Discussion Paper Series

No.130

Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present

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/

Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present

Kazuhiko Hayakawa*†

Graduate School of Economics

Hitotsubashi University

January 11, 2006

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

†I am deeply grateful to Taku Yamamoto, Katsuto Tanaka, Satoru Kanoh and seminar participants at Hitotsubashi University. I also gratefully acknowledge the financial support from the JSPS Fellowship. All remaining errors are mine.

Abstract

This paper addresses the many instruments problem, i.e. (1) the trade-off between the bias and the efficiency of the GMM estimator, and (2) inaccuracy of inference, in dynamic panel data models where unobservable heterogeneity may be large. We find that if we use all the instruments in levels, although the GMM estimator is robust to large heterogeneity, inference is inaccurate. In contrast, if we use the minimum number of instruments in levels in the sense that we use only one instrument for each period, the performance of the GMM estimator is heavily affected by the degree of heterogeneity, that is, both the asymptotic bias and the variance are proportional to the magnitude of heterogeneity. To address this problem, we propose a new form of instruments that are obtained from the so-called backward orthogonal deviation transformation. The asymptotic analysis shows that the GMM estimator with the minimum number of new instruments has smaller asymptotic bias than the estimators typically used such as the GMM estimator with all instruments in levels, the LIML estimators and the within-groups estimators, while the asymptotic variance of the proposed estimator is equal to the lower bound. Thus *both* the asymptotic bias and the variance of the proposed estimators become small simultaneously. Simulation results show that our new GMM estimator outperforms the conventional GMM estimator with all instruments in levels in term of the RMSE and in terms of accuracy of inference. An empirical application with Spanish firm data is also provided.

Keywords: Dynamic panel data, many instruments, generalized method of moments estimator, unobservable large heterogeneity.

JEL classification: C23.

1 Introduction

In cross-sectional data models, since the famous work of Angrist and Krueger (1991), the many instruments (MI) problem, i.e. (1) the trade-off between the bias and the efficiency of the two stage least squares (2SLS) estimator, and (2) inaccuracy of inference, has been intensively discussed, especially in connection with the weak instruments problem. For example, Bound, Jaeger and Baker (1995), Angrist, Imbens and Krueger (1999), Hahn and Inoue (2002), Hahn (2002), Hahn, Hausman and Kuersteiner (2004), Chao and Swanson (2005), Okui (2005b), Hansen, Hausman and Newey (2005), Anderson, Kunitomo and Matsushita (2005), and Andrews and Stock (2005) and the papers cited therein deal with this problem.[1]

Yet, while there are many studies on the MI problem in the context of cross sectional data models, little research has been done for the case of dynamic panel data models even though the MI problem also occurs in this type of model.[2] In fact, one of the important features of dynamic panel data models is that the number of available instruments increases as T, the dimension of the time series, gets larger.[3] One paper that deals with the MI problem in a dynamic panel model is Okui (2005b) which, based on Donald and Newey (2001) and Okui (2005a), develops a procedure to select the instruments so as to minimize the mean squared error (MSE) and improve the accuracy of inference. However, his method is computationally cumbersome and there still remain size distortions when α, an autoregressive parameter, is large. Furthermore, although Okui (2005b) does not pay much attention to the effects of large heterogeneity,[4] it is worth considering such effects, because in empirical analyses we may come across situations where heterogeneity is large. For example, Arellano (2002) set the ratio of the variance of the individual effects to the disturbances to be 9 in the simulation, where its simulation design was roughly calibrated to the real data of Bover and Watson (2004). The first purpose of the

[1] See also Kunitomo (1980), Morimune (1983) and Bekker (1994).

[2] An analysis of the MI problem in the context of static panel data models with predetermined variables is provided by Ziliak (1997).

[3] Since the MI problem becomes more serious when T is large, we focus on the case where T is greater than 10. The case when $T < 10$ is beyond the scope of the present paper.

[4] Throughout this paper, by "large heterogeneity" is meant that the variance of the unobservable individual effects is large relative to the variance of the disturbances.

present paper is to consider cases where heterogeneity is large and especially to consider the effects of large heterogeneity on generalized method of moments (GMM) estimators where instruments in levels are used. The second purpose is to suggest a way to overcome the drawbacks of Okui's method by proposing new instruments with which we can solve the MI problem even if heterogeneity is large.

The findings of this paper are as follows. If all the instruments in levels are used, although the GMM estimator is robust to large heterogeneity, the size distortion is substantial. In contrast, if we use the minimum number of instruments in levels, that is, only one instrument in each period, although the size is close to the nominal level, both the asymptotic bias and the variance are heavily affected by the degree of heterogeneity. These facts indicate that, as long as instruments in levels are used, we cannot obtain a GMM estimator with small bias and variance, and with less size distortion when heterogeneity is large. To overcome this problem, we consider the elimination of the individual effects from the instruments. Two methods are employed to remove the individual effects. The first is simply to take the first difference. The second method we propose is to use the backward orthogonal deviation (BOD) transformation. Asymptotic analysis shows that a GMM estimator with the minimum number of first-differenced instruments is no longer efficient, though it is robust to large heterogeneity. However, if we use the minimum number of instruments transformed by the BOD transformation, the GMM estimator is robust to the presence of large heterogeneity and has smaller asymptotic bias than the GMM estimator with all instruments in levels, the LIML estimator, and the within-groups estimator, while its asymptotic variance is equal to the efficiency bound. Thus *both* the asymptotic bias and variance of the proposed GMM estimator become small simultaneously. Furthermore, the simulation analysis shows that the size of the newly proposed GMM estimator is close to the nominal level.

The remainder of this paper is organized as follows. Section 2 provides the model and the basic GMM estimators. Section 3 considers the effect of large heterogeneity on the GMM estimator when all instruments in levels and the minimum number of instruments in levels are used. Section 4 considers the removal of the individual effects from the instruments and derives the asymptotic properties of the proposed GMM estimators. Section 5 reports the results of Monte Carlo simulations to assess the theoretical implications. Section 6 then applies the proposed estimator to the data of Bover and Watson (2004). Finally Section 7 concludes.

2 The model and the 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 $x_{it} = y_{i,t-1}$, $y_i = (y_{i,2}, ..., y_{i,T})'$, $x_i = (x_{i,2}, ..., x_{i,T})'$, $\iota_{T-1} = (1, ..., 1)'$ and $v_i = (v_{i,2}, ..., v_{i,T})$, (1) can be expressed in vector form as

$$y_i = \alpha x_i + \eta_i \iota_{T-1} + v_i \tag{2}$$

We impose the following assumptions which are the same as those in Alvarez and Arellano (2003).

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. *The initial observations satisfy*

$$y_{i1} = \frac{\eta_i}{1-\alpha} + w_{i1} \qquad\qquad for \quad i = 1, ..., N \tag{3}$$

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

Assumption 3. η_i *are i.i.d across individuals with* $E(\eta_i) = 0$, $var(\eta_i) = \sigma_\eta^2$, *and finite fourth order moment.*

Under these assumptions, y_{it} can be expressed as

$$y_{it} = \frac{\eta_i}{1-\alpha} + w_{it} = \mu_i + w_{it} \tag{4}$$

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

2.1 The basic GMM estimator

We shall provide the GMM estimator which is commonly used in the literature.[5] Following Arellano and Bover (1995), Alvarez and Arellano (2003), Hahn, Hausman and Kuersteiner (2002) and Okui (2005b), to remove individual effects from the model, we employ the following matrix, F, called the forward orthogonal deviation (FOD) transformation operator,

$$F = diag\left[\sqrt{\frac{T-2}{T-1}}, ..., \sqrt{\frac{1}{2}}\right]\begin{bmatrix} 1 & -\frac{1}{T-2} & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ 0 & 1 & -\frac{1}{T-3} & \cdots & -\frac{1}{T-3} & -\frac{1}{T-3} & -\frac{1}{T-3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{5}$$

This matrix has the feature that $FF' = I_{T-2}$, $F'F = Q_{T-1} = I_{T-1} - \iota_{T-1}\iota'_{T-1}/(T-1)$ and $F\iota_{T-1} = 0$. By premultiplying this matrix F in (2), we obtain

$$y_i^* = \alpha x_i^* + v_i^* \tag{6}$$

where $y_i^* = Fy_i$, $x_i^* = Fx_i$ and $v_i^* = Fu_i$. v_i^*. The $t-2$-th element of v_i^* would be

$$v_{it}^* = c_t\left[v_{i,t-1} - \frac{1}{T-t+1}(v_{i,t} + \cdots + v_{i,T})\right] \quad t = 3, ..., T \tag{7}$$

where

$$c_t^2 = \frac{T-t+1}{T-t+2} \tag{8}$$

In the literature it is common to use $z_{it}^{la} = (x_{i,2}, ..., x_{i,t-1})'$ as instruments.[6] Then the moment condition based on these instruments can be written as

$$E[Z_i^{la'} v_i^*] = 0 \tag{9}$$

[5] We do not employ the first difference (Arellano and Bond, 1991), the level (Arellano and Bover, 1995), and the system (Blundell and Bond, 1998) GMM estimators, because these GMM estimators suffer from large biases when T is large and when substantial heterogeneity is present. As shown in Hayakawa (2006), these GMM estimators are inconsistent when both N and T are large. Moreover, Bun and Kiviet (2006) and Hayakawa (2005) demonstrated that the finite sample bias of these estimators heavily depends on the degree of heterogeneity.

[6] See, for example, Alvarez and Arellano (2003), Hahn, Hausman and Kuersteiner (2002) and Okui (2005b).

where Z_i^{la} is a block diagonal matrix whose $(t-2)$th element is $z_{it}^{la'}$:

$$Z_i^{la} = \begin{bmatrix} z_{i,3}^{la'} & & & O \\ & z_{i,4}^{la'} & & \\ & & \ddots & \\ O & & & z_{i,T}^{la'} \end{bmatrix} \tag{10}$$

If we assume that v_{it} has a constant variance σ_v^2 given $\eta_i, y_{i,1}, ..., y_{i,t-1}$, the optimal weighting matrix is

$$E[Z_i^{la'} v_i^* v_i^{*'} Z_i^{la}] = \sigma_v^2 E[Z_i^{la'} Z_i^{la}] \tag{11}$$

This indicates that we do not need the two-step procedure to obtain an efficient GMM estimator.[7] Therefore, the efficient GMM estimator is defined as

$$\hat{\alpha}_{la} = \frac{x^{*'} P^{la} y^*}{x^{*'} P^{la} x^*} = \frac{\sum_{t=3}^{T} x_t^{*'} P_t^{la} y_t^*}{\sum_{t=3}^{T} x_t^{*'} P_t^{la} x_t^*} \tag{12}$$

where $x^* = (x_1^{*'}, ..., x_N^{*'})'$, $y^* = (y_1^{*'}, ..., y_N^{*'})'$, $P^{la} = Z^{la}(Z^{la'} Z^{la})^{-1} Z^{la'}$, $Z^{la} = (Z_1^{la'}, ..., Z_N^{la'})'$, $x_t^* = (x_{1t}^*, ..., x_{Nt}^*)'$, $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$, $P_t^{la} = Z_t^{la}(Z_t^{la'} Z_t^{la})^{-1} Z_t^{la'}$, and $Z_t^{la} = (z_{1t}^{la}, ..., z_{Nt}^{la})'$.

3 The effects of large heterogeneity

In this section we consider the effects of large heterogeneity on the GMM estimator with instruments in levels, especially in terms of the effect on its asymptotic biases and variances. Since the many instruments problem occurs when T is large, we consider the asymptotics where both N and T tend to infinity with $T/N \to c, (0 \leq c \leq 1)$.

3.1 GMM with all available instruments in levels

Alvarez and Arellano (2003) showed the following asymptotic result.

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

$$\hat{\alpha}_{la} \to^p \alpha \tag{13}$$

[7] Here, the term "efficient" refers to the large N and fixed T asymptotics.

Moreover, provided that $T/N \to c$, $0 \leq c \leq 1$,

$$\sqrt{N(T-2)}\left[\hat{\alpha}_{la} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \to^{d} N(0, 1-\alpha^2) \tag{14}$$

Note that Hahn and Kuersteiner (2002), using a Hajćk-type convolution theorem, establish that $N(0, 1-\alpha^2)$ is the minimal asymptotic distribution. Hence $(1-\alpha^2)$ is the lower bound of the asymptotic variance.

We find that the asymptotic bias and variance of $\hat{\alpha}_{la}$ are not affected by any potential large heterogeneity, since σ_η^2/σ_v^2 does not appear in the asymptotic distribution. This is because, as the proof of Lemma C2 in Alvarez and Arellano (2003) shows, the individual effects vanish as T gets larger. Hence, we can say that $\hat{\alpha}_{la}$ is robust to large heterogeneity. However, Okui (2005b) has shown that the size distortion of the test for the hypothesis $H_0 : \alpha = \alpha_0$ is very large and inference based on $\hat{\alpha}_{la}$ is therefore unreliable.

We suspect that the source of the size distortion is the bias which results from using all instruments. Therefore it would be expected that reducing the number of instruments would mitigate this problem since using fewer instruments reduces the bias of the estimator.

3.2 GMM with the minimum number of instruments in levels

In this subsection, we consider a GMM estimator that uses the minimum number of instruments, that is, $z_{it}^{lm} = x_{i,t-1}$. This means that we use only one instrument in each period. In this case, since the number of instruments does not grow as T gets larger, we would expect the bias to become small. The GMM estimator with instruments z_{it}^{lm} can be defined as

$$\hat{\alpha}_{lm} = \frac{\sum_{t=3}^{T} x_t^{*\prime} P_t^{lm} y_t^*}{\sum_{t=3}^{T} x_t^{*\prime} P_t^{lm} x_t^*} \tag{15}$$

where $P_t^{lm} = Z_t^{lm}(Z_t^{lm\prime} Z_t^{lm})^{-1} Z_t^{lm\prime}$, and $Z_t^{lm} = (z_{1t}^{lm}, ..., z_{Nt}^{lm})'$. The next theorem establishes the asymptotic properties of $\hat{\alpha}_{lm}$.

Theorem 2. *Let Assumptions 1, 2, and 3 hold. Then as both N and T tend to infinity,*

$$\hat{\alpha}_{lm} \to^{p} \alpha \tag{16}$$

and

$$\sqrt{N(T-2)}\left[\hat{\alpha}_{lm}-\left(\alpha-\frac{1}{N(T-2)}(1+\alpha)\rho_{lm}^{-1}\right)\right]\to^{d} N\left(0,(1-\alpha^2)\rho_{lm}^{-1}\right) \qquad (17)$$

where $k=\sigma_\eta^2/\sigma_v^2$ *and*

$$\rho_{lm}^{-1}=1+k\left(\frac{1+\alpha}{1-\alpha}\right)\geq 1 \qquad (18)$$

Remark 1 We find that there is a notable difference between $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$ with regards to the individual effects. Although the individual effects in $\hat{\alpha}_{la}$ vanish as T gets larger, this is not the case with $\hat{\alpha}_{lm}$. The large heterogeneity crucially affects the asymptotic bias and variance of $\hat{\alpha}_{lm}$. Both the asymptotic bias and the variance increase in proportion to k, the degree of heterogeneity.

Remark 2 In the case of $k=0$, i.e., $\sigma_\eta^2=0$, the asymptotic variance of $\hat{\alpha}_{lm}$ is equal to the lower bound of $1-\alpha^2$. In this case, the instruments that are not used in the estimation, i.e., $(x_{i2},...,x_{i,t-2})$, become redundant in the sense that using them does not improve efficiency.[8] Hence, when $k=0$, in terms of the bias, using the minimum number of instruments is preferable, since the magnitude of the asymptotic bias of $\hat{\alpha}_{lm}$ is $(1+\alpha)/N(T-2)$, while that of $\hat{\alpha}_{la}$ is $(1+\alpha)/N$.

