



**Discussion Paper Series**

No.177

**A Bias-Corrected Estimation for  
Dynamic Panel Models in Small Samples**

Hiroaki Chigira  
Taku Yamamoto

July 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/>

# A BIAS-CORRECTED ESTIMATOR FOR DYNAMIC PANEL MODELS IN SMALL SAMPLES\*

HIROAKI CHIGIRA<sup>‡</sup>

TAKU YAMAMOTO

Department of Economics  
Hitotsubashi University

July 2006

## ABSTRACT

This paper is concerned with the estimation of the autoregressive parameter of dynamic panel data models. We propose a bias-corrected GMM estimator whose bias is smaller than that of many existing GMM estimators. And we propose a small sample corrected estimator of the variance in order to reduce the size distortion of the Wald test. These estimators are easy to calculate and do not require preliminary estimates. The Monte Carlo experiments indicate that in terms of both bias and size distortion, the bias corrected estimator outperforms Blundell and Bond's (1998) system estimator even when using Windmeijer's (2005) correction of the estimated variance of the system estimator.

Key words: Generalized method of moments; bias correction; panel data.

JEL classifications: C12, C23

---

\*An earlier version of this paper was presented at the Center for Operations Research and Econometrics and Hitotsubashi University.

<sup>‡</sup>Corresponding author: Hiroaki Chigira. Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, JAPAN. E-mail: hchigira@mercury.ne.jp. The author gratefully acknowledges financial support from a JSPS Research Fellowship for Young Scientists.

# 1 INTRODUCTION

Dynamic panel models are used to examine dynamic relationships in panel data and include a lagged dependent variable as a regressor. A frequent problem encountered in estimating the parameters of dynamic panel models is that estimators are biased. For example, in pure time series models, the OLS estimator of an autoregressive parameter is known to have asymptotic bias of order  $1/T$ , where  $T$  is the sample size. The problem becomes more serious in panel data models because the presence of individual effects causes correlation between a lagged dependent variable and error components. Therefore, in the estimation of dynamic panel models, careful attention needs to be paid to small sample bias.

In order to eliminate such bias, Anderson and Hsiao (1981) suggest using the instrumental variable method, which is based on the first difference of the model with a lagged level variable as an instrument. Arellano and Bond (1991), however, show that there are additional instruments that could be used but are not exploited by Anderson and Hsiao (1981). Employing the generalized method of moments (GMM) approach, they demonstrate that using these instruments yields efficiency gains. We will call the GMM estimator proposed by Arellano and Bond (1991) the first differencing estimator. An alternative GMM estimator was proposed by Arellano and Bover (1995), which is based on the level of the model and uses lagged differenced variables as instruments. We will refer to this estimator as the level GMM estimator. Building on these two studies, Blundell and Bond (1998) argue that the use of both instruments - those proposed by Arellano and Bond (1991) and those proposed by Arellano and Bover (1995) - results in a dramatic gain in efficiency. In their Monte Carlo experiment, they show that their approach has smaller bias than the first differencing estimator. We will call Blundell and Bond's (1998) estimator the system GMM estimator.

Unfortunately, the system GMM estimator does not always work well. Bun and Kiviet (2005) show that the bias of the system estimator becomes large when the autoregressive parameter is close to unity and/or when the ratio of the variance of the individual effect to that of the disturbance departs from unity. They also show that the system estimator has a bias of the order  $1/N$ , where  $N$  is the number of individuals. This means that even when the sample size  $T$  is sufficiently large, the system estimator still exhibits a bias as long as  $N$  is small. Finally, Bun and Kiviet show that the biases of the first differencing

estimator and the level estimator are  $O(1/N)$ .

The GMM procedure suffers from an additional problem. Although a gain in efficiency is achieved by calculating the so-called two-step GMM estimator, the estimated asymptotic variance of the two-step GMM estimator could be severely downward-biased in small samples. This means that the Wald test based upon the two-step GMM procedure exhibits severe size distortions. Several methods to correct the downward bias of the variance of the two-step estimator have been proposed (see, e.g., Windmeijer, 2005). However, Bond and Windmeijer (2005) find in their Monte Carlo simulations that when the autoregressive parameter is close to unity, the system GMM procedure suffers from size distortion even when Windmeijer's (2005) correction is applied. One reason why the system GMM with Windmeijer's correction exhibits size distortions could be that a preliminary estimate, i.e., the one-step system GMM estimate, of the autoregressive parameter is needed for Windmeijer's correction. Because, as mentioned above, the one-step system estimator exhibits severe bias when the autoregressive parameter approaches unity, Windmeijer's correction would not work well in such a case.

There are ways other than the GMM approach which can reduce the bias and size distortion. Kiviet (1995), Hanh and Kuersteiner (2002), Bun and Carree (2005) and many others have constructed estimators for dynamic panel data models by modifying the within class estimator. Among them, Bun and Carree (2005) has few restrictions on the procedure. They argue that their estimator overcomes some of the drawbacks of GMM estimators. However, their estimator crucially depends on the assumption that the regressors are strictly exogenous with respect to the error term, while the GMM estimators do not require such a restrictive assumption.

The purpose of this paper is to propose a bias-corrected GMM estimator of the autoregressive parameter and a finite sample corrected estimator of the variance. The proposed estimators are based on the approach developed by Kurozumi and Yamamoto (2000) and overcome some of the drawbacks of existing GMM estimators. First, the bias of our bias-corrected estimator is  $o(\frac{1}{NT})$ , while that of the system estimator is  $O(1/N)$ . Second, in contrast with Windmeijer's (2005) approach, no preliminary estimate is necessary to obtain our small sample corrected estimator of the variance. In addition, because the bias-corrected estimator is based on a modified version of the system estimator, it is possible to relax the assumption of strict exogeneity. The Monte Carlo experiments indicate that the bias corrected estimator outperforms the system estimator with Windmeijer's

correction in terms of small sample bias and size distortion.

The remainder of this paper is organized as follows. Section 2 provides a review of several GMM estimators, including the first differencing estimator, the level estimator and the system estimator. Section 3 presents our bias-corrected estimator and our small sample corrected estimator of the variance. Section 4 proposes an additional instrument in order to further improve the efficiency of the estimator. Section 5 presents the Monte Carlo results of a comparison of the finite sample performance of the bias-corrected estimator and the system estimator with Windmeijer's (2005) correction. Section 6 then provides an empirical illustration consisting of the estimation of a panel data model of employment dynamics. Finally, Section 7 offers some concluding remarks.

## 2 THE MODEL AND THE GMM ESTIMATORS

Consider a dynamic panel data process of the form

$$y_{it} = \alpha y_{i,t-1} + \beta' w_{it} + \gamma_i + \varepsilon_{it}, \quad (1)$$

for  $i = 1, \dots, N$  and  $t = 2, \dots, T$ , where  $\alpha$  is an autoregressive parameter with  $|\alpha| < 1$ ,  $w_{it}$  is a set of exogenous variables,  $\gamma_i$  is the unobservable individual effect and  $\varepsilon_{it}$  is the disturbance.

Although we allow for the inclusion of exogenous regressor  $w_{it}$  in our Monte Carlo experiments and our empirical example, we develop our procedure using a model generated by an autoregressive model of order 1 without exogenous variables,

$$y_{it} = \alpha y_{i,t-1} + \gamma_i + \varepsilon_{it}. \quad (2)$$

We consider model (2) because our approach is based on Blundell and Bond (1998), who also use this model. We make the following standard assumption:

Assumption 1:

- (i)  $\gamma_i$  and  $\varepsilon_{it}$  are independently distributed across  $i$  with mean 0 and variance  $\sigma_\gamma^2$  and  $\sigma_\varepsilon^2$ , respectively.
- (ii)  $E(\varepsilon_{it}\gamma_i) = 0$ , for  $i = 1, \dots, N$  and  $t = 2, \dots, T$ .
- (iii)  $E(\varepsilon_{it}\varepsilon_{is}) = 0$ , for  $i = 1, \dots, N$  and  $\forall t \neq s$ .

(iv)  $E(y_{i1}\varepsilon_{it}) = 0$ , for  $i = 1, \dots, N$  and  $t = 2, \dots, T$ .

(v)  $y_{i1} = \frac{\gamma_i}{1-\alpha} + w_{i1}$  for  $i = 1, \dots, N$ ,

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

Assumptions 1(i) - (iv) are the same as in Blundell and Bond (1998). Assumption 1(v) is the same as in Alvarez and Arellano (2003).

Stacking equation (2), we obtain

$$y_i = \alpha x_i + u_i, \tag{3}$$

where  $y_i = (y_{i3}, \dots, y_{iT})'$ ,  $x_i = (y_{i2}, \dots, y_{i,T-1})'$  and  $u_i = (u_{i3}, \dots, u_{iT})$  with  $u_{it} = \gamma_i + \varepsilon_{it}$ .

Because the bias-corrected estimator which we propose is based on a modified version of the system estimator, we here briefly review the different GMM estimators, including the system estimator. The GMM estimators are based on Assumption 1.

### *The first differencing estimator*

In model (3), we find that the existence of the individual effect  $\gamma_i$  causes a severe correlation between the regressor  $x_i$  and the error term  $u_i$ . In order to eliminate the individual effect, Arellano and Bond (1991) take first differences of model (3):

$$\Delta y_i = \alpha \Delta x_i + \Delta u_i, \tag{4}$$

and then show that

$$E(Z_i^{D'} \Delta u_i) = 0, \tag{5}$$

where

$$Z_i^D = \begin{bmatrix} y_{i1} & & & & & & \\ & (y_{i1} \ y_{i2}) & & & & & 0 \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & & & & & & \\ & 0 & & & & (y_{i1} \ \dots \ y_{i,T-2}) & \end{bmatrix}. \tag{6}$$

Using (5) as the orthogonal conditions in the GMM, Arellano and Bond (1991) construct the two-step first differencing estimator for  $\alpha$ , which is given by

$$\hat{\alpha}^D = \frac{\Delta x' Z^D A^D Z^{D'} \Delta y}{\Delta x' Z^D A^D Z^{D'} \Delta x}, \tag{7}$$

where  $\Delta x = (\Delta x'_1, \dots, \Delta x'_N)'$ ,  $\Delta y = (\Delta y'_1, \dots, \Delta y'_N)'$ ,  $Z^D = (Z_1^{D'}, \dots, Z_N^{D'})'$  and

$$A^D = \left( \frac{1}{N} \sum_{i=1}^N Z_i^{D'} \hat{\Delta} u_i \hat{\Delta} u_i' Z_i^D \right)^{-1}$$

with residuals from the one-step consistent estimator,  $\hat{\Delta} u_i$ .

