	0.045	0.053	0.293	0.235	0.150
25	median	0.452	0.489	0.347	0.562	0.521	0.367	0.865	0.745	0.569
	iqr	0.039	0.049	0.049	0.061	0.048	0.049	0.051	0.066	0.083
	mae	0.048	0.026	0.153	0.062	0.029	0.133	0.365	0.245	0.072
50	median	0.420	0.489	0.229	0.589	0.524	0.256	0.904	0.743	0.458
	iqr	0.031	0.031	0.029	0.050	0.031	0.029	0.027	0.047	0.059
	mae	0.080	0.018	0.271	0.089	0.025	0.244	0.404	0.243	0.046
$\alpha = 0.8$										
10	median	0.796	0.798	0.722	0.877	0.866	0.802	0.977	0.974	0.956
	iqr	0.064	0.086	0.096	0.058	0.075	0.097	0.027	0.036	0.041
	mae	0.030	0.043	0.079	0.078	0.068	0.048	0.177	0.174	0.156
25	median	0.799	0.799	0.599	0.899	0.866	0.700	0.985	0.974	0.925
	iqr	0.036	0.051	0.055	0.032	0.044	0.061	0.011	0.022	0.033
	mae	0.017	0.025	0.201	0.099	0.066	0.100	0.185	0.174	0.125
50	median	0.799	0.802	0.456	0.915	0.870	0.575	0.989	0.974	0.881
	iqr	0.028	0.035	0.037	0.021	0.030	0.052	0.005	0.014	0.034
	mae	0.014	0.017	0.344	0.115	0.070	0.225	0.189	0.174	0.081

Note: 1000 replications; "iqr" is the interquartile range; "mae" denotes the median absolute error.

Table 5: The Asymptotic Bias of the GMM Estimators

		$k=0.2$				$k=1$				$k=10$			
α	c	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{bb}$	$\hat{\alpha}_{d1}$	$\hat{\alpha}_{l1}$	$\hat{\alpha}_{all}$	$\hat{\alpha}_{bb}$
0.2													
	0.1	0.130	0.212	0.179	0.156	0.130	0.256	0.207	0.156	0.130	0.543	0.426	0.156
	0.2	0.070	0.223	0.159	0.115	0.070	0.304	0.213	0.115	0.070	0.680	0.541	0.115
	0.25	0.042	0.229	0.150	0.097	0.042	0.326	0.215	0.097	0.042	0.722	0.580	0.097
	0.5	-0.073	0.256	0.108	0.013	-0.073	0.418	0.227	0.013	-0.073	0.832	0.690	0.013
	1	-0.229	0.304	0.041	-0.116	-0.229	0.543	0.242	-0.116	-0.229	0.906	0.774	-0.116
0.5													
	0.1	0.370	0.515	0.483	0.430	0.370	0.565	0.526	0.430	0.370	0.800	0.756	0.430
	0.2	0.269	0.528	0.468	0.370	0.269	0.615	0.546	0.370	0.269	0.875	0.829	0.370
	0.25	0.227	0.535	0.460	0.342	0.227	0.636	0.555	0.342	0.227	0.895	0.850	0.342
	0.5	0.071	0.565	0.429	0.227	0.071	0.714	0.586	0.227	0.071	0.941	0.899	0.227
	1	-0.100	0.615	0.383	0.071	-0.100	0.800	0.620	0.071	-0.100	0.969	0.929	0.071
0.8													
	0.1	0.490	0.817	0.800	0.686	0.490	0.862	0.848	0.686	0.490	0.964	0.959	0.686
	0.2	0.326	0.831	0.800	0.595	0.326	0.895	0.873	0.595	0.326	0.980	0.975	0.595
	0.25	0.271	0.837	0.800	0.557	0.271	0.906	0.881	0.557	0.271	0.984	0.979	0.557
	0.5	0.108	0.862	0.800	0.409	0.108	0.938	0.907	0.409	0.108	0.991	0.987	0.409
	1	-0.018	0.895	0.800	0.238	-0.018	0.964	0.926	0.238	-0.018	0.996	0.991	0.238

Note: $c = 0.1 : (T,N) = (10,100)$, $c = 0.2 : (T,N) = (10,50)$, $c = 0.25 : (T,N) = (25,100)$, $c = 0.5 : (T,N) = (25,50), (50,100)$, $c = 1 : (T,N) = (50,50)$, where c=T/N.