Remark 3 If the degree of heterogeneity is large, serious problems occur. Both the asymptotic bias and the variance increase. This indicates that if we use the minimum number of instruments to reduce the bias springing from the use of many instruments, then a bias due to large heterogeneity will appear. Especially if $T-2<\rho_{lm}^{-1}$, the asymptotic bias of $\hat{\alpha}_{lm}$ will be larger than that of $\hat{\alpha}_{la}$, even though $\hat{\alpha}_{lm}$ uses a smaller number of instruments than $\hat{\alpha}_{la}$. Hence, if a large degree of heterogeneity is present, reducing the number of instruments to reduce the bias may not work well. Furthermore, the asymptotic variance becomes quite large. Based on these findings, we conjecture that Okui's method does not work well if heterogeneity is large. If there is large heterogeneity, Okui's method tends to use more instruments to weaken the influence of individual effects.[9] However, if we use more instruments, the estimator will be more biased and inference will tend to be inaccurate.

[8] See Breusch et al (1999) for a discussion of the redundancy of the moment conditions in GMM.
[9] See Table 1 in Okui (2005b) for the optimal lag length of the instruments.

The results in this section indicate that if we use all the instruments in levels, although $\hat{\alpha}_{la}$ is robust to a large degree of heterogeneity, the size distortion is substantial. On the other hand, if we use the minimum number of instruments in levels, the effect of a large degree of heterogeneity on the estimator is large. Therefore, if a large degree of heterogeneity is present in the model, both $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$ are no longer desirable estimators. Neither of them has a small bias or a variance without size distortion. This suggests there is a need for new estimators which overcome the drawbacks mentioned above. We will present such a new estimator in the next section.

4 Removing the individual effects from the instruments

Since the asymptotic distribution of $\hat{\alpha}_{lm}$ is heavily affected by a large degree of heterogeneity through the instruments, we expect that if we use the instruments without the individual effects, the GMM estimator will be robust to the presence of large heterogeneity. In this section, we consider the removal of the individual effects from the instruments. We employ two methods to remove the individual effects. The first is simply to take the first difference. The second is to introduce a transformation called the Backward Orthogonal Deviation (BOD) transformation. BOD transformation is a modification of FOD transformation. Although the FOD transformation induces a deviation from the mean of all future values, the BOD transformation induces a deviation from the mean of all past values. To rid the instruments of the individual effects, we only have to multiply the following matrix:

$$B = diag\left[\sqrt{\frac{1}{2}}, ..., \sqrt{\frac{T-2}{T-1}}\right]\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ -\frac{1}{T-3} & -\frac{1}{T-3} & -\frac{1}{T-3} & \cdots & -\frac{1}{T-3} & 1 & 0 \\ -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & 1 \end{bmatrix} \quad (19)$$

By multiplying this matrix by x_i, we get the following:

$$x_{it}^b = b_t\left[x_{i,t} - \frac{1}{t-2}(x_{i,2} + \cdots + x_{i,t-1})\right] \quad (20)$$

$$= \quad b_t \left[w_{i,t-1} - \frac{1}{t-2}(w_{i,1} + \cdots + w_{i,t-2}) \right] \qquad\qquad t = 3, ..., T \tag{21}$$

where

$$b_t^2 = \frac{t-2}{t-1} \tag{22}$$

The GMM estimators with instruments $z_{it}^{dm} = \Delta x_{i,t-1}$ and $z_{it}^{bm} = x_{i,t-1}^b$ are defined as

$$\hat{\alpha}_{dm} = \frac{\sum_{t=4}^T x_t^{*\prime} P_t^{dm} y_t^*}{\sum_{t=4}^T x_t^{*\prime} P_t^{dm} x_t^*} \tag{23}$$

$$\hat{\alpha}_{bm} = \frac{\sum_{t=4}^T x_t^{*\prime} P_t^{bm} y_t^*}{\sum_{t=4}^T x_t^{*\prime} P_t^{bm} x_t^*} \tag{24}$$

where $P_t^{dm} = Z_t^{dm}(Z_t^{dm\prime} Z_t^{dm})^{-1} Z_t^{dm\prime}$, $Z_t^{dm} = (z_{1t}^{dm}, ..., z_{Nt}^{dm})'$, $P_t^{dm} = Z_t^{bm}(Z_t^{bm\prime} Z_t^{bm})^{-1} Z_t^{bm\prime}$, and $Z_t^{bm} = (z_{1t}^{bm}, ..., z_{Nt}^{bm})'$.

There are two notable features in x_{it}^b. The first is that x_{it}^b has no individual effects, and this is the main purpose of using the BOD transformation. The second is that since x_{it}^b is composed of all past values, we would expect that it contains more information than using only one instrument in levels or the first-differenced instrument, i.e. we expect that the GMM estimator with minimum number of x_{it}^b will be more efficient than $\hat{\alpha}_{lm}$ and $\hat{\alpha}_{dm}$. The following asymptotic analysis shows that this conjecture is correct.

Theorem 3. *Let Assumptions 1, 2, and 3 hold. Then as both N and T tend to infinity,*

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

and

$$\sqrt{N(T-3)} \left[\hat{\alpha}_{dm} - \left(\alpha - \frac{1}{N(T-3)}(1+\alpha)\rho_{dm}^{-1} \right) \right] \to^d N\left(0, (1-\alpha^2)\rho_{dm}^{-1}\right) \tag{26}$$

where

$$\rho_{dm}^{-1} \quad = \quad \left(\frac{2}{1-\alpha} \right) > 1 \tag{27}$$

Theorem 4. *Let Assumptions 1, 2, and 3 hold. Then as both N and T tend to infinity,*

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

and

$$\sqrt{N(T-3)}\left[\hat{\alpha}_{bm}-\left(\alpha-\frac{1}{N(T-3)}(1+\alpha)\right)\right]\to^{d}N(0,1-\alpha^{2}) \tag{29}$$

Remark 4 Compared with $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$, the asymptotic biases and variances of $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$ are not affected by k, and this is the main purpose of using instruments without individual effects. Therefore, we can say that $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$ are robust to the presence of a large degree of heterogeneity.

Remark 5 If we compare $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$, there is a notable difference both in their asymptotic biases and their variances. Since ρ_{dm}^{-1} is strictly larger than one, both the asymptotic bias and the variance of $\hat{\alpha}_{dm}$ are strictly larger than those of $\hat{\alpha}_{bm}$. Therefore, we can say that $\hat{\alpha}_{bm}$ is superior to $\hat{\alpha}_{dm}$.

Remark 6 The magnitude of the asymptotic bias of $\hat{\alpha}_{bm}$ is $(1+\alpha)/N(T-3)$, while the asymptotic biases of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$ and $\hat{\alpha}_{dm}$ are $(1+\alpha)/N$, $(1+\alpha)\rho_{lm}^{-1}/N(T-2)$, and $(1+\alpha)\rho_{dm}^{-1}/N(T-3)$, respectively. Also, as shown by Alvarez and Arellano (2003), the magnitude of the asymptotic biases of the within-groups estimator and the LIML estimator are $(1+\alpha)/(T-2)$ and $(1+\alpha)/(2N-(T-2))$, respectively. Thus, the magnitude of the asymptotic bias of $\hat{\alpha}_{bm}$ is smallest among these commonly-used estimators.

Remark 7 The asymptotic variance of $\hat{\alpha}_{dm}$ is strictly larger than the lower bound and can never be efficient. But the asymptotic variance of $\hat{\alpha}_{bm}$ is equal to the lower bound, and $\hat{\alpha}_{bm}$ is therefore asymptotically efficient. Also, it is notable that although $\hat{\alpha}_{la}$ becomes asymptotically efficient by using all instruments, $\hat{\alpha}_{bm}$ is asymptotically efficient by using the minimum number of instruments. This implies that the instruments which are not used, i.e., $(x_{i,3}^{b},...,x_{i,t-2}^{b})$, are asymptotically redundant.

Thus, the new estimator $\hat{\alpha}_{bm}$ addresses the trade-off between the bias and the variance: the asymptotic bias of $\hat{\alpha}_{bm}$ is smaller than that of other GMM estimators, whereas the asymptotic variance is equal to the lower bound.

We can say that the main advantage of $\hat{\alpha}_{bm}$ lies in its variance. To examine the variance properties in greater detail, we analytically compare the asymptotic variances of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$, $\hat{\alpha}_{dm}$, and $\hat{\alpha}_{bm}$ under the fixed T asymptotics. The asymptotic variances of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$, $\hat{\alpha}_{dm}$, and $\hat{\alpha}_{bm}$ under large N and fixed T asymptotics are given in the following lemma.

Lemma 1. *Under Assumptions 1, 2, and 3, the asymptotic variances of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$, $\hat{\alpha}_{dm}$, and $\hat{\alpha}_{bm}$ under large N and fixed T asymptotics are given by*

$$Avar(\hat{\alpha}_{la}) = (1-\alpha^2)\left[N\sum_{t=3}^{T}\psi_t^2\left(1-\frac{k(\frac{1+\alpha}{1-\alpha})}{1+k\{(\frac{1+\alpha}{1-\alpha})+(t-3)\}}\right)\right]^{-1} \tag{30}$$

$$Avar(\hat{\alpha}_{lm}) = (1-\alpha^2)\rho_{lm}^{-1}\left[N\sum_{t=3}^{T}\psi_t^2\right]^{-1} \tag{31}$$

$$Avar(\hat{\alpha}_{dm}) = (1-\alpha^2)\rho_{dm}^{-1}\left[N\sum_{t=4}^{T}\psi_t^2\right]^{-1} \tag{32}$$

$$Avar(\hat{\alpha}_{bm}) = (1-\alpha^2)\left[N\sum_{t=4}^{T}\psi_t^2\left(1-\frac{\alpha\phi_{t-3}}{t-3}\right)^2\Lambda_t^{-1}\right]^{-1} \tag{33}$$

where

$$\phi_j = \frac{1-\alpha^j}{1-\alpha} = 1+\alpha+\cdots+\alpha^{j-1} \tag{34}$$

$$\psi_t^2 = c_t^2\left[1-\frac{\alpha\phi_{T-t+1}}{T-t+1}\right]^2 \tag{35}$$

$$\Lambda_t = \left[1-\frac{2\alpha\phi_{t-3}}{t-3}+\frac{1}{(t-3)^2}\left\{\frac{(t-3)(1+\alpha)}{1-\alpha}-\frac{2\alpha(1-\alpha^{t-3})}{(1-\alpha)^2}\right\}\right] \tag{36}$$

Provided that $\sigma_\eta^2 = 0$, then the asymptotic variances of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$ reduce to

$$Avar(\hat{\alpha}_{la}) = Avar(\hat{\alpha}_{lm}) = (1-\alpha^2)\left[N\sum_{t=3}^{T}\psi_t^2\right]^{-1} \tag{37}$$

and provided that $\sigma_\eta^2 \to \infty$, then

$$Avar(\hat{\alpha}_{la}) = (1-\alpha^2)\left[N\sum_{t=3}^{T}\psi_t^2\left(1-\frac{(\frac{1+\alpha}{1-\alpha})}{(\frac{1+\alpha}{1-\alpha})+(t-3)}\right)\right]^{-1} < \infty \tag{38}$$

$$Avar(\hat{\alpha}_{lm}) \to \infty \tag{39}$$

We find that the asymptotic variances of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$ are exactly the same if $\sigma_\eta^2 = 0$. This coincides with the case where both N and T are large. However if $\sigma_\eta^2 \to \infty$, then the effect of large heterogeneity on $\hat{\alpha}_{lm}$ is unbounded, while $\hat{\alpha}_{la}$ receives a bounded influence.

We showed that under the double asymptotics where N and T tend to infinity, the asymptotic variances of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{bm}$ are the same. But by comparing the asymptotic variance of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{bm}$ when T is fixed, we find that the forms of the asymptotic variances are quite different. To examine this difference, we compare the asymptotic variances numerically. Figures 1 to 3 present the asymptotic variances for the case of $\alpha = 0.8$ and $N = 50$ and $k = 0.2, 1$, and 10 (Figures 1, 2, and 3, respectively). The horizontal axis shows T from $T = 10$ to $T = 29$, while the vertical axis depicts the magnitude of the asymptotic standard error calculated from Lemma 1, that is, the root of the asymptotic variances. An inspection of the figures shows that the asymptotic variance of $\hat{\alpha}_{lm}$ is heavily affected by the presence of large heterogeneity. Note the difference of the scale of the vertical axis in Figure 3. We also find that there is a significant difference between $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$. Although we find that there is a difference between $\hat{\alpha}_{la}$ and $\hat{\alpha}_{bm}$ when T is not so large, this difference shrinks as T gets larger. This fact coincides with the double asymptotic analysis.

5 Monte Carlo simulation

In this section we conduct Monte Carlo experiments to examine the performance of the estimators discussed above. We first consider a simple AR(1) model and then extend the analysis to consider the case where a predetermined variable is included.

5.1 Cases without covariates

We consider the following AR(1) model:

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + v_{it} \tag{40}$$

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 $N = 50, 100$, $T = 10, 15, 25$ 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 mean (mean), standard deviation (std), standard error (se), the root mean squared error (rmse), and the size of the Wald test for $H_0 : \alpha = \alpha_0$, where α_0 is the true value.[10]

These experiments fulfill five aims. The first is to discover how large the bias and the size distortion of $\hat{\alpha}_{la}$ are. The second is to examine how seriously the bias and variance of $\hat{\alpha}_{lm}$ are affected by the presence of a large degree of heterogeneity. The third is to see how large the differences in the bias and the variance of $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$ are. The fourth aim is to compare the variances of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{bm}$. And the final aim is to compare the power of $\hat{\alpha}_{lm}$, $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$.

We begin the examination of these five issues by first considering $\hat{\alpha}_{la}$. Tables 1 and 2 respectively report the simulation results for $\hat{\alpha}_{la}$ for the case of $N = 50$ and for $N = 100$. In the case of $T = 10$, the bias of $\hat{\alpha}_{la}$ is quite large and as σ_η^2/σ_v^2 gets larger the magnitude of the bias increases. In the case of $T = 25$, although the magnitude of the bias is smaller than in the case of $T = 10$, a large bias still remains. Although the magnitude of the bias increases as σ_η^2/σ_v^2 gets larger, it is still much smaller than in the case where T=10. This result supports the theoretical prediction that the individual effects vanish as T gets larger. With regards to the sizes, they are no longer close to the nominal level. Especially when $\alpha = 0.8$, the size distortion is substantial and we can say that inference is inaccurate.

The second aim of the Monte Carlo study is to examine the effect of large heterogeneity on $\hat{\alpha}_{lm}$. In the case of $\sigma_\eta^2/\sigma_v^2 = 0.2$, $\hat{\alpha}_{lm}$ performs very well. The RMSE of $\hat{\alpha}_{lm}$ is smaller than that of $\hat{\alpha}_{la}$. However, as σ_η^2/σ_v^2 gets larger, the performance of $\hat{\alpha}_{lm}$ dramatically worsens. Even in the case of $\sigma_\eta^2/\sigma_v^2 = 1$, the RMSE of $\hat{\alpha}_{lm}$ is larger than that of $\hat{\alpha}_{la}$. Especially in the case of $\sigma_\eta^2/\sigma_v^2 = 10$, with a few exceptions, the biases of $\hat{\alpha}_{lm}$ are larger than those of $\hat{\alpha}_{la}$ even though $\hat{\alpha}_{lm}$ uses a smaller number of instruments than $\hat{\alpha}_{la}$. These results coincide with the case where $T - 2 < \rho_{lm}^{-1}$ holds. In this case, the asymptotic bias of $\hat{\alpha}_{lm}$ is larger than that of $\hat{\alpha}_{la}$. For example, in the case of $T = 25$, $k = 10.$ and $\alpha = 0.8$, $1 + k(1 + \alpha)/(1 - \alpha) = 91 > T - 2 = 23$. With regards to the sizes, they are much closer to the nominal level than those of $\hat{\alpha}_{la}$.