As mentioned by Blundell and Bond (1998), instrument (6) becomes invalid when  $\alpha$  is close to unity and/or  $\sigma_\gamma^2/\sigma_\varepsilon^2$  increases. To illustrate this feature, we take  $T = 3$  as an example. Then the first differencing estimator is reduced to an instrumental variable estimator and the auxiliary regression which regresses the explanatory variable on the instrumental variable can be expressed as

$$\Delta y_{i2} = \pi y_{i1} + r_i.$$

After some algebra, the probability limit of  $\hat{\pi}$  is given by

$$\hat{\pi} \xrightarrow{p} (1 - \alpha) \frac{-(1 - \alpha)/(1 + \alpha)}{\sigma_\gamma^2/\sigma_\varepsilon^2 + (1 - \alpha)/(1 + \alpha)}.$$

We find that as  $\alpha$  gets close to unity or  $\sigma_\gamma^2/\sigma_\varepsilon^2$  becomes large,  $\text{plim } \hat{\pi}$  gets close to zero. Following Blundell and Bond (1998), we refer to this as the weak instrument problem.

### *The level GMM estimator*

Arellano and Bover (1995) suggest to eliminate the individual effect from the instrumental variables, while, as mentioned above, Arellano and Bond (1991) propose to eliminate it from the model. Explicitly, Arellano and Bover (1995) consider the level model (3) and then show that the variable

$$Z_i^L = \begin{bmatrix} \Delta y_{i2} & & 0 \\ & \ddots & \\ 0 & & \Delta y_{i,T-1} \end{bmatrix}, \quad (8)$$

which contains no individual effect because of first differencing, satisfies the orthogonal conditions

$$E(Z_i^{L'} u_i) = 0. \quad (9)$$

Using (9), Arellano and Bover's (1995) two-step level estimator is calculated as follows:

$$\hat{\alpha}^L = \frac{x' Z^L A^L Z^{L'} y}{x' Z^L A^L Z^{L'} x}, \quad (10)$$

where  $x = (x'_1, \dots, x'_N)'$ ,  $y = (y'_1, \dots, y'_N)'$ ,  $Z^L = (Z^L_1, \dots, Z^L_N)'$  and

$$A^L = \left( \frac{1}{N} \sum_{i=1}^N Z_i^{L'} \hat{u}_i \hat{u}'_i Z_i^L \right)^{-1}$$

with residuals from the one-step consistent estimator,  $\hat{u}_i$ .

We can easily see that the level estimator is free from the weak instrument problem in the case where  $T = 3$ . In this case, we obtain the auxiliary regression

$$y_{i2} = \pi \Delta y_{i2} + r_i.$$

As Blundell, Bond and Windmeijer (2000) have shown,

$$\hat{\pi} \xrightarrow{p} \frac{1}{2}.$$

Because  $\text{plim } \hat{\pi}$  does not approach zero, the instruments are always informative.

### *The system GMM estimator*

Blundell and Bond (1998) argue that the GMM estimator with both moment conditions (5) and (9) overcomes the weak instrument problem and achieves a gain in efficiency. Combining (2) and (4), they use a system to construct a GMM estimator utilizing these moment conditions. Their two-step system estimator is given by

$$\hat{\alpha}^S = \frac{x^{*'} Z^S A^S Z^{S'} y^*}{x^{*'} Z^S A^S Z^{S'} x^*}, \quad (11)$$

where  $x^* = [(\Delta x'_1, x'_1), \dots, (\Delta x'_N, x'_N)]'$ ,  $y^* = [(\Delta y'_1, y'_1), \dots, (\Delta y'_N, y'_N)]'$ ,  $Z^S = (Z^S_1, \dots, Z^S_N)'$ ,

$$Z_i^S = \begin{bmatrix} Z_i^D & 0 \\ 0 & Z_i^L \end{bmatrix} \quad (12)$$

and

$$A^S = \left( \frac{1}{N} \sum_{i=1}^N Z_i^{S'} \hat{u}_i^* \hat{u}_i^{*'} Z_i^S \right)^{-1}$$

with residuals from the one-step consistent estimator,  $\hat{u}_i^* = (\hat{\Delta u}'_i, \hat{u}'_i)'$ .



### 3 A BIAS-CORRECTED ESTIMATOR

#### 3.1 Bias Correction

In order to reduce the bias and the size distortion of the Wald statistic in small samples, we propose a correction method based on Kurozumi and Yamamoto (2000).

We estimate  $\alpha$  using observations for period ( $t = 3, \dots, T$ ). Let  $T$ , the sample size, be an even integer. We define the bias-corrected estimator  $\hat{\alpha}^{BC}$  as follows:

$$\hat{\alpha}^{BC} = 2\tilde{\alpha} - \frac{1}{2}(\tilde{\alpha}_I + \tilde{\alpha}_{II}), \quad (13)$$

where  $\tilde{\alpha}$  is a GMM estimator for the whole period and  $\tilde{\alpha}_I$  and  $\tilde{\alpha}_{II}$  are estimators based on a sample of the first period ( $t = 3, \dots, T/2 + 1$ ) and of the second period ( $t = T/2 + 2, \dots, T$ ), respectively. Next, we define the asymptotic bias of  $\tilde{\alpha}$ ,  $ABIAS(\tilde{\alpha})$ , as  $E(\tilde{\alpha} - \alpha)$ , ignoring the terms whose orders are lower than  $O(\frac{1}{NT})$ .

*Proposition 1: If the asymptotic bias of  $\tilde{\alpha}$  follows*

$$ABIAS(\tilde{\alpha}) = \frac{c}{NT} + o(\frac{1}{NT}), \quad (14)$$

where  $c$  is a finite-valued constant independent of  $T$  and  $N$ , the bias-corrected estimator (13) has no asymptotic bias, i.e.,

$$ABIAS(\hat{\alpha}^{BC}) = o(\frac{1}{NT}).$$

*Proof:* We obtain the following relation:

$$\begin{aligned} ABIAS(\hat{\alpha}^{BC}) &= ABIAS(2\tilde{\alpha} - \frac{1}{2}(\tilde{\alpha}_I + \tilde{\alpha}_{II})) \\ &= \frac{2c}{NT} - \frac{1}{2} \frac{c}{N(T/2)} - \frac{1}{2} \frac{c}{N(T/2)} + o(\frac{1}{NT}) \\ &= o(\frac{1}{NT}). \end{aligned}$$

The second equality holds because observations of the first period and second period have sample sizes  $N$  and  $T/2$ . This completes the proof of Proposition 1.

*Remark 1:* It should be noted that Proposition 1 can be proved even if

$$ABIAS(\tilde{\alpha}) = \frac{c}{f(N)T}, \quad (15)$$

where  $f(N)$  is a function of  $N$ , which is less restrictive than (14). The proof, however, is the same as that described above. We can set  $f(N) = 1$ ,  $f(N) = 1/N$ , and so on. If we set  $f(N) = \sqrt{N}$ , then we obtain the above proposition. For notational convenience, we use (14) rather than (15) hereafter.

*Remark 2:* We note that when  $T$  is an odd integer, say  $T = 2m + 1$ , we can obtain the bias-corrected estimator as follows:

$$\hat{\alpha}^{BC} = 2\tilde{\alpha} - \frac{m}{2m+1}\tilde{\alpha}_I - \frac{m+1}{2m+1}\tilde{\alpha}_{II},$$

where  $\tilde{\alpha}$  is a GMM estimator for the whole period and  $\tilde{\alpha}_I$  and  $\tilde{\alpha}_{II}$  are estimators based on a sample of the first period ( $t = 3, \dots, m+1$ ) and of the second period ( $t = m+2, \dots, T$ ), respectively. The reason why we let  $T$  be an even integer is just the notational convenience.

*Remark 3:* We find that Proposition 1 can also be proved when  $\tilde{\alpha}_I$  is an estimator based on a sample of one half of  $N$  individuals and  $\tilde{\alpha}_{II}$  is based on a sample of the other half. Note, however, that there are many ways to divide  $N$  individuals into halves. For example, if  $N = 6$ , we can divide  $(1, 2, 3, 4, 5, 6)$  into  $(1, 2, 3)$  and  $(4, 5, 6)$ , but also into  $(1, 3, 5)$  and  $(2, 4, 6)$ . Obviously, there are many other ways. The bias-corrected estimator clearly depends on which division is used, but it is difficult to decide which division is the appropriate one. We therefore do not consider such a division of individuals although it is theoretically possible to construct the bias-corrected estimator based on such a division.

Proposition 1 above is the same as theorem 1(i) in Kurozumi and Yamamoto (2000) except that they consider a possibly cointegrated vector autoregression and eliminate the quasi-asymptotic bias<sup>1</sup> of the least squares estimator for the autoregressive parameter matrices. They point out that (13) is similar to the jackknife estimator.

In the following subsection, we construct an estimator whose asymptotic bias is given by (14).

### 3.2 The Bias-corrected Estimator and the Estimator of its Variance

The bias-corrected estimator (13) has no asymptotic bias as long as (14) holds. Unfortunately, the asymptotic biases of the GMM estimators introduced in Section 2 are not

---

<sup>1</sup>For a definition of quasi-asymptotic bias, see Kurozumi and Yamamoto (2000).

given by (14). Explicitly, the first differencing estimator, the level estimator and the system estimator have, as shown by Bun and Kiviet (2005), a bias of the order  $1/N$  and do not satisfy (14). We therefore propose an estimator whose bias is given by (14).