[10] The standard errors are calculated under the large N and fixed T asymptotics, i.e. $se(\hat{\alpha}) = \sqrt{\hat{\sigma}_v^2(x^{*\prime}Px^*)^{-1}}$, and the size is based on the usual Wald test using a 5% level of significance.

Third, we compare $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$. Tables 3 and 4 report, respectively, the simulation results for $\hat{\alpha}_{dm}$ for $N = 50$ and $N = 100$, while Tables 5 and 6 report the simulation results for $\hat{\alpha}_{bm}$ for $N = 50$ and $N = 100$. For the purpose of comparison, we also consider the case where all available instruments are used. The corresponding GMM estimators are as follows:

$$\hat{\alpha}_{da} = \frac{\sum_{t=4}^{T} x_t^{*\prime} P_t^{da} y_t^*}{\sum_{t=4}^{T} x_t^{*\prime} P_t^{da} x_t^*} \tag{41}$$

$$\hat{\alpha}_{ba} = \frac{\sum_{t=4}^{T} x_t^{*\prime} P_t^{ba} y_t^*}{\sum_{t=4}^{T} x_t^{*\prime} P_t^{ba} x_t^*} \tag{42}$$

where $P_t^{da} = Z_t^{da}(Z_t^{da\prime} Z_t^{da})^{-1} Z_t^{da\prime}$, $Z_t^{da} = (z_{1t}^{da}, ..., z_{Nt}^{da})'$, $z_{it}^{da} = (\Delta x_{i3}, ..., \Delta x_{i,t-1})'$, $P_t^{ba} = Z_t^{ba}(Z_t^{ba\prime} Z_t^{ba})^{-1} Z_t^{ba\prime}$, $Z_t^{ba} = (z_{1t}^{ba}, ..., z_{Nt}^{ba})'$, and $z_{it}^{ba} = (x_{i3}^b, ..., x_{i,t-1}^b)'$.

Looking at the tables, we find that $\hat{\alpha}_{da}$ and $\hat{\alpha}_{ba}$ are numerically equivalent and that none of the four estimators are affected by σ_η^2/σ_v^2. With regards to the bias, there are almost no differences between $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$ when $\alpha = 0.2$ and 0.5. However, in the case of $\alpha = 0.8$, $\hat{\alpha}_{bm}$ has smaller bias than $\hat{\alpha}_{dm}$. Next we compare the variances of $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$. The tables indicate that there are significant differences in the magnitude of the variances of $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$. The variance of $\hat{\alpha}_{bm}$ is much smaller than that of $\hat{\alpha}_{dm}$. This result is in line with the theoretical prediction. In particular, if we compare $\hat{\alpha}_{da}$ and $\hat{\alpha}_{dm}$, the increase of the variance of $\hat{\alpha}_{dm}$ compared to that of $\hat{\alpha}_{da}$ is quite large, and as a result the RMSE of $\hat{\alpha}_{dm}$ is larger than that of $\hat{\alpha}_{da}$ in many cases. In contrast with $\hat{\alpha}_{dm}$, the degree of increase of the variance of $\hat{\alpha}_{bm}$ compared to that of $\hat{\alpha}_{ba}$ is very small. As a result, the RMSEs of $\hat{\alpha}_{bm}$ are smaller than those of $\hat{\alpha}_{ba}$ in all cases. Furthermore, the sizes are very close to the nominal level.

Fourth, we compare the variances of $\hat{\alpha}_{bm}$ and $\hat{\alpha}_{la}$. In the case of $T = 10$, the variance of $\hat{\alpha}_{bm}$ is a little larger than that of $\hat{\alpha}_{la}$. However, as T gets larger, the difference gets smaller. Especially in terms of the RMSE, $\hat{\alpha}_{bm}$ has smaller RMSE than $\hat{\alpha}_{la}$ in almost all the cases. The exception is when $T = 10$ and $\sigma_\eta^2/\sigma_v^2 = 0.2$. Taking into consideration the size distortion, we can conclude that $\hat{\alpha}_{bm}$ performs better than $\hat{\alpha}_{la}$.

Lastly, we compare the power of $\hat{\alpha}_{lm}$, $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$. We do not consider the estimators with all instruments since their sizes are far from the nominal level. Figures 4 to 12 show the result. In each case, $N = 50$ and $\alpha = 0.8$ are fixed.

Figures 4 to 6 depict the cases when $T = 10$ and $k = 0.2, 1, 10$, while Figure 7 to 9 depict the cases when $T = 15$ and $k = 0.2, 1, 10$; finally, Figures 10 to 12 depict the cases when $T = 25$ and $k = 0.2, 1, 10$. Looking at all these figures, we find that the power of $\hat{\alpha}_{lm}$ is crucially affected by the degree of heterogeneity, whereas the power of $\hat{\alpha}_{dm}$ and $\hat{\alpha}_{bm}$ is not. Furthermore, we find that $\hat{\alpha}_{bm}$ has higher power than $\hat{\alpha}_{dm}$.

Reducing the bias

The simulation results above show that the magnitude of the bias of $\hat{\alpha}_{bm}$ in the case of $T = 10$ is not negligible, although the size is close to the nominal level. Here we show that we can reduce the bias by using the matrix, which is different from (10) where the instruments are on the diagonal:

$$Z_i^{bm} = \begin{bmatrix} x_{i,3}^b & & & O \\ & x_{i,4}^b & & \\ & & \ddots & \\ O & & & x_{i,T-1}^b \end{bmatrix} \tag{43}$$

Let us define $\tilde{Z}_i^{bm}$ as follows:

$$\tilde{Z}_i^{bm\prime} = \begin{bmatrix} x_{i,3}^b & \cdots & x_{i,\bar{t}-1}^b & & O & \\ & O & & x_{i,\bar{t}}^b & \cdots & x_{i,T-1}^b \end{bmatrix} = \begin{bmatrix} Z_i^{(1)\prime} & O \\ O & Z_i^{(2)\prime} \end{bmatrix} \tag{44}$$

where $\bar{t} = [(T-3)/2] + 3$. $[\quad]$ denotes the integer part of the argument. Then it follows that the GMM estimator with $\tilde{Z}_i^{bm}$, $\tilde{\alpha}_{bm}$, is derived from two moment conditions. As $\tilde{\alpha}_{bm}$ uses a smaller number of moment conditions than $\hat{\alpha}_{bm}$, we expect that $\tilde{\alpha}_{bm}$ has smaller bias than $\hat{\alpha}_{bm}$ at the cost of efficiency.[11] Table 7 summarizes the simulation results. We find that $\tilde{\alpha}_{bm}$ is very close to the true value although its variance increases a little. In terms of the RMSE, $\hat{\alpha}_{bm}$ performs best in almost all the cases. Therefore, $\tilde{\alpha}_{bm}$ may be an option when we are interested in the value of a coefficient. In the simulation that follows, we focus only on $\hat{\alpha}_{bm}$ since it has a smaller RMSE than $\tilde{\alpha}_{bm}$.

[11] See Wooldridge (2005).

5.2 Cases with an additional regressor

In this subsection, we consider the case where a predetermined variable is included besides the lagged dependent variable. The aim of this design is to investigate the effect of an additional regressor. We consider the following dynamic panel data model with a covariate:

$$
\begin{aligned}
y_{it} &= \alpha y_{i,t-1} + \beta X_{it} + \eta_i + v_{it} \;=\; W_{it}'\delta + \eta_i + v_{it} \qquad t = 2, ..., T & (45)\\
X_{it} &= \rho X_{i,t-1} + \tau\eta_i + \theta v_{i,t-1} + \varepsilon_{it} & (46)
\end{aligned}
$$

where $W_{it} = (y_{i,t-1},\ X_{it})'$ and $\delta = (\alpha,\ \beta)'$. Initial observations are generated to be covariance stationary and we discard the first 10 periods. In this model, X_{it} is a predetermined variable. Also note that X_{it} is correlated with η_i. In the experiment, we set $\alpha = 0.8$, $\beta = 0.5$, $\rho = 0.5$, $\tau = 0.2$, and $\theta = 0.2$. N and T are $N = 50, 100$ and $T = 10, 15, 25$. In addition, we set $var(v_{it}) = var(\varepsilon_{it}) = 1$ and $\sigma_\eta^2 = 0.2, 1, 10$.

Define $y_i = (y_{i2}, ..., y_{iT})'$, $x_i = (y_{i1}, ..., y_{i,T-1})'$, $X_i = (X_{i2}, ..., X_{i,T})'$, $W_i = (W_{i2}, ..., W_{iT})'$ and $v_i = (v_{i2}, ..., v_{iT})'$. By multiplying F by y_i, W_i and v_i, (45) becomes

$$
y_{it}^* = \alpha x_{it}^* + \beta X_{it}^* + v_{it}^* = W_{it}^{*'}\delta + v_{it}^* \qquad t = 3, ..., T \qquad (47)
$$

Let z_{it} denote the generic instruments for W_{it}^* and let $Z_t = (z_{1t}, ..., z_{Nt})'$, $P_t = Z_t(Z_t'Z_t)^{-1}Z_t$. Then the GMM estimator has the following form:

$$
\hat{\delta} = \left(\sum_{t=t_0}^{T} W_t^{*'} P_t W_t^*\right)^{-1} \left(\sum_{t=t_0}^{T} W_t^{*'} P_t y_t^*\right) \qquad (48)
$$

where $t_0 = 3$ if the instruments do not contain $x_{i,t-1}^b$ and $t_0 = 4$ otherwise.

We consider three type of instruments for W_{it}^*. The first is $z_{it}^{lla} = (x_{i2}, ..., x_{i,t-1}, X_{i,1}, ..., X_{i,t-1})$ where all available instruments are exploited. The second is $z_{it}^{llm} = (x_{i,t-1}, X_{i,t-1})$ where the minimum number of instruments are used. The third is $z_{it}^{bbm} = (x_{i,t-1}^b, X_{it}^b)$ where $x_{i,t-1}^b$ and X_{it}^b are the $(t-3)$th and $(t-2)$th elements of $x_i^b = Bx_i$ and $X_i^b = B(X_{i,2}, \cdots, X_{i,T})'$, respectively. Let $\hat{\delta}_{la} = (\hat{\alpha}_{la}, \hat{\beta}_{la})'$, $\hat{\delta}_{lm} = (\hat{\alpha}_{lm}, \hat{\beta}_{lm})'$, and $\hat{\delta}_{bm} = (\hat{\alpha}_{bm}, \hat{\beta}_{bm})'$ denote the GMM estimators corresponding to z_{it}^{la}, z_{it}^{lm}, and z_{it}^{bm}. Tables 8 and 9 show the simulation results. With regards to the effect of large heterogeneity, the result is similar to the AR(1) case, that is, $\hat{\alpha}_{la}$ is not greatly affected by large heterogeneity and becomes more robust to large heterogeneity as

T gets larger. However, the size distortion is very large and the degree of distortion is more serious than in the AR(1) case. Unlike $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$ is sensitive to large heterogeneity. As σ_η^2/σ_v^2 gets larger, the RMSEs increase. Turning to $\hat{\alpha}_{bm}$, we find that by construction it is robust to large heterogeneity. The RMSEs of $\hat{\alpha}_{bm}$ are smaller than those of $\hat{\alpha}_{la}$ and $\hat{\alpha}_{lm}$ except for the case of $\sigma_\eta^2/\sigma_v^2 = 0.2$ for any N and T. Furthermore, the sizes of $\hat{\alpha}_{bm}$ are close to the nominal ones. When we are interested in the estimation and inference of β, either $\hat{\beta}_{bm}$ or $\hat{\beta}_{lm}$ work well since both estimators exhibit a similar performance.

6 An Empirical application

In this section, we apply our new estimator to a partial adjustment model for employment dynamics using the data employed by Arellano (2002). The data consist of a panel for 385 Spanish firms, starting in 1983 and spanning 14 years. For a more detailed description of the data, see Bover and Watson (2004). The model is given by

$$n_{it} = \alpha n_{i,t-1} + \beta w_{it} + \eta_i + v_{it} \tag{49}$$

where n_{it} is the logarithm of employment at firm i at time t and w_{it} is the logarithm of wages paid by firm i at time t. w_{it} is treated as a predetermined variable.

We computed $\hat{\delta}_{la}$, $\hat{\delta}_{bm}$ and $\tilde{\delta}_{bm}$ and their standard errors. The estimation results are presented in Table 10.[12] The results show that the GMM estimators proposed in this paper alleviate the bias of $\hat{\alpha}_{la}$. Based on the simulation studies in Section 5, which imply that the empirical sizes of $\hat{\alpha}_{bm}$ and $\tilde{\alpha}_{bm}$ are close to the nominal level, we should make inference by $\hat{\alpha}_{bm}$ and $\tilde{\alpha}_{bm}$.

7 Conclusion

In this paper, we addressed the many instruments problem in dynamic panel data models where unobservable heterogeneity may be large. We proposed a new form

[12] Time effects are removed by subtracting the cross-sectional averages of each period prior to the estimation.

of instruments with which we can overcome the many instruments problem. The proposed GMM estimator has smaller asymptotic bias than the conventional GMM, LIML and within-groups estimators, whereas its asymptotic variance is equal to the lower bound even if there is large heterogeneity in the model. Simulation results showed that in many cases the RMSEs of the proposed GMM estimators are smaller than the conventional GMM estimators. Furthermore, the size of the test for the parameter hypothesis was very close to the nominal size. The analysis of the new estimators was then extended to the case where additional regressors are included and it was found that the estimator performed well in such cases. Finally, we applied our new estimator to the data of Bover and Watson (2004) and were able to confirm that it alleviates the bias problem.

References

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

[2] Amemiya, T (1985). *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

[3] Anderson, T.W., N.Kunitomo, and Y. Matsushita (2005): "A New Light from Old Wisdoms . Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models," CIRJE Discussion Paper, University of Tokyo.

[4] Andrews, D. W. and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Paper No 1530, Yale University.

[5] Angrist, J.D., G.W. Imbens and A. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, Vol. 14, pp. 57-67.

[6] Angrist, J.D., and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, Vol. 106, pp. 979-1014.

[7] Arellano, M. (2002): "Modelling Optimal Instrumental Variables for Dynamic Panel Data Models", CEMFI Working Paper no. 0310.

[8] 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*, Vol. 58, pp. 277-297.

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

[10] Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, Vol. 62, pp. 657-681.

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

[12] Bound, J., D.A. Jaeger and R.M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, Vol. 90, pp. 443-450.

[13] Bover, O., and N. Watson (2004): "Are There Economies of Scales in the Demand for Money by Firms? Some Panel Data Estimates," *Journal of Monetary Economics*, Vol. 52, pp. 1569-1589.

[14] Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski (1999): "Redundancy of Moment Conditions," *Journal of Econometrics*, Vol. 91, pp. 89 - 111.

[15] 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.

[16] Chao, J. C. and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica* Vol. 73, pp. 1673-1692.

[17] Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, Vol. 69, pp. 1161-1191.

[18] Greene, W. H. (2002) *Econometric Analysis*, 5th ed., Prentice-Hall.

[19] Hahn, J (2002): "Optimal Inference with Many Instruments," *Econometric Theory*, Vol. 18, pp. 140-168.

[20] Hahn, J., J. Hausman and G. Kuersteiner (2002), "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," MIT Department of Economics Working Paper No. 01-24.

[21] Hahn, J., J. Hausman and G. Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations," *Econometrics Journal*, Vol. 7, pp. 272-306.

[22] Hahn, J. and A. Inoue (2002): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," *Econometric Reviews*, Vol. 21, pp. 309-336.

[23] 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*, Vol. 70, pp. 1639-1657.

[24] Hansen, C. and J. Hausman and W. Newey (2005): "Estimation with Many Instrumental Variables," Unpublished manuscript.

[25] 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.

[26] Hayakawa, K. (2006): "The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models when both N and T are large," Hi-Stat Discussion Paper No.129. Hitotsubashi University.

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

[28] Kunitomo, N. (1980): "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, Vol. 75, pp. 693-700.

[29] Morimune, K. (1983): "Approximate Distributions of k-class Estimators When the Degree of Overidentification is Large Compared with the Sample Size," *Econometrica*, Vol. 51, pp. 821-841.

[30] Okui, R. (2005a): "Instrumental Variable Estimation in the Presence of Many Moment Conditions," unpublished manuscript.

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

[32] Wooldridge, J (2005): "Instrumental Variables Estimation with Panel Data," *Econometric Theory*, Vol. 21, pp. 865-869.

[33] Ziliak, J.P. (1997): "Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics*, Vol. 15, pp. 419-431.

A Mathematical Proofs

Throughout the appendix, T^0 denotes:

$T-2$ when the instruments are z_{it}^{lm}; and

$T-3$ when the instruments are z_{it}^{dm} or z_{it}^{bm}.

Before we prove the theorems, we provide some lemmas.

Lemma 2. *Let κ_3 and κ_4 denote the third and fourth-order cumulants of v_{it}, and let P_t denote P_t^{lm} or P_t^{dm} or P_t^{bm}. In addition, let d_t and d_s be $N\times 1$ vectors containing the diagonal elements of P_t and P_s, respectively, so that $tr(P_t) = d_t'\iota_N = tr(P_s) = d_s'\iota_N = 1$, and $d_t'd_s \leq 1$. Then under Assumption 1 for $l \geq r \geq t$, $p \geq q \geq s$, and $t \geq s$,*

$$cov(v_l'P_tv_r, v_p'P_sv_q) = \begin{cases} 2\sigma_v^4 tr(P_tP_s) + \kappa_4 E(d_t'd_s) \leq 2\sigma_v^4 + \kappa_4 & if \quad l=r=p=q \\ \kappa_3 E(d_t'P_sv_q) & if \quad l=r=p\neq q<t \\ \sigma_v^4 tr(P_tP_s) \leq \sigma_v^4 & if \quad l=p\neq r=q \\ 0 & otherwise \end{cases} \tag{50}$$

where

$$|E(d_t'P_sv_q)| \leq \sigma_v \tag{51}$$

Proof of Lemma 2

Like Alvarez and Arellano (2003), we begin by showing the following:

$$cov_{t-1}(v_l'P_tv_r, v_p'P_sv_q) = \begin{cases} 2\sigma_v^4 + \kappa_4 tr(P_tP_s)d_t'd_s & if \quad l=r=p=q \\ \kappa_3 d_t'P_sv_q & if \quad l=r=p\neq q<t \\ \sigma_v^4 tr(P_tP_s) & if \quad l=p\neq r=q \\ 0 & otherwise \end{cases} \tag{52}$$

where E_{t-1} denotes an expectation conditional on η_i and $\{v_{i,t-1-j}\}_{j=1}^{\infty}$. To prove this, note that the conditional covariance can be expressed as

$$cov_{t-1}(v_l'P_tv_r, v_p'P_sv_q) = E_{t-1}(v_l'P_tv_rv_p'P_sv_q) - E_{t-1}(v_l'P_tv_r)E_{t-1}(v_p'P_sv_q) \tag{53}$$

Firstly the conditional mean terms in (53) are

$$E_{t-1}(v_l'P_tv_r) = E_{t-1}(v_p'P_sv_q) = \begin{cases} \sigma_v^2 & if \quad l=r \ or \ p=q \\ 0 & if \quad l\neq r \ or \ p\neq q \end{cases} \tag{54}$$

Next, with regards to the leading term of (53),we have

$$E_{t-1}(v_l'P_tv_rv_p'P_sv_q) = \begin{cases} E_{t-1}(v_l'P_tv_lv_l'P_sv_l) & if \quad l=r=p=q \\ E_{t-1}(v_l'P_tv_lv_l')P_sv_q & if \quad l=r=p\neq q<t \\ tr[P_tE_{t-1}(v_rv_r')P_sE_{t-1}(v_lv_l')] & if \quad l=p\neq r=q \\ 0 & otherwise \end{cases} \tag{55}$$

For the first type in (55)

$$\begin{aligned} E_{t-1}(v_l'P_tv_lv_l'P_tv_l) &= (3\sigma_v^4+\kappa_4)d_t'd_s+\sigma_v^4[tr(P_t)tr(P_s)-d_t'd_s]+2\sigma_v^4[tr(P_tP_s)-d_t'd_s] \\ &= \kappa_4d_t'd_s+\sigma_v^4+2\sigma_v^4tr(P_tP_s) \end{aligned}$$

For the second type in (55),

$$E_{t-1}(v_l'P_tv_lv_l')P_sv_q = \kappa_3d_t'P_sv_q \tag{56}$$

and for the third typein (55),

$$tr[P_tE_{t-1}(v_rv_r')P_sE_{t-1}(v_lv_l')] = \sigma_v^4tr(P_tP_s) \tag{57}$$

The results follow from the fact that

$$\begin{aligned} cov(v_l'P_tv_r, v_p'P_sv_q) &= E[cov_{t-1}(v_l'P_tv_r, v_p'P_sv_q)]+cov[E_{t-1}(v_l'P_tv_r), E_{t-1}(v_p'P_sv_q)] \\ &= E[cov_{t-1}(v_l'P_tv_r, v_p'P_sv_q)] \end{aligned} \tag{58}$$

The inequalities in the case of $l=r=p=q$ and $l=p\neq r=q$ in (50) are due to the Cauchy-Schwarz inequality

$$\begin{aligned} tr(P_tP_s) &= tr[z_t(z_t'z_t)^{-1}z_tz_s(z_s'z_s)^{-1}z_s] = tr\left(\frac{z_tz_t'z_sz_s'}{(z_t'z_t)(z_s'z_s)}\right) \\ &= \frac{z_t'z_s}{(z_t'z_t)(z_s'z_s)}tr(z_tz_s') = \frac{(z_t'z_s)^2}{(z_t'z_t)(z_s'z_s)} \leq 1 \end{aligned} \tag{59}$$

The proof of (51) will be omitted since it is the same as in Alvarez and Arellano (2003).

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

$$\frac{1}{NT^0}\sum_{t=3}^{T}w_{t-2}'P_t^{lm}w_{t-2} \to^p \rho_{lm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{60}$$

$$\frac{1}{NT^0}\sum_{t=4}^{T}w_{t-2}'P_t^{dm}w_{t-2} \to^p \rho_{dm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{61}$$

where

$$\rho_{lm} = \left[1 + \left(\frac{1+\alpha}{1-\alpha}\right) k\right]^{-1} \tag{62}$$

$$\rho_{dm} = \left(\frac{1-\alpha}{2}\right) \tag{63}$$

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

$$\frac{1}{NT^0} \sum_{t=4}^{T} w'_{t-2} P_t^{bm} w_{t-2} \to^p \left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{64}$$

Proof of (60)

After some algebra, we have

$$\begin{aligned} E\left(\frac{w'_{t-2}x_{t-1}}{N}\right) &= E(w_{i,t-2}y_{i,t-2}) = \frac{\sigma_v^2}{1-\alpha^2} && (65)\\ var\left(\frac{w'_{t-2}x_{t-1}}{N}\right) &= \frac{1}{N} var(w_{i,t-2}y_{i,t-2}) && (66)\\ E\left(\frac{x'_{t-1}x_{t-1}}{N}\right) &= E(y_{i,t-2}^2) = \frac{\sigma_\eta^2}{(1-\alpha)^2} + \frac{\sigma_v^2}{1-\alpha^2} && (67)\\ var\left(\frac{x'_{t-1}x_{t-1}}{N}\right) &= \frac{1}{N} var(y_{i,t-2}^2) && (68) \end{aligned}$$

Since we have assumed the finite fourth order moment of v_{it},

$$\begin{aligned} var(w_{i,t-2}y_{i,t-2}) &= E(w_{i,t-2}^4) + E(w_{i,t-2}^2)E(\mu_i^2) - \left(\frac{\sigma_v^2}{1-\alpha^2}\right)^2 \\ &= O(1) && (69)\\ var(y_{i,t-2}^2) &= E(\mu_i^4) + 2E(\mu_i^2)E(w_{i,t-2}^2) + E(w_{i,t-2}^4) - [E(y_{i,t-2}^2)]^2 \\ &= O(1) && (70) \end{aligned}$$

(66) and (68) tend to zero as N gets large. Thus, as N tends to infinity,

$$\begin{aligned} \frac{1}{NT^0} \sum_{t=3}^{T} w'_{t-2} P_t^{lm} w_{t-2} &\to^p \frac{1}{T^0} \sum_{t=3}^{T} E\left(w_{i,t-2}x_{i,t-1}\right)\left[E\left(x_{i,t-1}^2\right)\right]^{-1} E\left(x_{i,t-1}w_{i,t-2}\right) \\ &= \rho_{lm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \end{aligned}$$

Proof of (61)

We have

$$E\left(\frac{w'_{t-2}\Delta x_{t-1}}{N}\right) = E(w_{i,t-2}\Delta y_{i,t-2}) = \frac{\sigma_v^2}{1+\alpha} \tag{71}$$

$$var\left(\frac{w'_{t-2}\Delta x_{t-1}}{N}\right) = \frac{1}{N}var(w_{i,t-2}\Delta y_{i,t-2}) \quad (72)$$

$$E\left(\frac{\Delta x'_{t-1}\Delta x_{t-1}}{N}\right) = E(\Delta y_{i,t-2}^2) = \frac{2\sigma_v^2}{1+\alpha} \quad (73)$$

$$var\left(\frac{\Delta x'_{t-1}\Delta x_{t-1}}{N}\right) = \frac{1}{N}var(\Delta y_{i,t-2}^2) \quad (74)$$

Since we have assumed that v_{it} is finite up to the fourth order,

$$\begin{aligned} var(w_{i,t-2}\Delta y_{i,t-2}) &= E(w_{i,t-2}^4 + w_{i,t-2}^2 w_{i,t-3}^2 - 2w_{i,t-2}^2 w_{i,t-3}) - \left(\frac{\sigma_v^2}{1+\alpha}\right)^2 \\ &= O(1) \quad (75) \\ var(\Delta y_{i,t-2}^2) &= E(w_{i,t-2} - w_{i,t-3})^4 - \left(\frac{2\sigma_v^2}{1+\alpha}\right)^2 \\ &= O(1) \quad (76) \end{aligned}$$

Thus, as N gets large, (72) and (74) tend to zero. Therefore, as $N \to \infty$

$$\begin{aligned} \frac{1}{NT^0}\sum_{t=4}^{T} w'_{t-2}P_t^{dm}w_{t-2} &\to^p \frac{1}{T^0}\sum_{t=4}^{T} E(w_{i,t-2}\Delta y_{i,t-2})[E(\Delta y_{i,t-2}^2)]^{-1}E(\Delta y_{i,t-2}w_{i,t-2}) \\ &= \rho_{dm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \end{aligned}$$

Proof of (64)

To begin with, we provide some results which are useful in the proofs. Let $\phi_j = (1-\alpha^j)/(1-\alpha) = 1 + \alpha + \cdots + \alpha^{j-1}$ and $b_{t-1}^2 = (t-3)/(t-2)$. After some algebra, we have

$$E(w_{i,t-2}x_{i,t-1}^b) = b_{t-1}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left(1 - \frac{\alpha\phi_{t-3}}{t-3}\right) \quad (77)$$

$$\begin{aligned} E[(x_{i,t-1}^b)^2] &= b_{t-1}^2 E\left[w_{i,t-2} - \frac{1}{t-3}(w_{i,1} + \cdots + w_{i,t-3})\right]^2 \quad (78) \\ &= b_{t-1}^2\left[\frac{\sigma_v^2}{1-\alpha^2}\left(1 - \frac{2\alpha\phi_{t-3}}{t-3}\right) + \frac{1}{(t-3)^2}E(w_{i,1} + \cdots + w_{i,t-3})^2\right] \quad (79) \end{aligned}$$

Using the result of (A8) in Alvarez and Arellano (2003), we have

$$E(w_{i,1} + \cdots + w_{i,t-3})^2 = \frac{\sigma_v^2}{1-\alpha^2}\left[\frac{(t-3)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-3})}{(1-\alpha)^2}\right] \quad (80)$$

By substituting this term into (79), we get

$$E[(x_{i,t-1}^b)^2] = b_{t-1}^2\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left[1 - \frac{2\alpha\phi_{t-3}}{t-3} + \frac{1}{(t-3)^2}\left\{\frac{(t-3)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-3})}{(1-\alpha)^2}\right\}\right]$$

$$= b_{t-1}^2 \left(\frac{\sigma_v^2}{1-\alpha^2} \right) \Lambda_t \quad (81)$$

where Λ_t is given by (36).