Consider a two-step system estimator with the particular instrument matrix  $Z$ , which includes a linear transformation of  $Z^S$  and at the same time reduces the number of columns of  $Z^S$ :

$$\hat{\alpha} = \frac{x^{*'} Z A Z' y^*}{x^{*'} Z A Z' x^*}, \quad (16)$$

where  $Z = (Z_1, \dots, Z_N)'$ ,

$$Z_i = \begin{bmatrix} \begin{pmatrix} y_{i1} \\ \vdots \\ y_{i,T-2} \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix} \end{bmatrix} \quad (17)$$

and

$$A = \left( \frac{1}{NT} \sum_{i=1}^N Z_i' \hat{u}_i^* \hat{u}_i^{*'} Z_i \right)^{-1}$$

with residuals from the one-step consistent estimator,  $\hat{u}_i^* = (\hat{\Delta u}_i', \hat{u}_i')'$ . The GMM estimator (16) is based on the moment conditions

$$E(Z_i' u_i^*) = 0, \quad (18)$$

where  $u_i^* = (\Delta u_i', u_i')'$ . We derive (18) by using (5) and (9). Before showing that the asymptotic bias of estimator (16) is given by (14), we note that (16) is a consistent estimator of  $\alpha$ .

*Proposition 2: Estimator (16) consistently estimates  $\alpha$  as  $N$  and/or  $T$  grow.*

*Proof:* See Appendix.

The above proposition shows that (16) is consistent both in terms of  $N$  and  $T$ .<sup>2</sup> Now, we proceed to derive the asymptotic bias of (16) based on large  $N$  and  $T$  asymptotics.

Assumption 2:

---

<sup>2</sup>It should be noted that the first differencing estimator (6) and the system estimator (11) are inconsistent when both  $N$  and  $T$  are large (see Alvarez and Arellano, 2003, and Hayakawa, 2006a).

- (i)  $E|\hat{\omega} - \omega|^{2k} \leq O((NT)^{-k/2})$ , where  $\omega = \text{plim } \hat{\omega}$  and  $k = 1, 2$ .
- (ii)  $E|\hat{\omega}^{-1}|^2$  is bounded for some  $N > N_0$  and  $T > T_0$ .
- (iii)  $E|(\sum_{i=1}^N Z_i' \hat{u}_i^* \hat{u}_i^{*'} Z_i / NT)^{-1}|^{2k}$  is bounded for some  $N > N_0$  and  $T > T_0$  for  $k = 1, 2$ , where  $P = ZAZ'$  and  $\hat{\omega} = x^{*'} P x^* / NT$ .

We may note that Assumptions 2(i) and 2(ii) hold in some pure time series models. Explicitly, it is known that if  $\eta_t$  follows a time series autoregressive model of order 1 with normal disturbance,

$$E|\hat{\lambda} - \lambda|^{2k} = O(T^{-k}), \text{ where } \hat{\lambda} = \sum_{t=1}^T \eta_t^2 / T, \lambda = \text{plim } \hat{\lambda} \text{ and } k \geq 1 \text{ is a fixed integer}$$

and

$$E|\hat{\lambda}^{-1}|^2 \text{ is bounded for some } T > T_0.$$

*Proposition 3: Under Assumption 2, the asymptotic bias of estimator (16) is given by (14).*

*Proof:* See Appendix.

Replacing  $\tilde{\alpha}$  of equation (13) by estimator (16), we obtain the bias-corrected estimator

$$\hat{\alpha}^{BC} = 2\hat{\alpha} - \frac{1}{2}(\hat{\alpha}_I + \hat{\alpha}_{II}), \tag{19}$$

where

$$\hat{\alpha}_s = \frac{x_s^{*'} Z_s A_s Z_s' y_s^*}{x_s^{*'} Z_s A_s Z_s' x_s^*}, \quad s = I, II$$

with obvious notations  $y_s^*$ ,  $x_s^*$ ,  $Z_s$  and  $A_s$ . For example,  $y_I^* = (\Delta x_I', x_I')'$  where  $x_I = (x_{I1}', \dots, x_{IN}')'$  with  $x_{Ii} = (y_{i3}, \dots, y_{i,T/2+1})'$  and

$$A_I = \left( \frac{1}{NT/2} \sum_{i=1}^N Z_{Ii}' \hat{u}_{Ii}^* \hat{u}_{Ii}^{*'} Z_{Ii} \right)^{-1},$$

where  $\hat{u}_{Ii}^* = (\hat{\Delta} u_{Ii}', \hat{u}_{Ii}')'$  and  $\hat{\Delta} u_{Ii}$  and  $\hat{u}_{Ii}$  are the residuals from the one-step consistent estimator in the first period. Propositions 1 and 2 show that bias-corrected estimator (19) is consistent, while Propositions 1 and 3 show that bias-corrected estimator (19) has no asymptotic bias.

We now turn to the reduction of the size distortion. As mentioned in the introduction, the Wald test based upon the two-step GMM procedure usually has large size distortion in small samples. It is well known that one reason why the Wald test exhibits considerable size distortion is the severe downward-bias of the estimated asymptotic variance of the two-step GMM estimator. We therefore introduce a small sample corrected estimator of the variance which is based on corollary 1 in Kurozumi and Yamamoto (2000) and is given by

$$\begin{aligned}\widehat{Var}(\hat{\alpha}^{BC}) = & 4\widehat{Var}(\hat{\alpha}) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_I) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_{II}) \\ & - 2\widehat{Var}(\hat{\alpha})x^{*'}ZAA_{wI}A_I Z_1'x_I^* \widehat{Var}(\hat{\alpha}_I) \\ & - 2\widehat{Var}(\hat{\alpha})x^{*'}ZAA_{wII}A_{II} Z_{II}'x_{II}^* \widehat{Var}(\hat{\alpha}_{II}) \\ & + \widehat{Var}(\hat{\alpha})^2 x_I' A_I A_r A_{II} Z_{II}' x_{II}^*,\end{aligned}\tag{20}$$

where

$$A_{wI} = \frac{1}{NT/2} \sum_{i=1}^N Z_i' \hat{u}_i^* \hat{u}_{Ii}^{*'} Z_{Ii}, \quad A_{wII} = \frac{1}{NT/2} \sum_{i=1}^N Z_i' \hat{u}_i^* \hat{u}_{IIi}^{*'} Z_{IIi},\tag{21}$$

$$A_r = \frac{1}{NT/2} \sum_{i=1}^N Z_{Ii}' \hat{u}_{Ii}^* \hat{u}_{IIi}^{*'} Z_{IIi},\tag{22}$$

$$\widehat{Var}(\hat{\alpha}) = \left( \frac{1}{NT} x^{*'} Z A Z' x^* \right)^{-1}\tag{23}$$

and

$$\widehat{Var}(\hat{\alpha}_s) = \left( \frac{1}{NT/2} x_s^{*'} Z_s A_s Z_s' x_s^* \right)^{-1}, \quad s = I, II.\tag{24}$$

*Proposition 4:* The limiting distribution of the bias corrected estimator (13) is normal and estimator (20) is a consistent estimator of its variance.

*Proof:* See Appendix.

Estimator (20) may appear complex, but, as mentioned by Kurozumi and Yamamoto (2000), it is straightforward from the definition of  $\hat{\alpha}^{BC}$ . Proposition 4 implies that the conventional Wald test statistic,  $W$ , for the null of  $\alpha = \alpha_0$ , has a chi-square distribution with one degree of freedom:

$$W = \frac{TN(\hat{\alpha}^{BC} - \alpha_0)^2}{\widehat{Var}(\hat{\alpha}^{BC})} \xrightarrow{d} \chi_1^2.\tag{25}$$

It should be emphasized that, unlike in Windmeijer’s (2005) approach, no preliminary estimate is needed to calculate the small sample corrected estimator (20). In fact, there are no preliminary estimates in (20).

Note that the bias correction method proposed in Section 3.1 is applicable to any estimator as long as its asymptotic bias is given by (14) or (15). In fact, we can similarly construct such estimators by arranging the instrument matrix for the first differencing estimator or for the level estimator (see, e.g., Bun and Kiviet, 2005). This means that based on a modified version of the first differencing estimator or the level estimator, we can construct bias-corrected estimators. However, in our Monte Carlo experiments<sup>3</sup>, bias-corrected estimator (19), which is based on a modified version of the system estimator, worked better than bias-corrected estimators based on a modified version of the first differencing estimator or the level estimator. We therefore do not consider bias-corrected estimators based on a modified version of the first differencing estimator or the level estimator.

## 4 AN ADDITIONAL INSTRUMENT

In this section, we propose an additional instrument in order to achieve a gain in efficiency. Under Assumption 1, it can be shown that the following moment condition holds:

$$E(Z_i^A u_i) = 0, \tag{26}$$

where

$$Z_i^A = \begin{bmatrix} 0 \\ \Delta^2 y_{i3} \\ \vdots \\ \Delta^2 y_{i,T-1} \end{bmatrix}. \tag{27}$$

---

<sup>3</sup>Because of space limitations, we omit the Monte Carlo results. They are available from the authors upon request.

To use (26) as the orthogonal condition in the GMM, we augment (17) with  $Z_i^A$  as

$$Z_i = \begin{bmatrix} \begin{pmatrix} y_{i1} \\ \vdots \\ y_{i,T-2} \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix} & \begin{pmatrix} 0 \\ \Delta^2 y_{i3} \\ \vdots \\ \Delta^2 y_{i,T-1} \end{pmatrix} \end{bmatrix}. \quad (28)$$

The bias of the GMM estimator (16) using instruments (28) instead of (17) is given by (14). The derivation of the bias, however, is the same as that in Proposition 3 and we therefore omit it here.

Next, we provide an illustration to check whether (27) suffers from the weak instrument problem. Suppose we estimate  $\alpha$  using the GMM procedure based only on the moment condition (27). Taking  $T = 4$  as an example, we obtain the auxiliary regression

$$y_{i3} = \pi \Delta^2 y_{i3} + r_i.$$

Under Assumption 1, we obtain

$$\hat{\pi} \xrightarrow{p} \frac{-\alpha^2 + 2\alpha - 3}{-2(\alpha - 2)(\alpha - 3)}. \quad (29)$$

Therefore, even if  $\alpha$  is close to unity,  $\text{plim } \hat{\pi} \neq 0$ . Hence, the instrument does not suffer from the weak instrument problem.