Let ε_t denote the $N \times 1$ vector of errors of the population linear projection of w_{t-2} on Z_t^{bm}:

$$w_{t-2} = Z_t^{bm}\delta + \varepsilon_t \quad (82)$$

where $\delta = E(z_{it}^{bm} w_{i,t-2})/[E(z_{it}^{bm})^2] = E(x_{i,t-1}^b w_{i,t-2})/[E(x_{i,t-1}^b)^2]$. Then

$$\begin{aligned}
\varepsilon_{it} &= w_{i,t-2} - \delta x_{i,t-1}^b \quad (83)\\
&= \frac{\left[-\alpha\phi_{t-3} + \frac{(t-3)(1-\alpha)-2\alpha\phi_{t-3}}{(t-3)(1-\alpha)}\right] w_{i,t-2}}{(t-3)\Lambda_t} \\
&\quad + \frac{\left[\frac{(1-\alpha)(t-3)-\alpha(1-\alpha)\phi_{t-3}}{(1-\alpha)(t-3)}\right](1-\alpha)(w_{i,1} + \cdots + w_{i,t-3})}{(t-3)\Lambda_t} \\
&= \frac{\lambda_1 w_{i,t-2} + \lambda_2(1-\alpha)(w_{i,1} + \cdots + w_{i,t-3})}{\lambda_3} \\
&= \frac{\lambda_1 v_{i,t-2} + (\bar{\lambda}_1 + \lambda_2)v_{i,t-3} + \cdots + (\alpha^{t-5}\bar{\lambda}_1 + \lambda_2)v_{i,2} + (\alpha^{t-4}\bar{\lambda}_1 + \lambda_2)w_{i,1}}{\lambda_3} \quad (84)
\end{aligned}$$

where

$$\lambda_1 = -\alpha\phi_{t-3} + \frac{(t-3)(1-\alpha) - 2\alpha\phi_{t-3}}{(t-3)(1-\alpha)} \quad (85)$$

$$\lambda_2 = \frac{(1-\alpha)(t-3) - \alpha(1-\alpha)\phi_{t-3}}{(1-\alpha)(t-3)} \quad (86)$$

$$\lambda_3 = (t-3)\Lambda_3 \quad (87)$$

$$\bar{\lambda}_1 = \alpha(\lambda_1 - \lambda_2) = \alpha^2 \left(-\phi_{t-3} + \frac{(t-3) - \phi_{t-3}}{(t-3)(1-\alpha)} \right) \quad (88)$$

Since (84) is a linear combination of independent variables,

$$\begin{aligned}
E(\varepsilon_{it}^2) &= \frac{\sigma_v^2\left[(1-\alpha^2)\lambda_1^2 + \bar{\lambda}_1^2 + \{(t-4)(1-\alpha^2)+1\}\lambda_2^2 + 2(1+\alpha-\alpha^{t-3})\bar{\lambda}_1\lambda_2\right]}{(1-\alpha^2)\lambda_3^2} \\
&= O\left(\frac{1}{t}\right) \quad (89)
\end{aligned}$$

Now we consider the decomposition:

$$\begin{aligned}
w_{t-2}' P_t^{bm} w_{t-2} &= w_{t-2}' w_{t-2} - w_{t-2}'(I_N - P_t^{bm}) w_{t-2} \quad (90)\\
&= w_{t-2}' w_{t-2} - \varepsilon_t'(I_N - P_t^{bm})\varepsilon_t \quad (91)
\end{aligned}$$

The second equality is due to the fact that $(I_N - P_t^{bm})w_{t-2} = (I_N - P_t^{bm})(Z_t^{bm}\delta + \varepsilon_t)$.

Hence we have

$$\frac{1}{NT^0}\sum_{t=3}^{T} E(w_{t-2}' P_t^{bm} w_{t-2}) = E(w_{i,t-2}^2) - \frac{1}{NT^0}\sum_{t=3}^{T} E(\varepsilon_t'(I_N - P_t^{bm})\varepsilon_t) \quad (92)$$

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

$$\frac{1}{NT^0}\sum_{t=3}^{T} E(\varepsilon_t'(I_N - P_t^{bm})\varepsilon_t) \leq \frac{1}{NT^0}\sum_{t=3}^{T} E(\varepsilon_t'\varepsilon_t) = \frac{1}{T^0}\sum_{t=3}^{T} E(\varepsilon_{i,t}^2) = \frac{1}{T^0}O(\log T) \to 0 \quad (93)$$

Hence, as $T \to \infty$,

$$\frac{1}{NT^0}\sum_{t=3}^{T} E(w_{t-2}' P_t^{bm} w_{t-2}) \to E(w_{i,t-2}^2) = \frac{\sigma_v^2}{1-\alpha^2} \quad (94)$$

With regards to the proofs that the variance of $(NT^0)^{-1}\sum_{t=3}^{T} w_{t-2}'w_{t-2}$ and $(NT^0)^{-1}\sum_{t=3}^{T} \varepsilon_t'\varepsilon_t$ tend to zero, see Alvarez and Arellano (2003).

Lemma 4. *Under Assumptions 1, 2, and 3, the following results hold:*

$$E(x^{*'} P^{lm} v^*) = \left(\frac{\sigma_v^2}{1-\alpha}\right)\left(\frac{\phi_{T-1}}{T-1} - 1\right) \quad (95)$$

$$E(x^{*'} P^{dm} v^*) = E(x^{*'} P^{bm} v^*) = \left(\frac{\sigma_v^2}{1-\alpha}\right)\left(\frac{\phi_{T-2}}{T-2} - 1\right) \quad (96)$$

Proof of (95)

Let P_t denote P_t^{lm} or P_t^{dm} or P_t^{bm} Decompose x_t^* as

$$x_t^* = \psi_t w_{t-2} - c_t \tilde{v}_{t-1,T-1} \quad (97)$$

$$\psi_t = c_t\left(1 - \frac{\alpha\phi_{T-t+1}}{T-t+1}\right) \quad (98)$$

$$\tilde{v}_{t-1,T-1} = \frac{1}{T-t+1}(\phi_{T-t+1}v_{t-1} + \cdots + \phi_1 v_{T-1}) \quad (99)$$

Following Alvarez and Arellano (2003), by using $\phi_j = \phi_{j-1} + \alpha^{j-1}$, and $\phi_1 + \cdots + \phi_{j-1} = (j - \phi_j)/(1-\alpha)$, we get

$$\begin{aligned}
E(x_t^{*'} P_t v_t^*) &= E(\psi_t w_{t-2}' P_t v_t^*) - E(c_t \tilde{v}_{t-1,T-1}' P_t v_t^*) \\
&= -E(c_t \tilde{v}_{t-1,T-1}' P_t v_t^*) \\
&= \frac{-\sigma_v^2}{T-t+2} tr(P_t)\left[\phi_{T-t+1} - \frac{\phi_{T-t} + \cdots + \phi_1}{T-t+1}\right] \\
&= \frac{-\sigma_v^2}{T-t+2} tr(P_t)\left[\frac{\phi_{T-t+1}}{T-t+1} - \frac{\phi_{T-t+2}}{T-t+2}\right] \quad (100)
\end{aligned}$$

Since $tr(P_t) = 1$, the result follows.

Lemma 5. *Let Assumptions 1, 2, 3 hold. Then, as both N and T tend to infinity,*

$$var\left(\frac{x^{*\prime}P^{lm}v^*}{\sqrt{NT^0}}\right) \rightarrow \rho_{lm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{101}$$

$$var\left(\frac{x^{*\prime}P^{dm}v^*}{\sqrt{NT^0}}\right) \rightarrow \rho_{dm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{102}$$

$$var\left(\frac{x^{*\prime}P^{bm}v^*}{\sqrt{NT^0}}\right) \rightarrow \left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{103}$$

Proof of (101)-(103)

Let P_t denote the generic projection matrices which can be P_t^{lm}, P_t^{dm}, P_t^{bm}. By using the fact that $v_t^* = (v_{t-1} - \bar{v}_{t-1,T})/c_t$, we get the following decomposition:

$$\frac{1}{\sqrt{NT^0}}x^{*\prime}Pv^* = \left(\frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} w_{t-2}'P_t v_{t-1} - \Upsilon_{11NT} - \Upsilon_{12NT}\right) - (\Upsilon_{21NT} - \Upsilon_{22NT}) \tag{104}$$

where

$$\bar{v}_{t-1,T} = \frac{v_{t-1} + \cdots + v_T}{T-t+2} \tag{105}$$

$$\Upsilon_{11NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} w_{t-2}'P_t\bar{v}_{t-1,T} \tag{106}$$

$$\Upsilon_{12NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} \frac{c_t\alpha\phi_{T-t+1}}{T-t+1}w_{t-2}'P_t v_t^* \tag{107}$$

$$\Upsilon_{21NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} \tilde{v}_{t-1,T-1}'P_t v_{t-1} \tag{108}$$

$$\Upsilon_{22NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} \tilde{v}_{t-1,T-1}'P_t\bar{v}_{t-1,T} \tag{109}$$

The variance of the leading term in (104) is

$$var\left(\frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} w_{t-2}'P_t v_{t-1}\right) = \frac{1}{NT^0}\sum_{t=t_0}^{T} var(w_{t-2}'P_t v_{t-1}) = \frac{1}{NT^0}\sum_{t=t_0}^{T} E(w_{t-2}'P_t w_{t-2}) \tag{110}$$

This is because for $t > s$, $cov(w_{t-2}'P_t v_{t-1}, w_{s-2}'P_s v_{s-1}) = 0$. Thus, the leading term converges to the form obtained by Lemma 3. Next, since $E(w_{t-2}'P_t\bar{v}_{t-1,T}) = E(\bar{v}_{s-1,T}'P_s w_{s-2}) = 0$,

$$var(\Upsilon_{11NT}) = \frac{1}{NT^0}\sum_{t=t_0}^{T}\sum_{s=t_0}^{T} E(w_{t-2}'P_t\bar{v}_{t-1,T}\bar{v}_{s-1,T}'P_s w_{s-2})$$

$$
\begin{aligned}
&= \frac{1}{NT^0}\sum_{t=t_0}^{T}\sum_{s=t_0}^{T} cov(w_{t-2}' P_t \bar{v}_{t-1,T}, \bar{v}_{s-1,T}' P_s w_{s-2}) \\
&\leq \frac{1}{NT^0}\sum_{t=t_0}^{T}\sum_{s=t_0}^{T} \sqrt{E(w_{t-2}' P_t \bar{v}_{t-1,T})^2}\sqrt{E(\bar{v}_{s-1,T}' P_s w_{s-2})^2} \qquad (111)
\end{aligned}
$$

Since the maximum eigenvalue of P_t is equal to 1, we have

$$
\begin{aligned}
E(w_{t-2}' P_t \bar{v}_{t-1,T})^2 &= \frac{\sigma_v^2}{T-t+2} E(w_{t-2}' P_t w_{t-2}) \\
&\leq \frac{\sigma_v^2}{T-t+2} E(w_{t-2}' w_{t-2}) \\
&= \frac{N\sigma_v^2}{(T-t+2)} E(w_{i,t-2}^2) = \left(\frac{\sigma_v^4}{1-\alpha^2}\right)\left(\frac{N}{T-t+2}\right) \qquad (112)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
var(\Upsilon_{11NT}) &\leq \left(\frac{\sigma_v^4}{1-\alpha^2}\right)\frac{1}{T^0}\sum_{t=t_0}^{T}\sum_{s=t_0}^{T}\sqrt{\frac{1}{T-t+2}}\sqrt{\frac{1}{T-s+2}} \\
&= \left(\frac{\sigma_v^4}{1-\alpha^2}\right)\frac{1}{T^0}\sum_{t=t_0}^{T}\sqrt{\frac{1}{T-t+2}}\sum_{s=t_0}^{T}\sqrt{\frac{1}{T-s+2}} \\
&= \left(\frac{\sigma_v^4}{1-\alpha^2}\right)\frac{O(\log T)}{T^0} \to 0 \qquad (113)
\end{aligned}
$$

Next, by using the result that for $t > s$, $E_{t-1}(v_{t-1}^* v_{s-1}^{*'}) = 0$, and $cov(w_{t-2}' P_t v_t^*, w_{s-2}' P_s v_s^*) = 0$, we have

$$
\begin{aligned}
var(\Upsilon_{12NT}) &= \frac{1}{NT^0}\sum_{t=t_0}^{T}\frac{\alpha^2\phi_{T-t+1}^2}{(T-t+1)(T-t+2)} var(w_{t-2}' P_t v_t^*) \\
&= \frac{\sigma_v^2}{NT^0}\sum_{t=t_0}^{T}\frac{\alpha^2\phi_{T-t+1}^2}{(T-t+1)(T-t+2)} E(w_{t-2}' P_t w_{t-2}) \\
&\leq \left(\frac{\sigma_v^4}{1-\alpha^2}\right)\frac{1}{T^0}\sum_{t=t_0}^{T}\frac{\alpha^2\phi_{T-t+1}^2}{(T-t+1)(T-t+2)} \to 0 \qquad (114)
\end{aligned}
$$

We then turn to consider $var(\Upsilon_{21NT})$

$$
\begin{aligned}
var(\Upsilon_{21NT}) &= \frac{1}{NT^0} var\left[\sum_{t=t_0}^{T}\frac{1}{T-t+1} v_{t-1}' P_t(\phi_{T-t+1} v_{t-1} + \cdots + \phi_1 v_{T-1})\right] \\
&= a_{0NT} + a_{1NT} \qquad (115)
\end{aligned}
$$

where

$$
a_{0NT} = \frac{1}{NT^0}\sum_{t=t_0}^{T}\frac{\phi_{T-t+1}^2 var(v_{t-1}' P_t v_{t-1}) + \cdots + \phi_1^2 var(v_{t-1}' P_t v_{T-1})}{(T-t+1)^2}
$$

$$= \frac{1}{NT^0}\sum_{t=t_0}^{T}\frac{\phi_{T-t+1}^2[2\sigma_v^4 tr(P_tP_s)+\kappa_4E(d_t'd_s)]+(\phi_{T-t}^2+\cdots+\phi_1^2)tr(P_tP_s)\sigma_v^4}{(T-t+1)^2}$$

and

$$\begin{aligned} a_{1NT} &= \frac{2}{NT^0}\sum_{t=t_0}^{T-1}\left[\frac{\phi_{T-t}^2 cov(v_{t-1}'P_tv_t, v_t'P_{t+1}v_t)}{(T-t+1)(T-t)}\right. \\ &\qquad \left. +\cdots+\frac{\phi_1^2 cov(v_{t-1}'P_tv_{T-1}, v_{T-1}'P_Tv_{T-1})}{(T-t+1)}\right] \\ &= \frac{2}{NT^0}\sum_{t=t_0}^{T-1}\left[\frac{\phi_{T-t}^2\kappa_3E(d_{t+1}'P_tv_{t-1})}{(T-t+1)(T-t)}+\cdots+\frac{\phi_1^2\kappa_3E(d_T'P_tv_{t-1})}{(T-t+1)}\right] \end{aligned}$$

Using the fact that $\phi_j^2 < 1/(1-\alpha)^2$,

$$\begin{aligned} a_{0NT} &\leq \frac{1}{NT^0}\sum_{t=t_0}^{T}\frac{\phi_{T-t+1}^2[2\sigma_v^4+\kappa_4]+(\phi_{T-t}^2+\cdots+\phi_1^2)\sigma_v^4}{(T-t+1)^2} \\ &\leq \frac{1}{(1-\alpha)^2}\frac{1}{NT^0}\sum_{t=t_0}^{T}\frac{[2\sigma_v^4+\kappa_4]+(T-t)\sigma_v^4}{(T-t+1)^2} \\ &= \frac{2\sigma_v^4+\kappa_4}{(1-\alpha)^2}\frac{1}{NT^0}\sum_{t=3}^{T}\frac{1}{(T-t+1)^2}+\frac{\sigma_v^4}{(1-\alpha)^2}\frac{1}{NT^0}\sum_{t=3}^{T}\frac{T-t}{(T-t+1)^2} \\ &\to 0 \end{aligned} \tag{116}$$

In view of the triangle inequality and the fact that $|E(d_{t+j}'P_tv_t)| \leq \sigma_v$,

$$\begin{aligned} |a_{1NT}| &\leq \frac{2|\kappa_3|\sigma_v}{(1-\alpha)^2}\frac{1}{NT^0}\sum_{t=t_0}^{T-1}\left[\frac{1}{(T-t+1)(T-t)}+\cdots+\frac{1}{(T-t+1)}\right] \\ &= \frac{2|\kappa_3|\sigma_v}{(1-\alpha)^2}\frac{1}{NT^0}O(logT)\to 0 \end{aligned} \tag{117}$$

Finally we consider the term Υ_{22NT}. We decompose the variance of Υ_{22NT} as follows:

$$var(\Upsilon_{22NT}) = \frac{1}{NT^0}var\left(\sum_{t=t_0}^{T}\bar{v}_{t-1,T}'P_t\tilde{v}_{t-1,T-1}\right) = b_{0NT}+b_{1NT} \tag{118}$$

where

$$b_{0NT} = \frac{1}{NT^0}\sum_{t=t_0}^{T}var(\bar{v}_{t-1,T}'P_t\tilde{v}_{t-1,T-1}) \tag{119}$$

and

$$b_{1NT} = \frac{2}{NT^0}\sum_{s}\sum_{s>t}cov(\bar{v}_{t-1,T}'P_t\tilde{v}_{t-1,T-1}, \bar{v}_{s-1,T}'P_s\tilde{v}_{s-1,T-1}) \tag{120}$$

From (A73) in Alvarez and Arellano (2003), we have

$$var(\bar{v}'_{t-1,T}P_t\tilde{v}_{t-1,T-1}) = O\left(\frac{1}{(T-t+1)^2}\right) \tag{121}$$

Next, with regard to the term b_{1NT}, we have

$$\begin{aligned} |b_{1NT}| &\leq \frac{2}{NT^0}\sum_s\sum_{s>t}|cov(\bar{v}'_{t-1,T}P_t\tilde{v}_{t-1,T-1},\bar{v}'_{s-1,T}P_s\tilde{v}_{s-1,T-1})| \\ &\leq \frac{2}{NT^0}\sum_s\sum_{s>t}\sqrt{var(\bar{v}'_{t-1,T}P_t\tilde{v}_{t-1,T-1})}\sqrt{var(\bar{v}'_{s-1,T}P_s\tilde{v}_{s-1,T-1})} \\ &\leq \frac{2}{NT^0}\sum_s O\left(\frac{1}{T-t}\right)\sum_t O\left(\frac{1}{T-s}\right) = O\left(\frac{(\log T)^2}{NT}\right) \to 0 \end{aligned} \tag{122}$$

Lemma 6. *Let Assumptions 1, 2, 3 hold. Then as both N and T tend to infinity,*

$$\frac{x^{*\prime}P^{lm}x^*}{NT^0} \to^p \rho_{lm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{123}$$

$$\frac{x^{*\prime}P^{dm}x^*}{NT^0} \to^p \rho_{dm}\left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{124}$$

$$\frac{x^{*\prime}P^{bm}x^*}{NT^0} \to^p \left(\frac{\sigma_v^2}{1-\alpha^2}\right) \tag{125}$$

Proof of (123)

Using the decomposition of x_t^*, we have

$$\begin{aligned} \frac{x^{*\prime}Px^*}{NT^0} &= \frac{1}{NT^0}\sum_{t=t_0}^T\psi_t^2 w'_{t-2}P_t w_{t-2} - \frac{2}{NT^0}\sum_{t=t_0}^T c_t\psi_t w'_{t-2}P_t\tilde{v}_{t-1,T-1} \\ &\quad + \frac{1}{NT^0}\sum_{t=t_0}^T c_t^2\tilde{v}'_{t-1,T-1}P_t\tilde{v}_{t-1,T-1} \end{aligned}$$

Since $\psi_t^2 = 1 - O(1/(T-t+1))$, the first term of the right-hand side converges to the form obtained by Lemma 3. The second term has zero mean and, by using similar arguments as those used for Υ_{11NT}, it can be shown that its variance tends to zero. The third term is analogous to Υ_{22NT}. Its mean is given by

$$\begin{aligned} E\left(\frac{1}{NT^0}\sum_{t=t_0}^T c_t^2\tilde{v}'_{t-1,T-1}P_t\tilde{v}_{t-1,T-1}\right) &= \frac{1}{NT^0}\sum_{t=t_0}^T c_t^2 E(\tilde{v}^2_{i,t-1,T-1}) \\ &= \frac{\sigma_v^2}{NT^0}\sum_{t=t_0}^T\frac{\phi^2_{T-t+1}+\cdots+\phi_1^2}{(T-t+2)(T-t+1)} \\ &= O\left(\frac{\log T}{NT}\right) \to 0 \end{aligned} \tag{126}$$

and its variance is shown to tend to zero in the same way as Υ_{22NT}.

Proof of Theorems 2 and 3

Consistency directly follows from Lemmas 4, 5, and 6. Next, we show the asymptotic normality of $\hat{\alpha}_{lm}$ and $\hat{\alpha}_{dm}$. Here, let P denote P^{lm} or P^{dm}, and z_{it} let denote z_{it}^{lm} or z_{it}^{dm}. In addition, ρ denotes ρ_{lm} or ρ_{dm}. Since we have shown that the variances of Υ_{11NT}, Υ_{12NT}, Υ_{21NT}, and Υ_{22NT} tend to zero,

$$\frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} x_t^{*'} P_t v_t^* - \mu_{NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} w_{t-2}' P_t v_{t-1} + o_p(1) \to^d \left(0, \rho\left(\frac{\sigma_v^4}{1-\alpha^2}\right)\right) \tag{127}$$

where

$$\mu_{NT} = \frac{1}{\sqrt{NT^0}} E(x^{*'} P v^*) \tag{128}$$

This is because

$$\begin{aligned}
\frac{1}{\sqrt{NT^0}}\sum_{t=t_0}^{T} w_{t-2}' z_t (z_t' z_t)^{-1} z_t' v_{t-1} &= \frac{1}{\sqrt{T^0}}\sum_{t=t_0}^{T}\left(\frac{w_{t-2}' z_t}{N}\right)\left(\frac{z_t' z_t}{N}\right)^{-1}\frac{z_t' v_{t-1}}{\sqrt{N}} \\
&\to^d \frac{1}{\sqrt{T^0}}\sum_{t=t_0}^{T}\frac{E(w_{i,t-2} z_{it})}{E(z_{it}^2)}\xi_t \\
&= \frac{E(w_{i,t-2} z_{it})}{E(z_{it}^2)}\frac{1}{\sqrt{T}}\sum_{t=t_0}^{T}\xi_t \\
&\to^d \frac{E(w_{i,t-2} z_{it})}{E(z_{it}^2)} N(0, var(\xi_t))
\end{aligned} \tag{129}$$

where $\xi_t \sim N(0, \sigma_v^2 E(z_{it}^2))$. Then (127) holds. Therefore:

$$\begin{aligned}
\left(\frac{x^{*'} P x^*}{NT^0}\right)^{-1}\left(\frac{1}{\sqrt{NT^0}} x^{*'} P v^* - \mu_{NT}\right) &= \sqrt{NT^0}(\hat{\alpha} - \alpha) - \left(\frac{x^{*'} P x^*}{NT^0}\right)^{-1}\mu_{NT} \\
&= \sqrt{NT^0}(\hat{\alpha} - \alpha) + (1+\alpha)\frac{\rho^{-1}}{\sqrt{NT^0}}\left(1 - \frac{\phi_{T-1}}{T-1}\right) \\
&= \sqrt{NT^0}\left[\hat{\alpha} - \left(\alpha - \frac{(1+\alpha)\rho^{-1}}{NT^0}\right)\right] + O\left(\frac{1}{\sqrt{NT^3}}\right) \\
&\to^d N\left(0, (1-\alpha^2)\rho^{-1}\right)
\end{aligned}$$

Proof of Theorem 4

It is straightforward to show consistency from Lemmas 4, 5, 6. Next, we show the asymptotic normality of $\hat{\alpha}_{bm}$. To begin with, note that

$$\frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} x_t^{*'} P_t^{bm} v_t^* - \mu_{NT} = \frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} w_{t-2}' P_t^{bm} v_{t-1} + o_p(1)$$

$$= \frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} w'_{t-2}v_{t-1} - \frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} w'_{t-2}(I_N - P_t)v_{t-1} + o_p(1) \tag{130}$$

The second term in (130) is $o_p(1)$. This is because it has zero mean, and from Lemma 3 we know that its variance is

$$\begin{aligned} var\left(\frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} w'_{t-2}(I_N - P_t)v_{t-1}\right) &= \frac{1}{NT^0}\sum_{t=4}^{T} var[w'_{t-2}(I_N - P_t)v_{t-1}] \\ &= \frac{\sigma_v^2}{NT^0}\sum_{t=4}^{T} E[w'_{t-2}(I_N - P_t)w_{t-2}] \\ &= \frac{\sigma_v^2}{NT^0}\sum_{t=4}^{T} E[\varepsilon'_t(I_N - P_t)\varepsilon_t] \\ &= \frac{1}{T^0}O(\log T) \to 0 \end{aligned} \tag{131}$$

Therefore,

$$\frac{1}{\sqrt{NT^0}}\sum_{t=4}^{T} x_t^{*\prime} P_t^{bm} v_t^* - \mu_{NT} = \frac{1}{\sqrt{N(T-3)}}\sum_{t=4}^{T} w'_{t-2}v_{t-1} + o_p(1) \to^d N\left(0, \frac{\sigma_v^4}{(1-\alpha^2)}\right) \tag{132}$$

Using Cramer's theorem, we get the following result:

$$\begin{aligned} \left(\frac{x^{*\prime}P^{bm}x^*}{NT^0}\right)^{-1}\left(\frac{x^{*\prime}P^{bm}v^*}{\sqrt{NT^0}} - \mu_{NT}\right) &= \sqrt{NT^0}(\hat{\alpha}-\alpha) + (1+\alpha)\frac{1}{\sqrt{NT^0}}\left(1 - \frac{\phi_{T-1}}{T-1}\right) \\ &= \sqrt{NT^0}\left[\hat{\alpha} - \left(\alpha - \frac{(1+\alpha)}{NT^0}\right)\right] + O\left(\frac{1}{\sqrt{NT^3}}\right) \\ &\to^d N\left(0, 1-\alpha^2\right) \end{aligned}$$

Proof of Lemma 1

The asymptotic variance with a large N and a fixed T is given by

$$Avar(\hat{\alpha}) = \sigma_v^2\left[N\sum_{t=t_0}^{T} E(x_{i,t}^* z'_{it})[E(z_{it}z'_{it})]^{-1}E(z'_{it}x_{it}^*)\right]^{-1} \tag{133}$$

Since it is straightforward to obtain (31), (32) and (33) by using the proofs of Lemma 3, the proofs of (31), (32) and (33) will be omitted. We only consider the case of (30). After some algebra, we get

$$E(x_{it}^* z_{it}^{la\prime}) = \psi_t\left(\frac{\sigma_v^2}{1-\alpha^2}\right)(\alpha^{t-3}, \cdots, 1) \tag{134}$$

Next,

$$[E(z_{it}^{la}z_{it}^{la'})]^{-1} = \begin{bmatrix} \sigma_\mu^2 + \left(\frac{1}{1-\alpha^2}\right)\sigma_v^2 & \cdots & \sigma_\mu^2 + \left(\frac{\alpha^{t-3}}{1-\alpha^2}\right)\sigma_v^2 \\ \vdots & & \vdots \\ \sigma_\mu^2 + \left(\frac{\alpha^{t-3}}{1-\alpha^2}\right)\sigma_v^2 & \cdots & \sigma_\mu^2 + \left(\frac{1}{1-\alpha^2}\right)\sigma_v^2 \end{bmatrix}^{-1} \quad (135)$$

$$= \frac{1}{\sigma_v^2}\left[(\sqrt{\lambda}\iota_{t-2})(\sqrt{\lambda}\iota_{t-2})' + V_{t-2}\right]^{-1} \quad (136)$$

where $\mu_i = \eta_i/(1-\alpha)$ with variance σ_μ^2, $\lambda = \sigma_\mu^2/\sigma_v^2$, ι_{t-2} is a $(t-2)$ dimensional column vector of ones, and V_{t-2} is

$$V_{t-2} = \frac{1}{1-\alpha^2}\begin{bmatrix} 1 & \cdots & \alpha^{t-3} \\ \vdots & & \vdots \\ \alpha^{t-3} & \cdots & 1 \end{bmatrix} \quad (137)$$

By using the decomposition of V_{t-2} and the fact that[13]

$$[A + bb']^{-1} = A^{-1} - \left[\frac{1}{1 + b'A^{-1}b}\right] A^{-1}bb'A^{-1} \quad (138)$$

we get

$$[E(z_{it}^{la}z_{it}^{la'})]^{-1} = \sigma_v^{-2}\left[L'L - \frac{\lambda}{1 + \lambda\iota_{t-2}'L'L\iota_{t-2}}L'L\iota_{t-2}\iota_{t-2}'L'L\right] \quad (139)$$

where

$$V_{t-2}^{-1} = L'L \quad (140)$$

and

$$L = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix} \quad (141)$$

By using (134) and (139), the result follows.

[13]See Amemiya (1985, p.164), Hamilton (1994, p.120) and Greene (2001, p.822).

Table 1: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$ ($N = 50$)

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.171	0.195	0.171	0.196	0.176	0.199	0.448	0.485	0.455	0.491	0.463	0.496	0.670	0.744	0.704	0.767	0.735	0.783
std	0.064	0.069	0.046	0.050	0.034	0.038	0.073	0.084	0.050	0.058	0.033	0.039	0.102	0.142	0.061	0.085	0.035	0.052
se	0.063	0.068	0.046	0.050	0.032	0.036	0.071	0.081	0.048	0.056	0.032	0.039	0.095	0.127	0.056	0.079	0.031	0.047
rmse	0.070	0.069	0.055	0.050	0.042	0.038	0.090	0.085	0.067	0.059	0.050	0.039	0.165	0.152	0.114	0.092	0.074	0.055
size	0.079	0.056	0.098	0.037	0.125	0.061	0.122	0.062	0.149	0.059	0.215	0.050	0.283	0.081	0.401	0.067	0.536	0.071
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.161	0.185	0.168	0.190	0.172	0.194	0.435	0.460	0.449	0.477	0.457	0.486	0.606	0.619	0.680	0.701	0.725	0.748
std	0.071	0.093	0.051	0.073	0.033	0.051	0.081	0.131	0.053	0.097	0.034	0.061	0.127	0.270	0.071	0.166	0.036	0.099
se	0.069	0.094	0.049	0.069	0.033	0.050	0.082	0.126	0.053	0.088	0.033	0.060	0.113	0.234	0.063	0.145	0.033	0.085
rmse	0.081	0.094	0.060	0.073	0.043	0.051	0.104	0.137	0.074	0.099	0.055	0.063	0.232	0.325	0.140	0.194	0.083	0.112
size	0.097	0.048	0.120	0.055	0.139	0.056	0.111	0.057	0.148	0.069	0.242	0.044	0.404	0.088	0.455	0.066	0.619	0.079
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.159	0.122	0.165	0.142	0.170	0.172	0.424	0.325	0.442	0.387	0.458	0.430	0.578	0.467	0.666	0.572	0.723	0.663
std	0.074	0.220	0.051	0.165	0.034	0.116	0.094	0.288	0.058	0.209	0.034	0.145	0.126	0.323	0.072	0.243	0.038	0.169
se	0.075	0.215	0.051	0.160	0.034	0.112	0.089	0.288	0.055	0.204	0.034	0.138	0.122	0.362	0.066	0.240	0.034	0.152
rmse	0.085	0.233	0.062	0.175	0.045	0.120	0.121	0.337	0.082	0.238	0.054	0.161	0.256	0.464	0.153	0.333	0.085	0.218
size	0.092	0.045	0.093	0.046	0.143	0.061	0.143	0.037	0.192	0.035	0.223	0.060	0.458	0.052	0.510	0.051	0.616	0.067

Table 2: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{la}$, $\hat{\alpha}_{lm}$ ($N = 100$)