Before proceeding to the experiments, we summarize the bias correction procedure proposed in this paper: For the whole period, the first period and the second period, we calculate GMM estimator (16) using instruments (28). Then, we substitute these three GMM estimators, i.e., (16) for the whole period, the first period and the second period, into (19), which gives the bias-corrected estimator  $\hat{\alpha}^{BC}$ . The small sample corrected estimator of the variance of  $\hat{\alpha}^{BC}$  is given by (20). In order to reduce the size distortion, we calculate the Wald test statistic (25) which is based on (20).

## 5 EXPERIMENTS AND RESULTS

In this section, we carry out Monte Carlo experiments to investigate the small sample properties of the bias-corrected estimator (19) and the Wald test statistic (25).

## 5.1 A Process with No Exogenous Variable

### *The Monte Carlo Design*

In this subsection we employ the following data generating process (DGP):

$$y_{it} = \alpha y_{i,t-1} + \gamma_i + \varepsilon_{it},$$

where  $\alpha \in \{0, .3, .5, .8, .9\}$ ,  $\gamma_i \sim N(0, \sigma_\gamma^2)$  is independent across  $i$ ,  $\varepsilon_{it} \sim N(0, 1)$  is independent across  $i$  and  $t$ , and  $\sigma_\gamma^2 \in \{.25, 1, 4\}$ . We generate  $\gamma_i$  and  $\varepsilon_{it}$  such that they are independent of each other. The sample sizes are  $T \in \{6, 10, 20\}^4$  and  $N \in \{50, 100\}$ . We compute the bias-corrected estimator based on formulation (19) using instruments (28). The Wald test statistic is given by (25). For comparison, we compute the two-step system GMM estimator (11).<sup>5</sup> When we calculate the Wald test statistic based on the two-step system GMM estimator, we apply Windmeijer's (2005) correction<sup>6</sup> to the estimated variance of the two-step system GMM estimator in order to avoid a downward bias of the estimated variance. Throughout the experiments, 1000 samples of size  $N$  and  $T + 50$  were generated with the last  $T$  observations used for estimation and testing purposes. The nominal size of the Wald test was set to 5%.

### *The Monte Carlo Results*

Table 1 presents the finite sample properties of the system estimator and the bias-corrected estimator in the case of  $\sigma_\gamma^2 = 1$ . We can see that the system estimator has nonnegligible bias. The mean of the system estimator, i.e.  $\sum_{r=1}^{1000} \hat{\alpha}_r^S / 1000$ , departs from the true value of  $\alpha$  especially when  $\alpha$  is moderately large and/or the sample size  $N$  is small. This experimental result is consistent with the analytical result of Bun and Kiviet (2005). The bias of the bias-corrected estimator is smaller than that of the system estimator. Also, the size distortions of the Wald test statistic based on the bias-corrected

---

<sup>4</sup>Because we divide the whole period ( $t = 3, \dots, T$ ) into a first period ( $t = 3, \dots, T/2 + 1$ ) and a second period ( $t = T/2 + 2, \dots, T$ ),  $T$  has to be at least 6. If  $T < 6$ , a possible way to obtain the bias-corrected estimator is to divide all individuals ( $i = 1, \dots, N$ ) into two groups, as mentioned in Remark 3.

<sup>5</sup>When  $T = 20$ , we compute the one-step system GMM estimator because we often fail to compute  $A^S$ . That is, we often fail to invert  $\frac{1}{N} \sum_{i=1}^N Z_i^{S'} \hat{u}_i^* \hat{u}_i^{*'} Z_i^S$  numerically in the case of large  $T$  and small  $N$ .

<sup>6</sup>Windmeijer (2005) proposes a finite sample correction for the estimated variance of the two-step GMM estimator based on the estimate of the difference between the finite sample and the asymptotic variance.



estimator are moderate in comparison with those based on the system estimator with Windmeijer's (2005) correction. On the other hand, the standard deviation, which is a measure of efficiency, of the system estimator is smaller than that of the bias-corrected estimator. This may be due to the number of moment conditions. In fact, the number of moment conditions used in the system estimator is  $(T - 1)(T - 2)/2 + (T - 2)$ , whereas that used in the bias-corrected estimator is 3. Because the GMM estimators are known to show efficiency gains as the number of instruments increases, the standard deviation of the system estimator is smaller than that of the bias-corrected estimator. However, there does not seem to be much of a difference between these standard deviations unless  $\alpha$  is close to unity.

Further, Table 1 indicates that the bias and the size distortion of the system estimator become large as  $T$  becomes large. For example, when  $\alpha = 0.3$ ,  $N = 50$  and  $T = 6$ , the mean of  $\hat{\alpha}^S$  is 0.3 and the empirical size is 0.09. When  $T = 20$ , however, the mean is 0.14 and the empirical size is 0.97. This experimental result suggests that the system estimator performs well only when  $T$  is moderately small. One reason why the system estimator has large bias in the case of large  $T$  may be that the system estimator is inconsistent when both  $N$  and  $T$  tend to infinity, as shown by Hayakawa (2006a). In contrast, the bias-corrected estimator performs better as  $T$  becomes large. This is because the asymptotic bias of the bias-corrected estimator is  $o(\frac{1}{NT})$  as shown in Proposition 1.

Tables 2 and 3 show the finite sample properties when  $\sigma_\gamma^2 = 0.25$  and 4, respectively. These tables indicate that the bias and the size distortion of the system estimator become worse than in the case of  $\sigma_\gamma^2 = 1$ , which is consistent with the analytical result obtained by Bun and Kiviet (2005). In contrast, the results of the bias-corrected estimator are not very different from those in the case of  $\sigma_\gamma^2 = 1$  except for the case of  $\alpha = 0.9$ .

In order to evaluate the effect of using the additional instrument introduced in Section 4, the column "s.d.<sup>A</sup>" in Table 3 provides the standard deviation of the bias-corrected estimator using the instruments (17) instead of (28). Because the instruments (17) do not include the additional instrument, the comparison of the "s.d." of the bias-corrected estimator and "s.d.<sup>A</sup>" allows us to evaluate the effect of the additional instrument. Table 3 indicates that when we do not use the additional instrument, the standard deviation of the bias-corrected estimator becomes considerably large when  $\alpha = 0.8$  or 0.9. This means that the additional instrument helps to improve the efficiency of the bias-corrected estimator.

Table 1: The numerical performance for the case of  $\sigma_\gamma^2 = 1$

$\alpha$	N	T	System Estimator			BC Estimator		
			mean	s.d.	size	mean	s.d.	size
0.0	50	6	0.00	0.10	0.10	0.00	0.11	0.06
		10	-0.04	0.07	0.26	0.00	0.07	0.06
		20	-0.11	0.04	0.88	0.00	0.05	0.06
	100	6	0.00	0.06	0.07	-0.01	0.08	0.06
		10	-0.01	0.05	0.12	0.00	0.05	0.05
		20	-0.05	0.03	0.70	0.00	0.03	0.05
0.3	50	6	0.30	0.11	0.09	0.29	0.14	0.07
		10	0.23	0.08	0.30	0.30	0.09	0.05
		20	0.14	0.05	0.97	0.30	0.05	0.07
	100	6	0.31	0.07	0.08	0.29	0.10	0.06
		10	0.28	0.05	0.13	0.30	0.06	0.05
		20	0.22	0.04	0.84	0.30	0.03	0.05
0.5	50	6	0.50	0.12	0.10	0.49	0.16	0.07
		10	0.43	0.09	0.28	0.50	0.10	0.06
		20	0.32	0.05	0.99	0.50	0.06	0.07
	100	6	0.51	0.08	0.09	0.49	0.11	0.06
		10	0.47	0.06	0.14	0.50	0.06	0.05
		20	0.40	0.04	0.88	0.50	0.04	0.05
0.8	50	6	0.83	0.12	0.16	0.80	0.22	0.08
		10	0.78	0.08	0.15	0.79	0.13	0.08
		20	0.68	0.06	0.82	0.80	0.07	0.08
	100	6	0.82	0.09	0.18	0.79	0.15	0.08
		10	0.80	0.06	0.12	0.80	0.08	0.06
		20	0.73	0.05	0.60	0.80	0.05	0.05
0.9	50	6	0.94	0.08	0.21	0.92	0.26	0.07
		10	0.92	0.05	0.22	0.89	0.18	0.09
		20	0.88	0.04	0.23	0.89	0.10	0.10
	100	6	0.94	0.08	0.24	0.90	0.20	0.09
		10	0.93	0.04	0.24	0.89	0.12	0.07
		20	0.90	0.03	0.20	0.90	0.06	0.09

Note: The headings “System Estimator” and “BC Estimator” stand for the two-step system GMM estimator and the bias-corrected estimator given by (19). The column “mean” reports the sample mean of the estimators, i.e.,  $\sum_{r=1}^{1000} \hat{\alpha}_r / 1000$ . The column “s.d.” reports the standard deviation of the estimators, i.e.,  $(\sum_{r=1}^{1000} \hat{\alpha}_r^2 / 1000 - (\sum_{r=1}^{1000} \hat{\alpha}_r / 1000)^2)^{1/2}$ . The column “size” reports the empirical size of the Wald test for the null hypothesis  $\alpha = 0, .3, .5, .8, .9$ , at a significance level of 5%. Simulations are based on 1000 replications. The individual effects  $\gamma_i$  and the disturbance  $\varepsilon_{it}$  are generated from  $i.i.d.N(0, 1)$ .