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\alpha}_{la}$	$\hat{\alpha}_{lm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.186	0.198	0.186	0.199	0.186	0.199	0.473	0.493	0.475	0.493	0.479	0.497	0.718	0.769	0.743	0.784	0.761	0.792
std	0.046	0.048	0.033	0.036	0.024	0.026	0.053	0.057	0.034	0.039	0.023	0.027	0.075	0.091	0.044	0.056	0.025	0.034
se	0.046	0.048	0.033	0.036	0.023	0.026	0.052	0.057	0.035	0.040	0.023	0.027	0.073	0.090	0.043	0.055	0.024	0.033
rmse	0.048	0.048	0.036	0.036	0.027	0.026	0.059	0.058	0.042	0.039	0.031	0.027	0.111	0.096	0.072	0.058	0.047	0.035
size	0.069	0.057	0.065	0.045	0.089	0.045	0.082	0.058	0.092	0.042	0.135	0.046	0.201	0.052	0.255	0.059	0.366	0.056
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.179	0.189	0.183	0.196	0.186	0.199	0.460	0.476	0.470	0.486	0.477	0.492	0.676	0.698	0.725	0.748	0.754	0.769
std	0.052	0.070	0.035	0.049	0.025	0.035	0.062	0.097	0.039	0.063	0.025	0.045	0.094	0.204	0.054	0.118	0.027	0.070
se	0.050	0.067	0.035	0.049	0.024	0.035	0.061	0.090	0.039	0.063	0.025	0.043	0.091	0.172	0.050	0.104	0.026	0.061
rmse	0.056	0.070	0.039	0.049	0.028	0.035	0.074	0.100	0.049	0.064	0.034	0.045	0.155	0.229	0.093	0.129	0.053	0.077
size	0.073	0.054	0.070	0.052	0.094	0.045	0.099	0.066	0.110	0.052	0.149	0.063	0.271	0.074	0.322	0.067	0.429	0.063
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.178	0.144	0.183	0.167	0.185	0.178	0.458	0.367	0.472	0.428	0.477	0.457	0.653	0.487	0.715	0.606	0.752	0.691
std	0.056	0.166	0.037	0.124	0.024	0.089	0.066	0.260	0.042	0.164	0.027	0.114	0.108	0.336	0.057	0.238	0.028	0.160
se	0.054	0.161	0.037	0.118	0.025	0.085	0.066	0.235	0.041	0.157	0.025	0.107	0.101	0.347	0.053	0.224	0.027	0.137
rmse	0.060	0.175	0.040	0.128	0.028	0.091	0.078	0.292	0.050	0.179	0.035	0.122	0.183	0.459	0.102	0.307	0.055	0.194
size	0.069	0.048	0.066	0.053	0.089	0.051	0.088	0.049	0.106	0.046	0.161	0.051	0.313	0.058	0.358	0.051	0.409	0.072

Table 3: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{da}$, $\hat{\alpha}_{dm}$, ($N = 50$)

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.164	0.194	0.168	0.194	0.173	0.201	0.430	0.484	0.443	0.487	0.458	0.495	0.597	0.710	0.671	0.747	0.724	0.778
std	0.077	0.101	0.053	0.074	0.036	0.052	0.093	0.136	0.057	0.090	0.036	0.063	0.145	0.224	0.075	0.132	0.039	0.084
se	0.083	0.108	0.054	0.075	0.035	0.052	0.100	0.144	0.059	0.094	0.035	0.062	0.144	0.245	0.072	0.140	0.035	0.082
rmse	0.085	0.101	0.062	0.074	0.044	0.052	0.117	0.137	0.081	0.091	0.055	0.063	0.250	0.241	0.149	0.142	0.085	0.087
size	0.059	0.039	0.084	0.044	0.116	0.048	0.090	0.037	0.140	0.039	0.225	0.049	0.269	0.057	0.407	0.041	0.546	0.071
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.160	0.189	0.167	0.194	0.172	0.199	0.430	0.480	0.446	0.489	0.457	0.495	0.584	0.696	0.671	0.750	0.723	0.776
std	0.080	0.107	0.053	0.072	0.034	0.050	0.095	0.139	0.058	0.090	0.036	0.060	0.150	0.223	0.079	0.139	0.037	0.080
se	0.083	0.108	0.054	0.075	0.035	0.052	0.100	0.144	0.059	0.094	0.035	0.062	0.144	0.240	0.072	0.140	0.036	0.082
rmse	0.090	0.107	0.063	0.073	0.044	0.050	0.118	0.140	0.079	0.091	0.056	0.060	0.264	0.246	0.151	0.147	0.086	0.084
size	0.081	0.047	0.097	0.048	0.113	0.041	0.091	0.037	0.118	0.038	0.226	0.045	0.318	0.071	0.420	0.058	0.576	0.055
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.162	0.195	0.167	0.194	0.171	0.197	0.432	0.483	0.445	0.487	0.460	0.494	0.579	0.683	0.668	0.745	0.724	0.782
std	0.075	0.101	0.051	0.070	0.034	0.051	0.097	0.135	0.059	0.090	0.034	0.061	0.144	0.222	0.076	0.138	0.039	0.079
se	0.083	0.108	0.054	0.075	0.035	0.052	0.100	0.144	0.059	0.094	0.035	0.062	0.145	0.242	0.073	0.141	0.035	0.082
rmse	0.084	0.101	0.061	0.070	0.045	0.051	0.119	0.136	0.081	0.091	0.053	0.061	0.264	0.251	0.152	0.149	0.085	0.081
size	0.055	0.043	0.072	0.036	0.126	0.050	0.091	0.034	0.138	0.044	0.196	0.042	0.318	0.059	0.439	0.066	0.562	0.043

Table 4: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{da}$, $\hat{\alpha}_{dm}$, $(N = 100)$

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$	$\hat{\alpha}_{da}$	$\hat{\alpha}_{dm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.181	0.198	0.182	0.198	0.185	0.198	0.458	0.488	0.469	0.492	0.476	0.494	0.661	0.742	0.717	0.771	0.752	0.786
std	0.056	0.072	0.037	0.051	0.025	0.037	0.069	0.096	0.042	0.066	0.025	0.042	0.112	0.175	0.058	0.103	0.028	0.059
se	0.060	0.077	0.039	0.054	0.025	0.037	0.074	0.104	0.044	0.068	0.026	0.045	0.115	0.186	0.057	0.107	0.028	0.062
rmse	0.059	0.072	0.041	0.051	0.029	0.037	0.081	0.097	0.053	0.067	0.034	0.042	0.178	0.185	0.101	0.107	0.056	0.061
size	0.047	0.027	0.058	0.034	0.076	0.060	0.075	0.030	0.098	0.042	0.134	0.036	0.203	0.051	0.281	0.053	0.385	0.048
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.178	0.196	0.183	0.197	0.186	0.200	0.456	0.484	0.469	0.490	0.477	0.495	0.659	0.727	0.718	0.766	0.752	0.786
std	0.058	0.075	0.038	0.051	0.026	0.035	0.070	0.097	0.043	0.067	0.026	0.045	0.112	0.166	0.060	0.104	0.028	0.059
se	0.060	0.077	0.039	0.054	0.025	0.037	0.074	0.104	0.044	0.068	0.026	0.045	0.115	0.184	0.057	0.107	0.028	0.062
rmse	0.062	0.075	0.042	0.051	0.029	0.035	0.083	0.098	0.053	0.067	0.035	0.045	0.180	0.181	0.102	0.109	0.055	0.061
size	0.066	0.042	0.059	0.040	0.083	0.036	0.083	0.042	0.093	0.055	0.137	0.059	0.225	0.055	0.268	0.065	0.390	0.053
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.181	0.197	0.184	0.199	0.185	0.199	0.463	0.491	0.474	0.497	0.477	0.497	0.664	0.744	0.719	0.774	0.753	0.788
std	0.057	0.073	0.037	0.051	0.024	0.036	0.068	0.095	0.042	0.068	0.027	0.044	0.116	0.170	0.059	0.102	0.028	0.059
se	0.060	0.077	0.039	0.054	0.025	0.037	0.073	0.104	0.044	0.068	0.026	0.045	0.117	0.188	0.057	0.107	0.028	0.062
rmse	0.060	0.073	0.040	0.051	0.028	0.036	0.077	0.095	0.050	0.068	0.035	0.044	0.179	0.179	0.101	0.105	0.054	0.060
size	0.043	0.038	0.056	0.037	0.076	0.040	0.059	0.038	0.090	0.052	0.149	0.045	0.207	0.039	0.284	0.045	0.356	0.039

Table 5: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{ba}$, $\hat{\alpha}_{bm}$, $(N = 50)$

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.164	0.194	0.168	0.195	0.173	0.200	0.430	0.486	0.443	0.488	0.458	0.496	0.597	0.736	0.671	0.772	0.724	0.792
std	0.077	0.081	0.053	0.054	0.036	0.037	0.093	0.102	0.057	0.062	0.036	0.039	0.145	0.183	0.075	0.090	0.039	0.046
se	0.083	0.086	0.054	0.056	0.035	0.037	0.100	0.108	0.059	0.064	0.035	0.039	0.144	0.193	0.072	0.094	0.035	0.045
rmse	0.085	0.081	0.062	0.054	0.044	0.037	0.117	0.103	0.081	0.063	0.055	0.039	0.250	0.194	0.149	0.094	0.085	0.047
size	0.059	0.037	0.084	0.042	0.116	0.048	0.090	0.036	0.140	0.050	0.225	0.049	0.269	0.052	0.407	0.052	0.546	0.050
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.160	0.190	0.167	0.195	0.172	0.198	0.430	0.485	0.446	0.491	0.457	0.495	0.584	0.718	0.671	0.769	0.723	0.791
std	0.080	0.084	0.053	0.056	0.034	0.035	0.095	0.102	0.058	0.063	0.036	0.038	0.150	0.182	0.079	0.094	0.037	0.045
se	0.083	0.086	0.054	0.056	0.035	0.037	0.100	0.108	0.059	0.064	0.035	0.039	0.144	0.190	0.072	0.094	0.036	0.046
rmse	0.090	0.084	0.063	0.056	0.044	0.035	0.118	0.103	0.079	0.063	0.056	0.038	0.264	0.200	0.151	0.099	0.086	0.046
size	0.081	0.059	0.097	0.051	0.113	0.042	0.091	0.035	0.118	0.043	0.226	0.047	0.318	0.061	0.420	0.052	0.576	0.045
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.162	0.193	0.167	0.195	0.171	0.197	0.432	0.486	0.445	0.492	0.460	0.498	0.579	0.712	0.668	0.769	0.724	0.791
std	0.075	0.078	0.051	0.053	0.034	0.036	0.097	0.103	0.059	0.064	0.034	0.037	0.144	0.182	0.076	0.096	0.039	0.048
se	0.083	0.086	0.054	0.056	0.035	0.037	0.100	0.108	0.059	0.064	0.035	0.039	0.145	0.193	0.073	0.095	0.035	0.046
rmse	0.084	0.078	0.061	0.053	0.045	0.036	0.119	0.104	0.081	0.064	0.053	0.037	0.264	0.202	0.152	0.101	0.085	0.049
size	0.055	0.034	0.072	0.036	0.126	0.035	0.091	0.039	0.138	0.048	0.196	0.046	0.318	0.049	0.439	0.053	0.562	0.068

Table 6: Means, Standard Deviations, Standard Errors, RMSE, and the Size of $\hat{\alpha}_{ba}$, $\hat{\alpha}_{bm}$, $(N = 100)$

	$\alpha = 0.2$						$\alpha = 0.5$						$\alpha = 0.8$					
	$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$		$T = 10$		$T = 15$		$T = 25$	
	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$	$\hat{\alpha}_{ba}$	$\hat{\alpha}_{bm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.181	0.197	0.182	0.197	0.185	0.199	0.458	0.489	0.469	0.493	0.476	0.497	0.661	0.759	0.717	0.786	0.752	0.794
std	0.056	0.058	0.037	0.039	0.025	0.026	0.069	0.072	0.042	0.044	0.025	0.026	0.112	0.129	0.058	0.066	0.028	0.032
se	0.060	0.061	0.039	0.040	0.025	0.026	0.074	0.077	0.044	0.046	0.026	0.027	0.115	0.138	0.057	0.068	0.028	0.033
rmse	0.059	0.058	0.041	0.039	0.029	0.026	0.081	0.073	0.053	0.044	0.034	0.026	0.178	0.136	0.101	0.067	0.056	0.032
size	0.047	0.042	0.058	0.049	0.076	0.048	0.075	0.045	0.098	0.049	0.134	0.043	0.203	0.048	0.281	0.046	0.385	0.044
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.178	0.195	0.183	0.197	0.186	0.200	0.456	0.487	0.469	0.494	0.477	0.498	0.659	0.755	0.718	0.783	0.752	0.796
std	0.058	0.060	0.038	0.039	0.026	0.026	0.070	0.074	0.043	0.044	0.026	0.027	0.112	0.130	0.060	0.068	0.028	0.031
se	0.060	0.061	0.039	0.040	0.025	0.026	0.074	0.077	0.044	0.046	0.026	0.027	0.115	0.139	0.057	0.068	0.028	0.033
rmse	0.062	0.060	0.042	0.039	0.029	0.026	0.083	0.075	0.053	0.044	0.035	0.028	0.180	0.138	0.102	0.070	0.055	0.031
size	0.066	0.049	0.059	0.044	0.083	0.051	0.083	0.042	0.093	0.050	0.137	0.047	0.225	0.041	0.268	0.057	0.390	0.045
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.181	0.197	0.184	0.199	0.185	0.199	0.463	0.493	0.474	0.499	0.477	0.499	0.664	0.761	0.719	0.785	0.753	0.796
std	0.057	0.058	0.037	0.038	0.024	0.024	0.068	0.071	0.042	0.045	0.027	0.028	0.116	0.135	0.059	0.066	0.028	0.031
se	0.060	0.061	0.039	0.040	0.025	0.026	0.073	0.077	0.044	0.046	0.026	0.027	0.117	0.141	0.057	0.068	0.028	0.033
rmse	0.060	0.058	0.040	0.038	0.028	0.024	0.077	0.071	0.050	0.045	0.035	0.028	0.179	0.140	0.101	0.067	0.054	0.031
size	0.043	0.034	0.056	0.037	0.076	0.042	0.059	0.039	0.090	0.046	0.149	0.051	0.207	0.043	0.284	0.039	0.356	0.039

Table 7: Alternative Instrumental Variables Matrix with New Instruments

	$\alpha = 0.2$				$\alpha = 0.5$				$\alpha = 0.8$			
	$T = 10$		$T = 15$		$T = 10$		$T = 15$		$T = 10$		$T = 15$	
	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$	$\hat{\alpha}^{bm}$	$\tilde{\alpha}^{bm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$												
mean	0.194	0.199	0.195	0.197	0.486	0.500	0.488	0.500	0.736	0.801	0.772	0.797
std	0.081	0.083	0.054	0.055	0.102	0.105	0.062	0.066	0.183	0.205	0.090	0.094
se	0.086	0.081	0.056	0.055	0.108	0.105	0.064	0.063	0.193	0.204	0.094	0.097
rmse	0.081	0.083	0.054	0.055	0.103	0.105	0.063	0.066	0.194	0.205	0.094	0.094
size	0.037	0.050	0.042	0.048	0.036	0.042	0.050	0.054	0.052	0.041	0.052	0.044
$\sigma_\eta^2/\sigma_v^2 = 1$												
mean	0.190	0.201	0.195	0.199	0.485	0.497	0.491	0.497	0.718	0.791	0.769	0.793
std	0.084	0.078	0.056	0.055	0.102	0.107	0.063	0.065	0.182	0.221	0.094	0.099
se	0.086	0.081	0.056	0.055	0.108	0.105	0.064	0.063	0.190	0.213	0.094	0.098
rmse	0.084	0.078	0.056	0.055	0.103	0.107	0.063	0.065	0.200	0.221	0.099	0.099
size	0.059	0.044	0.051	0.048	0.035	0.056	0.043	0.049	0.061	0.041	0.052	0.046
$\sigma_\eta^2/\sigma_v^2 = 10$												
mean	0.193	0.199	0.195	0.200	0.486	0.500	0.492	0.497	0.712	0.796	0.769	0.794
std	0.078	0.085	0.053	0.055	0.103	0.106	0.064	0.065	0.182	0.216	0.096	0.099
se	0.086	0.081	0.056	0.055	0.108	0.105	0.064	0.063	0.193	0.208	0.095	0.097
rmse	0.078	0.085	0.053	0.055	0.104	0.106	0.064	0.065	0.202	0.216	0.101	0.100
size	0.034	0.058	0.036	0.050	0.039	0.047	0.048	0.056	0.049	0.039	0.053	0.049