Table 2: The numerical performance for the case of  $\sigma_\gamma^2 = 0.25$

$\alpha$	N	T	System Estimator			BC Estimator		
			mean	s.d.	size	mean	s.d.	size
0.0	50	6	-0.01	0.10	0.10	0.00	0.11	0.06
		10	-0.05	0.06	0.28	0.00	0.07	0.06
		20	-0.11	0.04	0.91	0.00	0.04	0.06
	100	6	0.00	0.06	0.08	0.00	0.07	0.06
		10	-0.02	0.05	0.12	0.00	0.05	0.06
		20	-0.06	0.03	0.75	0.00	0.03	0.05
0.3	50	6	0.29	0.11	0.09	0.30	0.13	0.07
		10	0.22	0.08	0.37	0.30	0.08	0.07
		20	0.13	0.05	0.99	0.30	0.05	0.07
	100	6	0.30	0.07	0.08	0.30	0.09	0.06
		10	0.27	0.05	0.14	0.30	0.05	0.05
		20	0.21	0.04	0.89	0.30	0.03	0.05
0.5	50	6	0.48	0.11	0.09	0.49	0.14	0.07
		10	0.40	0.08	0.40	0.50	0.08	0.06
		20	0.28	0.05	1.00	0.50	0.05	0.06
	100	6	0.50	0.08	0.08	0.50	0.09	0.06
		10	0.46	0.05	0.17	0.50	0.06	0.05
		20	0.38	0.04	0.95	0.50	0.03	0.05
0.8	50	6	0.76	0.12	0.11	0.79	0.17	0.07
		10	0.68	0.09	0.41	0.80	0.10	0.07
		20	0.55	0.06	1.00	0.80	0.06	0.05
	100	6	0.78	0.09	0.11	0.79	0.12	0.06
		10	0.75	0.06	0.17	0.80	0.06	0.06
		20	0.64	0.05	0.98	0.80	0.04	0.05
0.9	50	6	0.88	0.11	0.09	0.89	0.21	0.05
		10	0.83	0.08	0.23	0.90	0.13	0.07
		20	0.74	0.06	0.97	0.90	0.07	0.07
	100	6	0.89	0.09	0.10	0.89	0.15	0.06
		10	0.86	0.06	0.12	0.90	0.08	0.05
		20	0.79	0.05	0.90	0.90	0.04	0.06

Note: The experiment design is the same as that described in Table 1 except that the individual effects  $\gamma_i$  are generated from  $i.i.d.N(0, 0.25)$ .

Table 3: The numerical performance for the case of  $\sigma_\gamma^2 = 4$

$\alpha$	N	T	System Estimator			BC Estimator			
			mean	s.d.	size	mean	s.d.	size	s.d. <sup>A</sup>
0.0	50	6	0.02	0.11	0.13	0.00	0.13	0.06	0.13
		10	-0.01	0.07	0.22	0.00	0.08	0.05	0.08
		20	-0.08	0.04	0.70	0.00	0.05	0.06	0.05
	100	6	0.01	0.07	0.09	0.00	0.09	0.06	0.09
		10	0.00	0.05	0.12	0.00	0.06	0.05	0.06
		20	-0.04	0.03	0.49	0.00	0.03	0.05	0.04
0.3	50	6	0.35	0.13	0.15	0.30	0.17	0.07	0.17
		10	0.30	0.09	0.22	0.30	0.10	0.05	0.10
		20	0.21	0.05	0.66	0.30	0.06	0.06	0.06
	100	6	0.33	0.09	0.11	0.29	0.12	0.06	0.12
		10	0.31	0.06	0.16	0.30	0.07	0.04	0.07
		20	0.26	0.04	0.45	0.30	0.04	0.04	0.04
0.5	50	6	0.59	0.14	0.23	0.50	0.21	0.08	0.22
		10	0.54	0.09	0.28	0.50	0.12	0.06	0.12
		20	0.44	0.06	0.50	0.50	0.07	0.07	0.07
	100	6	0.56	0.10	0.19	0.50	0.15	0.08	0.15
		10	0.53	0.07	0.23	0.50	0.08	0.05	0.08
		20	0.48	0.05	0.35	0.50	0.05	0.04	0.05
0.8	50	6	0.92	0.09	0.55	0.86	0.29	0.14	0.39
		10	0.90	0.06	0.71	0.80	0.20	0.10	0.21
		20	0.85	0.04	0.61	0.79	0.11	0.10	0.11
	100	6	0.91	0.08	0.51	0.82	0.25	0.13	0.29
		10	0.89	0.05	0.68	0.79	0.13	0.08	0.14
		20	0.86	0.04	0.75	0.80	0.07	0.06	0.07
0.9	50	6	0.98	0.05	0.65	0.97	0.26	0.20	0.54
		10	0.98	0.02	0.88	0.93	0.24	0.17	0.32
		20	0.96	0.02	0.96	0.89	0.16	0.14	0.17
	100	6	0.98	0.05	0.64	0.95	0.24	0.19	0.45
		10	0.97	0.02	0.86	0.91	0.18	0.15	0.22
		20	0.97	0.02	0.98	0.90	0.10	0.15	0.10

Note: The column “s.d.<sup>A</sup>” reports the standard deviation of the bias-corrected estimator using instruments (17) instead of (28). The experiment design is the same as that described in Table 1 except that the individual effects  $\gamma_i$  are generated from  $i.i.d.N(0, 4)$ .

## 5.2 A Process with an Exogenous Variable

### *The Monte Carlo Design*

In this subsection, we employ a DGP that includes an exogenous variable and is specified as follows

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta w_{it} + \gamma_i + \varepsilon_{it} \\ w_{it} &= \rho w_{i,t-1} + \tau \gamma_i + v_{it}, \end{aligned}$$

where  $\rho$  is a scalar with  $|\rho| < 1$  and  $v_{it}$  is independently distributed across  $i$  and  $t$ . Including  $\tau \gamma_i$  in the DGP of  $w_{it}$ , we allow a correlation between the exogenous variable  $w_{it}$  and the individual effect  $\gamma_i$ . We find that there are moment conditions

$$E(Z_i' u_i^*) = 0, \tag{30}$$

where

$$Z_i = \begin{bmatrix} \begin{pmatrix} y_{i1} \\ \vdots \\ y_{i,T-2} \end{pmatrix} & \begin{pmatrix} w_{i2} \\ \vdots \\ w_{i,T-1} \end{pmatrix} & 0 & & & \\ 0 & & \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix} & \begin{pmatrix} \Delta w_{i3} \\ \vdots \\ \Delta w_{i,T} \end{pmatrix} & \begin{pmatrix} 0 \\ \Delta^2 y_{i3} \\ \vdots \\ \Delta^2 y_{i,T-1} \end{pmatrix} & \begin{pmatrix} \Delta^2 w_{i3} \\ \Delta^2 w_{i4} \\ \vdots \\ \Delta^2 w_{i,T} \end{pmatrix} \end{bmatrix}.$$

We compute the bias-corrected estimator based on formulation (19) using the above instruments instead of (28). Similarly, we augment the instruments (12) with  $w_{it}$  when we compute the system estimator. It should be noted that moment conditions (30) are weaker than the assumption of strict exogeneity of  $w_{it}$ .

We set  $\beta = 1$ ,  $\rho = 0.5$ ,  $\tau = 0.1$  and  $\nu_{it} \sim i.i.d.N(0, 1)$ . We set  $\alpha$ ,  $\gamma_i$ ,  $\varepsilon_{it}$ ,  $N$  and  $T$  to be the same as in Section 5.1.

### *The Monte Carlo Results*

Table 4 shows the experimental result when  $\sigma_\gamma^2 = 1$ . Comparing Table 4 with Table 1, we can see that when there is an exogenous variable, the size distortion of the system estimator of  $\alpha$  becomes worse in some cases. In contrast, the bias and the size distortion of the bias-corrected estimator are sufficiently small when an exogenous variable is included

in the DGP. The standard deviations of both the system estimator and the bias-corrected estimator get smaller than in the case where there is no exogenous variable. For  $\beta$ , the bias-corrected estimator has almost no bias and size distortion, while the system estimator suffers from size distortion.

As in the case where the DGP does not include an exogenous variable, the bias and the size distortion of the system estimator increase with  $T$ . In contrast, the bias-corrected estimator performs well regardless of  $T$ .

Tables 5 and 6 show the experimental results when  $\sigma_\gamma^2 = 0.25$  and 4, respectively. Tables 4, 5 and 6 indicate that the numerical performance of the system estimator heavily depends on  $\sigma_\gamma^2$ . In particular, when  $\sigma_\gamma^2 = 4$ , the system estimator of  $\alpha$  has substantial upward bias and large size distortion. In contrast, the bias and size distortion of the bias-corrected estimator is moderate even when  $\sigma_\gamma^2 = 4$ .

Table 4: The numerical performance for the case of  $\sigma_\gamma^2 = 1$

$\alpha$	N	T	System Estimator				BC Estimator							
			mean( $\alpha$ )	mean( $\beta$ )	s.d.( $\alpha$ )	s.d.( $\beta$ )	size( $\alpha$ )	size( $\beta$ )	mean( $\alpha$ )	mean( $\beta$ )	s.d.( $\alpha$ )	s.d.( $\beta$ )	size( $\alpha$ )	size( $\beta$ )
0.0	50	6	0.00	1.01	0.07	0.10	0.12	0.10	0.00	0.99	0.08	0.12	0.07	0.06
		10	-0.03	0.99	0.05	0.07	0.09	0.04	0.00	1.00	0.05	0.08	0.06	0.07
	100	20	-0.09	0.99	0.03	0.04	0.97	0.27	0.00	1.00	0.03	0.05	0.05	0.06
		6	0.00	1.00	0.04	0.07	0.07	0.06	0.00	1.00	0.05	0.08	0.06	0.05
		10	-0.02	1.00	0.03	0.05	0.24	0.18	0.00	1.00	0.04	0.05	0.06	0.05
		20	-0.05	0.99	0.02	0.03	0.84	0.25	0.00	1.00	0.02	0.03	0.07	0.06
0.3	50	6	0.30	1.01	0.07	0.10	0.12	0.11	0.30	0.99	0.08	0.12	0.06	0.06
		10	0.27	0.99	0.05	0.07	0.10	0.04	0.30	1.00	0.05	0.08	0.06	0.08
	100	20	0.19	0.98	0.03	0.04	0.99	0.30	0.30	1.00	0.03	0.05	0.05	0.06
		6	0.30	1.00	0.05	0.07	0.08	0.06	0.30	1.00	0.05	0.08	0.05	0.05
		10	0.28	1.00	0.03	0.05	0.21	0.17	0.30	1.00	0.03	0.05	0.06	0.06
		20	0.24	0.99	0.02	0.03	0.88	0.28	0.30	1.00	0.02	0.03	0.06	0.06
0.5	50	6	0.50	1.01	0.07	0.10	0.12	0.11	0.50	0.99	0.08	0.12	0.05	0.06
		10	0.47	0.99	0.05	0.07	0.09	0.04	0.50	1.00	0.05	0.08	0.07	0.08
	100	20	0.40	0.98	0.03	0.04	0.98	0.33	0.50	1.00	0.03	0.05	0.06	0.06
		6	0.51	1.00	0.05	0.07	0.08	0.06	0.50	1.00	0.05	0.08	0.05	0.05
		10	0.49	1.00	0.03	0.05	0.18	0.18	0.50	1.00	0.03	0.05	0.06	0.06
		20	0.44	0.98	0.02	0.03	0.87	0.29	0.50	1.00	0.02	0.03	0.06	0.06
0.8	50	6	0.83	1.02	0.06	0.10	0.21	0.11	0.80	1.00	0.08	0.12	0.07	0.07
		10	0.80	1.00	0.10	0.21	0.05	0.04	0.80	1.00	0.05	0.08	0.07	0.08
	100	20	0.75	0.99	0.03	0.04	0.70	0.27	0.80	1.00	0.03	0.05	0.07	0.06
		6	0.82	1.01	0.04	0.07	0.16	0.06	0.80	1.00	0.05	0.08	0.06	0.05
		10	0.81	1.00	0.03	0.05	0.17	0.20	0.80	1.00	0.03	0.05	0.06	0.05
		20	0.78	0.99	0.02	0.03	0.46	0.26	0.80	1.00	0.02	0.03	0.06	0.05
0.9	50	6	0.93	1.02	0.04	0.09	0.34	0.10	0.91	1.00	0.08	0.12	0.08	0.07
		10	0.92	1.01	0.03	0.06	0.36	0.24	0.90	1.01	0.05	0.08	0.08	0.08
	100	20	0.90	1.00	0.02	0.04	0.25	0.24	0.90	1.00	0.03	0.05	0.08	0.06
		6	0.93	1.01	0.03	0.07	0.34	0.07	0.90	1.00	0.05	0.08	0.07	0.06
		10	0.92	1.01	0.02	0.05	0.39	0.18	0.90	1.00	0.03	0.05	0.07	0.06
		20	0.91	1.00	0.02	0.03	0.29	0.25	0.90	1.00	0.02	0.03	0.06	0.06

Note: The experiment design is the same as that described in Table 1 except that an exogenous variable,  $w_{it}$ , is included in the DGP.

Table 5: The numerical performance for the case of  $\sigma_\gamma^2 = 0.25$

$\alpha$	N	T	System Estimator			BC Estimator							
			mean( $\alpha$ )	s.d.( $\alpha$ )	size( $\beta$ )	mean( $\alpha$ )	s.d.( $\alpha$ )	size( $\beta$ )					
0.0	50	6	-0.01	0.07	0.10	0.12	0.10	0.00	0.99	0.07	0.11	0.07	0.06
		10	-0.04	0.05	0.07	0.12	0.04	0.00	1.00	0.05	0.08	0.07	0.08
	100	20	-0.10	0.03	0.04	0.99	0.28	0.00	1.00	0.03	0.05	0.05	0.06
		6	0.00	0.04	0.07	0.07	0.06	0.00	1.00	0.05	0.07	0.06	0.05
	10	10	-0.02	0.03	0.05	0.26	0.17	0.00	1.00	0.03	0.05	0.08	0.05
		20	-0.05	0.02	0.03	0.90	0.27	0.00	1.00	0.02	0.03	0.06	0.06
0.3	50	6	0.28	0.07	0.10	0.11	0.11	0.30	0.99	0.07	0.11	0.05	0.06
		10	0.26	0.05	0.07	0.15	0.04	0.30	1.00	0.05	0.07	0.07	0.08
	100	20	0.18	0.03	0.04	1.00	0.33	0.30	1.00	0.03	0.05	0.06	0.06
		6	0.30	0.04	0.07	0.06	0.06	0.30	1.00	0.05	0.07	0.07	0.05
	10	10	0.27	0.03	0.05	0.27	0.18	0.30	1.00	0.03	0.05	0.07	0.06
		20	0.23	0.02	0.03	0.95	0.31	0.30	1.00	0.02	0.03	0.06	0.05
0.5	50	6	0.48	0.07	0.09	0.11	0.11	0.50	0.99	0.07	0.11	0.05	0.07
		10	0.45	0.05	0.07	0.17	0.04	0.50	1.00	0.05	0.07	0.08	0.08
	100	20	0.37	0.03	0.04	1.00	0.38	0.50	1.00	0.03	0.04	0.06	0.06
		6	0.50	0.04	0.07	0.06	0.06	0.50	1.00	0.05	0.07	0.06	0.05
	10	10	0.47	0.03	0.04	0.27	0.18	0.50	1.00	0.03	0.05	0.07	0.06
		20	0.43	0.02	0.03	0.97	0.35	0.50	1.00	0.02	0.03	0.06	0.05
0.8	50	6	0.79	0.06	0.09	0.11	0.10	0.80	0.99	0.06	0.11	0.06	0.08
		10	0.77	0.04	0.07	0.12	0.05	0.80	1.00	0.04	0.07	0.07	0.08
	100	20	0.70	0.03	0.04	1.00	0.40	0.80	1.00	0.02	0.04	0.07	0.07
		6	0.80	0.04	0.06	0.08	0.06	0.80	1.00	0.04	0.07	0.05	0.05
	10	10	0.78	0.03	0.04	0.20	0.18	0.80	1.00	0.03	0.04	0.06	0.06
		20	0.74	0.02	0.03	0.95	0.36	0.80	1.00	0.02	0.03	0.07	0.05
0.9	50	6	0.90	0.05	0.09	0.12	0.09	0.90	0.99	0.06	0.11	0.05	0.07
		10	0.89	0.17	0.29	0.07	0.10	0.90	1.00	0.04	0.07	0.07	0.08
	100	20	0.84	0.03	0.04	0.93	0.32	0.90	1.00	0.02	0.04	0.07	0.06
		6	0.91	0.03	0.06	0.10	0.06	0.90	1.00	0.04	0.07	0.05	0.06
	10	10	0.89	0.02	0.05	0.12	0.18	0.90	1.00	0.03	0.04	0.07	0.05
		20	0.86	0.02	0.03	0.78	0.29	0.90	1.00	0.02	0.03	0.07	0.05

Note: The experiment design is the same as that described in Table 2 except that an exogenous variable,  $w_{it}$ , is included in the DGP.



Table 6: The numerical performance for the case of  $\sigma_\gamma^2 = 4$

$\alpha$	N	T	System Estimator				BC Estimator							
			mean( $\alpha$ )	mean( $\beta$ )	s.d.( $\alpha$ )	s.d.( $\beta$ )	size( $\alpha$ )	size( $\beta$ )	mean( $\alpha$ )	mean( $\beta$ )	s.d.( $\alpha$ )	s.d.( $\beta$ )	size( $\alpha$ )	size( $\beta$ )
0.0	50	6	0.03	1.02	0.08	0.11	0.18	0.12	0.00	1.00	0.09	0.14	0.07	0.06
		10	0.00	1.00	0.05	0.08	0.07	0.04	0.00	1.00	0.06	0.09	0.06	0.08
	100	6	-0.06	1.00	0.03	0.05	0.77	0.27	0.00	1.00	0.03	0.05	0.06	0.06
		10	0.01	1.01	0.05	0.08	0.10	0.07	0.00	1.00	0.06	0.09	0.06	0.05
		6	0.00	1.01	0.04	0.05	0.20	0.21	0.00	1.00	0.04	0.06	0.05	0.04
		10	-0.03	1.00	0.03	0.03	0.55	0.26	0.00	1.00	0.02	0.03	0.06	0.05
0.3	50	6	0.35	1.03	0.08	0.11	0.23	0.12	0.29	1.00	0.10	0.14	0.06	0.05
		10	0.32	1.01	0.06	0.08	0.09	0.04	0.30	1.00	0.06	0.09	0.06	0.08
	100	6	0.25	1.00	0.04	0.05	0.63	0.27	0.30	1.00	0.03	0.05	0.05	0.06
		10	0.33	1.01	0.05	0.08	0.12	0.07	0.29	1.01	0.07	0.09	0.05	0.05
		6	0.32	1.01	0.04	0.05	0.25	0.23	0.30	1.00	0.04	0.06	0.05	0.05
		10	0.28	1.00	0.03	0.03	0.43	0.27	0.30	1.00	0.02	0.03	0.06	0.06
0.5	50	6	0.57	1.03	0.08	0.11	0.32	0.12	0.49	1.00	0.10	0.14	0.07	0.05
		10	0.54	1.01	0.06	0.08	0.17	0.05	0.50	1.00	0.07	0.09	0.07	0.08
	100	6	0.48	1.01	0.04	0.05	0.42	0.28	0.50	1.00	0.03	0.05	0.05	0.06
		10	0.54	1.01	0.06	0.08	0.18	0.07	0.49	1.01	0.07	0.09	0.06	0.05
		6	0.53	1.02	0.04	0.05	0.39	0.24	0.50	1.00	0.04	0.06	0.05	0.04
		10	0.50	1.01	0.03	0.03	0.34	0.29	0.50	1.00	0.02	0.03	0.05	0.05
0.8	50	6	0.89	1.04	0.05	0.10	0.66	0.12	0.82	1.01	0.11	0.14	0.14	0.07
		10	0.87	1.03	0.03	0.07	0.84	0.31	0.79	1.01	0.07	0.09	0.09	0.08
	100	6	0.85	1.03	0.02	0.04	0.83	0.37	0.80	1.00	0.03	0.05	0.08	0.06
		10	0.87	1.03	0.04	0.07	0.54	0.08	0.80	1.01	0.07	0.09	0.10	0.06
		6	0.87	1.03	0.03	0.05	0.87	0.28	0.79	1.01	0.04	0.06	0.07	0.05
		10	0.85	1.02	0.02	0.03	0.89	0.38	0.80	1.00	0.02	0.03	0.05	0.05
0.9	50	6	0.96	1.03	0.02	0.09	0.84	0.11	0.93	1.01	0.08	0.14	0.21	0.07
		10	0.96	1.03	0.02	0.06	0.89	0.25	0.91	1.02	0.07	0.10	0.16	0.09
	100	6	0.95	1.03	0.01	0.04	1.00	0.38	0.90	1.00	0.03	0.05	0.11	0.08
		10	0.96	1.03	0.02	0.07	0.79	0.08	0.92	1.02	0.07	0.09	0.18	0.06
		6	0.96	1.03	0.01	0.05	0.98	0.25	0.90	1.01	0.05	0.06	0.12	0.06
		10	0.95	1.02	0.01	0.03	1.00	0.41	0.90	1.00	0.02	0.04	0.07	0.06

Note: The experiment design is the same as that described in Table 3 except an exogenous variable  $w_{it}$  is included in the DGP.

Table 7: Estimation results

N	T	System Estimator				BC Estimator			
		$\alpha$	$\beta$	s.e.( $\alpha$ )	s.e.( $\beta$ )	$\alpha$	$\beta$	s.e.( $\alpha$ )	s.e.( $\beta$ )
385	14	1.01	-0.41	0.05	0.09	0.78	-0.11	0.03	0.05

## 6 EMPIRICAL ILLUSTRATION

In this section, we provide an empirical illustration to demonstrate the application of the bias-corrected estimator. We consider a simple model for employment dynamics which was already used by Hayakawa (2006b) and is given by

$$n_{it} = \alpha n_{i,t-1} + \beta w_{it} + \gamma_i + \varepsilon_{it},$$

where  $n_{it}$  is the log of employment in firm  $i$  at time  $t$  and  $w_{it}$  is the log of wages.

We use a dataset that is taken from Arellano (2002). This dataset is a balanced panel of 385 Spanish firms consisting of annual data for the period 1983-1996 with  $T = 14$ . Following Arellano (2002), prior to the estimation, we remove time effects by using the deviations from period-specific cross-sectional means.

The estimation results are presented in Table 7, where “s.e.” indicates the standard error of each estimator. The standard error of the system estimator is calculated using Windmeijer’s (2005) method. The standard error of the bias-corrected estimator is given by (20). The results indicate that the system estimator is unstable in the sense that the estimate of  $\alpha$  exceeds unity, while the bias-corrected estimator gives a much smaller estimate. This result is consistent with our Monte Carlo experiments which showed that the system estimator was biased upward. Thus, the system estimator does not appear to perform well. In terms of standard error, there is no large difference between these estimators.

## 7 CONCLUDING REMARKS

In this paper, we proposed a bias-corrected estimator for the autoregressive parameter in dynamic panel data models. We constructed the bias-corrected estimator by applying a bias correction based on Kurozumi and Yamamoto (2000) to the GMM estimator which

is a modified version of Blundell and Bond's (1998) system estimator. The bias of our estimator is smaller than that of several commonly used GMM estimators. In order to reduce the size distortion of the Wald test, we introduced a small sample corrected estimator of the variance. Further, we proposed an additional instrument to achieve a gain in efficiency.

The Monte Carlo experiments indicated that the bias-corrected estimator outperforms the system estimator with Windmeijer's (2005) correction in terms of small sample bias and size distortion. It turned out that the additional instrument helps to improve the efficiency of the bias-corrected estimator. The bias-corrected estimator gains an advantage over the system estimator as  $N$  becomes small and/or  $\sigma_\gamma^2/\sigma_\varepsilon^2$  departs from unity. Further, we find that the bias and the size distortion of the system estimator become large as  $T$  becomes large, while those of the bias-corrected estimator become small as  $T$  becomes large. The empirical illustration yielded results corroborating that while the system estimator is severely biased, the bias-corrected estimator has only moderate bias.

Another advantage of the bias-corrected estimator and the small sample corrected estimator of the variance is that they do not require any preliminary estimates. The bias-corrected estimator is simply the linear combination of the GMM estimator for different sample periods and therefore can be easily calculated. In sum, the bias-corrected estimator has two attractive features: It outperforms the system estimator in terms of bias and size distortion in finite samples and is easy to calculate.

## References

- Alvarez, J. and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121-1159.
- Arellano, M. (2002): "Modelling Optimal Instrumental Variables for Dynamic Panel Data Models," CEMFI Working Paper No. 0310.
- 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.
- Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Component Models," *Journal of Econometrics*, 68, 29-45.
- Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.
- Blundell, R., Bond, S. and F. Windmeijer (2000): "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," IFS Working Papers W00/12, Institute for Fiscal Studies.
- Bond, S. and 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.
- Bun, M.J.G. and M.A. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23, 200-210.
- Bun, M.J.G. and 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.
- Hahn, J. and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both  $n$  and  $T$  Are Large," *Econometrica*, 70, 1639-1657.

- Hayakawa, K. (2006a): “The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T are Large ,” Hi-Stat Discussion Paper Series No.129, Institute of Economic Research, Hitotsubashi University.
- Hayakawa, K. (2006b): “Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present,” Hi-Stat Discussion Paper Series No.130, Institute of Economic Research, Hitotsubashi University.
- Kiviet, J.F. (1995): “On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models,” *Journal of Econometrics*, 68, 53-78.
- Kurozumi, E. and T. Yamamoto (2000): “Modified Lag Augmented Vector Autoregressions,” *Econometric Reviews*, 19, 207-231.
- Windmeijer, F (2005): “Finite Sample Correction for the Variance of Linear Two-step GMM Estimators,” *Journal of Econometrics*, 126, 25-51.
- Yamamoto, T. and N. Kunitomo (1984): “Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models,” *Annals of the Institute of Statistical Mathematics*, 36, 419-430.

## Proof of Proposition 2

To prove Proposition 2, we provide the following lemma:

*Lemma:* The sample analogue of the moment conditions (18) converges in probability to zero as  $N$  and  $T$  grow:

$$(Z'u^*/NT) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=2}^{T-1} \begin{bmatrix} y_{i,t-1} \Delta u_{it} \\ \Delta y_{it} u_{i,t+1} \end{bmatrix} \xrightarrow{p} 0,$$

where  $u^* = (u_1^*, \dots, u_N^*)'$ .

*Proof:* Because the expectation of the random variables  $y_{i,t-1} \Delta u_{it}$  and  $y_{i,t-1} \Delta u_{it}$  is zero, the ergodic theorem ensures that  $\text{plim}(Z'u^*/NT) = 0$ .

Using the above lemma, we obtain

$$\begin{aligned} \hat{\alpha} &= \frac{x^{*'} Z A Z' y^*}{x^{*'} Z A Z' x^*} \\ &= \alpha + \frac{x^{*'} Z A Z' u^*}{x^{*'} Z A Z' x^*} \\ &= \alpha + \frac{(x^{*'} Z / NT) A (Z' u^* / NT)}{(x^{*'} Z / NT) A (Z' x^* / NT)} \\ &\xrightarrow{p} \alpha + 0. \end{aligned} \tag{31}$$

## Proof of Proposition 3

Our proof that the asymptotic variance of (16) satisfies (14) is similar to Lemma 2 of Yamamoto and Kunitomo (1984). First, we express  $\hat{\alpha} - \alpha$  as

$$\begin{aligned} \hat{\alpha} - \alpha &= \frac{x^{*'} P u^* / NT}{\hat{\omega}} \\ &= \frac{x^{*'} P u^* / NT}{\omega \{1 + (\hat{\omega} / \omega - 1)\}} \\ &= (x^{*'} P u^* / NT) \left\{ \frac{1}{\omega} \left( 1 - \frac{\hat{\omega} - \omega}{\omega} \right) \right\} + (x^{*'} P u^* / NT) \left\{ \frac{1}{\hat{\omega}} \left( \frac{(\hat{\omega} - \omega)^2}{\omega^2} \right) \right\}. \end{aligned} \tag{32}$$

Assumption 2 and Schwarz's inequality ensure that the expectation of the second term of (32) is bounded above by the term  $O(\frac{1}{NT})$ . We therefore consider only the first term of (32):

$$(x^{*'} P u^* / NT) \left\{ \frac{1}{\omega} \left( 1 - \frac{\hat{\omega} - \omega}{\omega} \right) \right\} = \frac{2}{\omega} (x^{*'} P u^* / NT) - \frac{1}{\omega^2} (x^{*'} P u^* \hat{\omega} / NT).$$

We rewrite  $x^{*'}Pu^*$  as follows:

$$\begin{aligned} x^{*'}Pu^* &= \frac{x^{*'}Z}{\sqrt{NT}} \left( \sum_{i=1}^N Z_i \hat{u}_i^* \hat{u}_i^{*'} Z_i / NT \right)^{-1} \frac{Z'u^*}{\sqrt{NT}} \\ &= \begin{pmatrix} \frac{a_1}{\sqrt{NT}} & \frac{a_1}{\sqrt{NT}} \\ \frac{a_1}{\sqrt{NT}} & \frac{a_1}{\sqrt{NT}} \end{pmatrix} \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} \begin{pmatrix} \frac{c_1}{\sqrt{NT}} \\ \frac{c_2}{\sqrt{NT}} \end{pmatrix}, \end{aligned}$$

where  $a_1 = \sum_{i=1}^N \sum_{t=2}^{T-1} \Delta y_{it} y_{i,t-1}$ ,  $a_2 = \sum_{i=1}^N \sum_{t=2}^{T-1} y_{i,t+1} \Delta y_{it}$ ,  $b_1, \dots, b_4$  are the elements of

$$\left( \sum_{i=1}^N Z_i \hat{u}_i^* \hat{u}_i^{*'} Z_i / NT \right)^{-1}, \quad c_1 = \sum_{i=1}^N \sum_{t=2}^{T-1} y_{i,t-1} \Delta u_{i,t-1} \quad \text{and} \quad c_2 = \sum_{i=1}^N \sum_{t=2}^{T-1} \Delta y_{it} u_{i,t+1}.$$

Using Assumption 2 and Schwarz's inequality, we obtain

$$\begin{aligned} E|x^{*'}Pu^*| &= E|(a_1 b_1 c_1 + a_2 b_3 c_1 + a_1 b_2 c_2 + a_2 b_4 c_2) / NT| \\ &\leq \left( \sqrt{E(a_1^2 c_1^2) E(b_1^2)} + \sqrt{E(a_2^2 c_1^2) E(b_3^2)} + \sqrt{E(a_1^2 c_2^2) E(b_2^2)} + \sqrt{E(a_2^2 c_2^2) E(b_4^2)} \right) / NT \\ &\leq \sqrt{O((NT)^2)} / NT. \end{aligned}$$

Similarly we obtain  $E|x^{*'}Pu^* \hat{\omega}| \leq O(1)$ . This complete the proof of this proposition.

Proof of Proposition 4

Following the proof of Theorem 1 (ii) in Kurozumi and Yamamoto (2000), we first show that the limiting distribution of the bias-corrected estimator (13) is normal. Note that the bias corrected estimator (13) can be expressed as

$$\hat{\alpha}^{BC} = 2 \frac{x^{*'} Z A Z' y^*}{x^{*'} Z A Z' y^*} - \frac{1}{2} \left( \frac{x_I^{*'} Z_I A_I Z_I' y_I^*}{x_I^{*'} Z_I A_I Z_I' y_I^*} + \frac{x_{II}^{*'} Z_{II} A_{II} Z_{II}' y_{II}^*}{x_{II}^{*'} Z_{II} A_{II} Z_{II}' y_{II}^*} \right).$$

Thus, we obtain

$$\sqrt{NT}(\hat{\alpha}^{BC} - \alpha) = 2\sqrt{NT} \frac{x^{*'} Z A Z' u^*}{x^{*'} Z A Z' x^*} - \frac{\sqrt{2}}{2} \sqrt{\frac{NT}{2}} \left( \frac{x_I^{*'} Z_I A_I Z_I' u_I^*}{x_I^{*'} Z_I A_I Z_I' x_I^*} + \frac{x_{II}^{*'} Z_{II} A_{II} Z_{II}' u_{II}^*}{x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^*} \right).$$

We can express the above equation as follows:

$$\begin{aligned} &\sqrt{NT}(\hat{\alpha}^{BC} - \alpha) \\ &= \left[ \sqrt{2} \left( \frac{1}{NT} x^{*'} Z A Z' x^* \right)^{-1} \quad -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_I^{*'} Z_I A_I Z_I' x_I^* \right)^{-1} \quad -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^* \right)^{-1} \right] \\ &\quad \times \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{2}{NT}} x^{*'} Z A Z' u^* \\ \sqrt{\frac{2}{NT}} x_I^{*'} Z_I A_I Z_I' u_I^* \\ \sqrt{\frac{2}{NT}} x_{II}^{*'} Z_{II} A_{II} Z_{II}' u_{II}^* \end{bmatrix}. \end{aligned} \tag{33}$$

Because

$$\frac{1}{NT} x^{*'} Z A Z' x^* \xrightarrow{p} q, \quad \frac{2}{NT} x_I^{*'} Z_I A_I Z_I' x_I^* \xrightarrow{p} q, \quad \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^* \xrightarrow{p} q$$

and

$$\begin{bmatrix} \sqrt{\frac{2}{NT}}x^{*I}ZAZ_I'u_I^* \\ \sqrt{\frac{2}{NT}}x^{*II}ZAZ_{II}'u_{II}^* \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \xi_I \\ \xi_{II} \end{bmatrix} \sim N\left(0, \begin{bmatrix} q & r \\ r & q \end{bmatrix}\right), \quad (34)$$

we obtain

$$\begin{aligned} \sqrt{NT}(\hat{\alpha}^{BC} - \alpha) &\xrightarrow{d} \frac{\sqrt{2}}{2}q(\xi_I + \xi_{II}) \\ &\sim N(0, q^{-1} + q^{-2}r). \end{aligned} \quad (35)$$

Note that the remaining individual effect in  $u^*$  causes the correlation between  $u_I^*$  and  $u_{II}^*$ , resulting in  $r$ .

Next, we construct a consistent estimator of the variance of (35) in roughly the same way as corollary 1 in Kurozumi and Yamamoto (2000). Using (34), we can express the variance of the limiting distribution of (33), except for the first matrix, as follows:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} q & r \\ r & q \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2q & q & q \\ q & q & 0 \\ q & 0 & q \end{bmatrix} + \begin{bmatrix} 2r & r & r \\ r & 0 & r \\ r & r & 0 \end{bmatrix} = Q + R \quad (\text{say}).$$

We estimate parameter matrix  $Q$  as follows. Because the (1,1) element of  $Q$  is associated with the whole period, it will be estimated from the sample of the whole period. Similarly, the (2,2) and the (3,3) element will be estimated from the samples of the first period and the second period, respectively. Because the (1,2) element of  $Q$  is associated with the whole and the first period, it will be estimated from the samples of the whole and the first period. For the same reason, the (1,3) element will be estimated from the sample of the whole and the second period. In the same way, we estimate parameter matrix  $R$ .



Thus, the estimator of the variance is given by

$$\begin{aligned}
& \widehat{Var}(\hat{\alpha}^{BC}) \\
&= \left[ \sqrt{2} \left( \frac{1}{NT} x^{*'} Z A Z' x^* \right)^{-1} \quad -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_I^{*'} Z_I A_I Z_I' x_I^* \right)^{-1} \quad -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^* \right)^{-1} \right] \\
& \quad \left\{ \begin{aligned} & \left[ \begin{array}{ccc} \frac{2}{NT} x^{*'} Z A Z' x^* & \frac{2}{NT} x^{*'} Z A A_{wI} A_I Z_I' x_I^* & \frac{2}{NT} x^{*'} Z A A_{wII} A_{II} Z_{II}' x_{II}^* \\ \frac{2}{NT} x^{*'} Z A A_{wI} A_I Z_I' x_I^* & \frac{2}{NT} x_I^{*'} Z_I A_I Z_I' x_I^* & 0 \\ \frac{2}{NT} x^{*'} Z A A_{wII} A_{II} Z_{II}' x_{II}^* & 0 & \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^* \end{array} \right] \\ & + \left[ \begin{array}{ccc} \frac{2}{NT} x^{*'} Z A A_r A Z' x^* & \frac{2}{NT} x^{*'} Z A A_r A_I Z_I' x_I^* & \frac{2}{NT} x^{*'} Z A A_r A_{II} Z_{II}' x_{II}^* \\ \frac{2}{NT} x^{*'} Z A A_r A_I Z_I' x_I^* & 0 & \frac{2}{NT} x_I^{*'} Z_I A_I A_r A_{II} Z_{II}' x_{II}^* \\ \frac{2}{NT} x^{*'} Z A A_r A_{II} Z_{II}' x_{II}^* & \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} A_r A_I Z_I' x_I^* & 0 \end{array} \right] \end{aligned} \right\} \\
& \quad \left[ \begin{array}{c} \sqrt{2} \left( \frac{1}{NT} x^{*'} Z A Z' x^* \right)^{-1} \\ -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_I^{*'} Z_I A_I Z_I' x_I^* \right)^{-1} \\ -\frac{\sqrt{2}}{2} \left( \frac{2}{NT} x_{II}^{*'} Z_{II} A_{II} Z_{II}' x_{II}^* \right)^{-1} \end{array} \right] \\
&= 4\widehat{Var}(\hat{\alpha}) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_I) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_{II}) \tag{36} \\
& \quad - 2\widehat{Var}(\hat{\alpha}) x^{*'} Z A A_{wI} A_I Z_I' x_I^* \widehat{Var}(\hat{\alpha}_I) - 2\widehat{Var}(\hat{\alpha}) x^{*'} Z A A_{wII} A_{II} Z_{II}' x_{II}^* \widehat{Var}(\hat{\alpha}_{II}) \\
& \quad + \left\{ 4\widehat{Var}(\hat{\alpha})^2 x^{*'} Z A A_r A Z' x^* \right. \\
& \quad + \frac{1}{2}\widehat{Var}(\hat{\alpha}_I) x_I^{*'} Z_I A_I A_r A_{II} Z_{II}' x_{II}^* \widehat{Var}(\hat{\alpha}_{II}) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_{II}) x_{II}^{*'} Z_{II} A_{II} A_r A_I Z_I' x_I^* \widehat{Var}(\hat{\alpha}_I) \\
& \quad \left. - 2\widehat{Var}(\hat{\alpha}) x^{*'} Z A A_r A_I Z_I' x_I^* \widehat{Var}(\hat{\alpha}_I) - 2\widehat{Var}(\hat{\alpha}) x^{*'} Z A A_r A_{II} Z_{II}' x_{II}^* \widehat{Var}(\hat{\alpha}_{II}) \right\},
\end{aligned}$$

where  $A_{wI}$ ,  $A_{wII}$ ,  $A_r$  and  $\widehat{Var}(\hat{\alpha}_s)$ ,  $s = I, II$  are given in (21), (22), (23) and (24). However, calculating (36) needs a bit of effort. Thus, we replace the terms between braces in (36) by

$$\widehat{Var}(\hat{\alpha})^2 x_I^{*'} Z_I A_I A_r A_{II} Z_{II}' x_{II}^*, \tag{37}$$

which is asymptotically equivalent to the terms between the braces. There are alternative estimators other than (37), which are easy to compute and asymptotically equivalent to the terms between the braces, such as,  $\widehat{Var}(\hat{\alpha})^2 x_I^{*'} Z A A_r A Z' x^*$ . However, in our Monte Carlo experiments,<sup>7</sup> the Wald test statistic using these alternative estimators exhibits large size distortions. Therefore, we recommend to use (37). Consequently, we obtain

---

<sup>7</sup>The Monte Carlo results are available from authors upon request.

the corrected estimator of the variance

$$\begin{aligned}
\widehat{Var}(\hat{\alpha}^{BC}) = & 4\widehat{Var}(\hat{\alpha}) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_I) + \frac{1}{2}\widehat{Var}(\hat{\alpha}_{II}) \\
& - 2\widehat{Var}(\hat{\alpha})x^*ZAA_wIA_I Z'_1 x_I^* \widehat{Var}(\hat{\alpha}_I) \\
& - 2\widehat{Var}(\hat{\alpha})x^*ZAA_wIIA_{II} Z'_{II} x_{II}^* \widehat{Var}(\hat{\alpha}_{II}) \\
& + \widehat{Var}(\hat{\alpha})^2 x_I^* Z_I A_I A_r A_{II} Z'_{II} x_{II}^*.
\end{aligned}$$

We can see that  $\widehat{Var}(\hat{\alpha}^{BC}) \xrightarrow{p} q^{-1} + q^{-2}r$ . This completes the proof of this proposition.