Table 8: Means, Standard Deviations, Standard Errors, RMSE, and the Size of the GMM Estimators with a Predetermined Variable ($N = 50$)

	$IV: z_{it}^{lla}$						$IV: z_{it}^{llm}$						$IV: z_{it}^{bbm}$					
	$T=10$		$T=15$		$T=25$		$T=10$		$T=15$		$T=25$		$T=10$		$T=15$		$T=25$	
	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.685	0.460	0.716	0.482	0.742	0.501	0.755	0.482	0.773	0.490	0.788	0.499	0.731	0.451	0.769	0.484	0.788	0.496
std	0.060	0.067	0.039	0.044	0.021	0.029	0.078	0.074	0.053	0.047	0.032	0.031	0.125	0.129	0.062	0.070	0.032	0.038
se	0.056	0.059	0.034	0.040	0.019	0.028	0.079	0.070	0.049	0.045	0.029	0.030	0.124	0.129	0.064	0.067	0.032	0.038
rmse	0.130	0.078	0.093	0.047	0.062	0.029	0.090	0.076	0.060	0.048	0.034	0.031	0.143	0.138	0.069	0.072	0.035	0.038
size	0.537	0.136	0.673	0.092	0.845	0.067	0.072	0.065	0.086	0.064	0.072	0.055	0.071	0.059	0.056	0.060	0.068	0.053
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.665	0.448	0.708	0.484	0.737	0.501	0.700	0.452	0.746	0.485	0.773	0.496	0.731	0.450	0.771	0.485	0.788	0.495
std	0.073	0.066	0.040	0.043	0.023	0.028	0.133	0.086	0.087	0.052	0.052	0.030	0.127	0.126	0.064	0.068	0.033	0.037
se	0.062	0.060	0.036	0.040	0.020	0.028	0.122	0.083	0.074	0.049	0.046	0.031	0.124	0.128	0.064	0.068	0.032	0.038
rmse	0.153	0.084	0.101	0.046	0.067	0.028	0.166	0.099	0.102	0.054	0.059	0.031	0.144	0.136	0.070	0.070	0.035	0.037
size	0.573	0.150	0.724	0.082	0.878	0.037	0.131	0.081	0.111	0.062	0.094	0.049	0.079	0.039	0.067	0.056	0.071	0.044
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.654	0.443	0.702	0.480	0.737	0.502	0.651	0.430	0.710	0.471	0.756	0.496	0.723	0.445	0.767	0.481	0.788	0.497
std	0.074	0.070	0.041	0.043	0.022	0.028	0.162	0.099	0.103	0.056	0.066	0.032	0.121	0.129	0.062	0.067	0.032	0.037
se	0.064	0.060	0.037	0.040	0.020	0.028	0.159	0.096	0.102	0.054	0.064	0.031	0.125	0.129	0.064	0.068	0.032	0.038
rmse	0.163	0.090	0.106	0.048	0.067	0.029	0.220	0.121	0.136	0.063	0.080	0.032	0.143	0.140	0.071	0.070	0.034	0.037
size	0.616	0.172	0.748	0.097	0.883	0.062	0.126	0.090	0.124	0.080	0.092	0.051	0.072	0.044	0.069	0.060	0.063	0.046

Table 9: Means, Standard Deviations, Standard Errors, RMSE, and the Size of the GMM Estimators with Predetermined Variable ($N = 100$)

	$IV: z_{it}^{lla}$						$IV: z_{it}^{llm}$						$IV: z_{it}^{bbm}$					
	$T=10$		$T=15$		$T=25$		$T=10$		$T=15$		$T=25$		$T=10$		$T=15$		$T=25$	
	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{la}$	$\hat{\beta}_{la}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{lm}$	$\hat{\beta}_{lm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$	$\hat{\alpha}_{bm}$	$\hat{\beta}_{bm}$
$\sigma_\eta^2/\sigma_v^2 = 0.2$																		
mean	0.732	0.471	0.749	0.486	0.765	0.498	0.781	0.491	0.788	0.495	0.795	0.499	0.759	0.467	0.782	0.489	0.796	0.498
std	0.047	0.048	0.029	0.032	0.016	0.021	0.058	0.050	0.036	0.033	0.021	0.022	0.092	0.096	0.049	0.050	0.022	0.026
se	0.045	0.044	0.027	0.030	0.015	0.020	0.057	0.050	0.035	0.032	0.021	0.021	0.097	0.098	0.048	0.049	0.023	0.027
rmse	0.082	0.056	0.059	0.035	0.039	0.021	0.061	0.051	0.038	0.034	0.021	0.022	0.100	0.101	0.052	0.051	0.023	0.026
size	0.312	0.108	0.467	0.089	0.644	0.056	0.067	0.054	0.062	0.062	0.048	0.048	0.057	0.050	0.073	0.043	0.042	0.041
$\sigma_\eta^2/\sigma_v^2 = 1$																		
mean	0.709	0.463	0.740	0.485	0.759	0.498	0.743	0.474	0.769	0.490	0.784	0.498	0.754	0.465	0.785	0.490	0.794	0.498
std	0.053	0.049	0.032	0.032	0.016	0.020	0.101	0.066	0.060	0.038	0.036	0.021	0.091	0.094	0.046	0.047	0.022	0.026
se	0.050	0.045	0.029	0.030	0.016	0.020	0.094	0.061	0.057	0.035	0.034	0.022	0.096	0.098	0.047	0.049	0.023	0.027
rmse	0.105	0.061	0.068	0.036	0.044	0.020	0.116	0.071	0.068	0.039	0.039	0.021	0.102	0.100	0.049	0.048	0.023	0.026
size	0.434	0.141	0.520	0.098	0.734	0.045	0.094	0.074	0.079	0.076	0.069	0.051	0.057	0.037	0.052	0.040	0.044	0.038
$\sigma_\eta^2/\sigma_v^2 = 10$																		
mean	0.698	0.458	0.736	0.482	0.757	0.498	0.677	0.443	0.731	0.476	0.766	0.496	0.758	0.470	0.784	0.487	0.793	0.497
std	0.056	0.050	0.034	0.032	0.018	0.020	0.144	0.082	0.097	0.046	0.059	0.022	0.087	0.090	0.046	0.048	0.023	0.026
se	0.053	0.046	0.030	0.030	0.016	0.020	0.142	0.079	0.090	0.041	0.055	0.022	0.095	0.097	0.048	0.050	0.023	0.027
rmse	0.116	0.065	0.072	0.037	0.046	0.021	0.189	0.100	0.119	0.052	0.068	0.023	0.096	0.095	0.048	0.050	0.024	0.026
size	0.459	0.158	0.546	0.112	0.723	0.046	0.120	0.081	0.107	0.095	0.104	0.051	0.046	0.037	0.040	0.050	0.054	0.046

Table 10: Estimation Results of the Employment Equation

	$\hat{\alpha}_{la}$	$\hat{\alpha}_{bm}$	$\tilde{\alpha}_{bm}$
Estimates	0.541	0.800	0.827
Std.	0.020	0.051	0.070

	$\hat{\beta}_{la}$	$\hat{\beta}_{bm}$	$\tilde{\beta}_{bm}$
Estimates	-0.477	-0.829	-0.924
Std.	0.032	0.077	0.100

Figures 1 to 3: Asymptotic Variances with T fixed ($N = 50$, $\alpha = 0.8$)

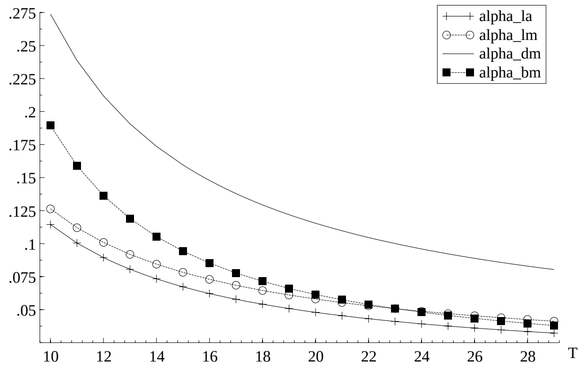

Figure 1: $\sigma_\eta^2/\sigma_v^2 = 0.2$

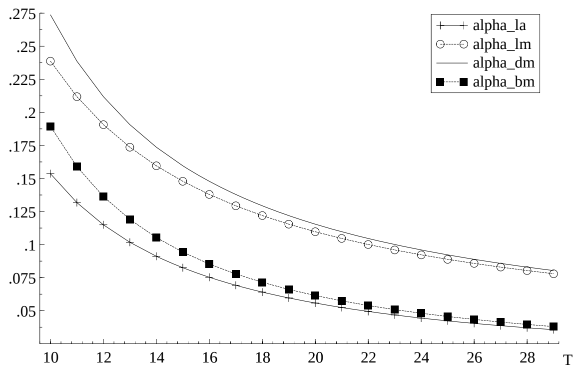

Figure 2: $\sigma_\eta^2/\sigma_v^2 = 1$

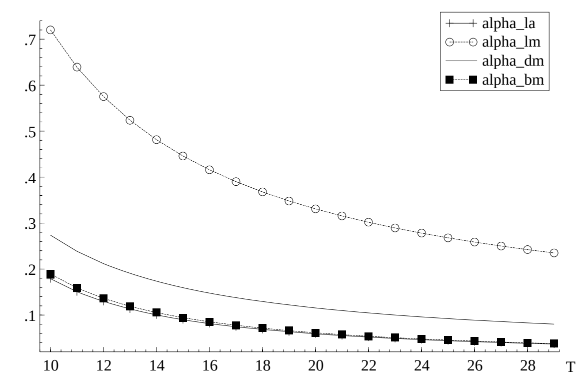

Figure 3: $\sigma_\eta^2/\sigma_v^2 = 10$

Figures 4 to 6: Power Plot, $H_0 : \alpha = 0.8$ ($T = 10$, $N = 50$)

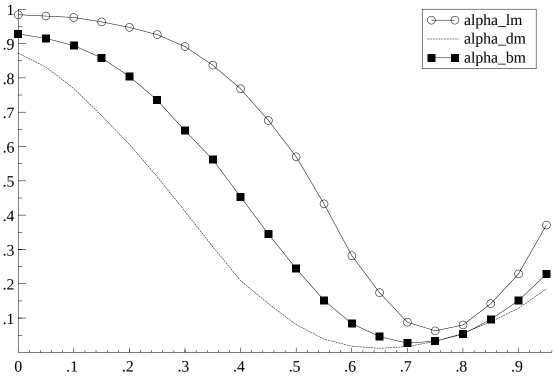

Figure 4: $\sigma_\eta^2/\sigma_v^2 = 0.2$

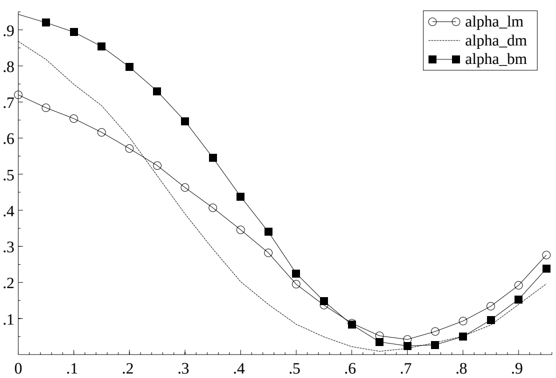

Figure 5: $\sigma_\eta^2/\sigma_v^2 = 1$

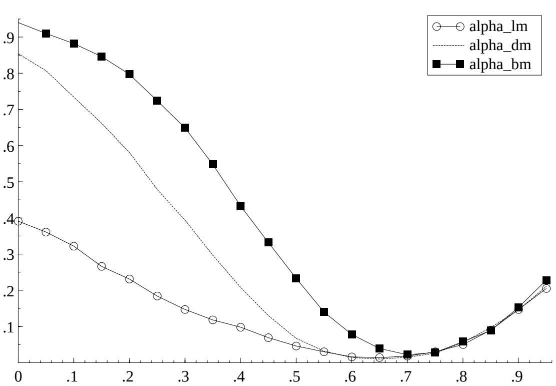

Figure 6: $\sigma_\eta^2/\sigma_v^2 = 10$

Figures 7 to 9: Power Plot, $H_0 : \alpha = 0.8$

$(T = 15,\, N = 50)$

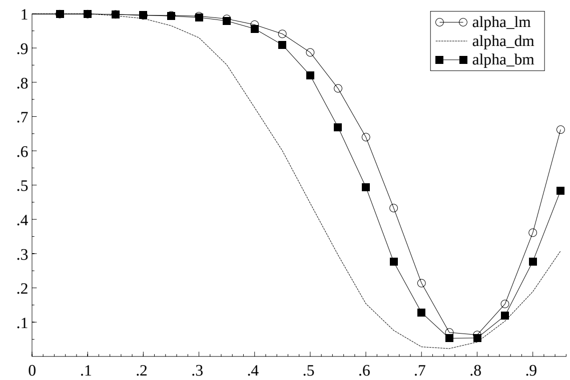

Figure 7: $\sigma_\eta^2/\sigma_v^2 = 0.2$

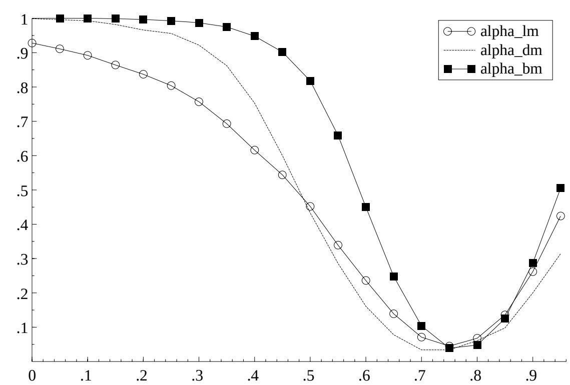

Figure 8: $\sigma_\eta^2/\sigma_v^2 = 1$

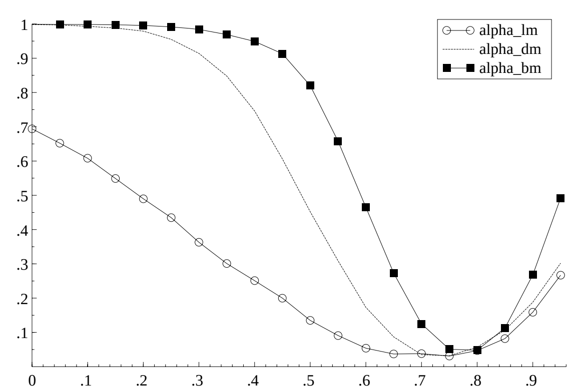

Figure 9: $\sigma_\eta^2/\sigma_v^2 = 10$

Figures 10 -to 12: Power Plot, $H_0 : \alpha = 0.8$

$(T = 25,\, N = 50)$

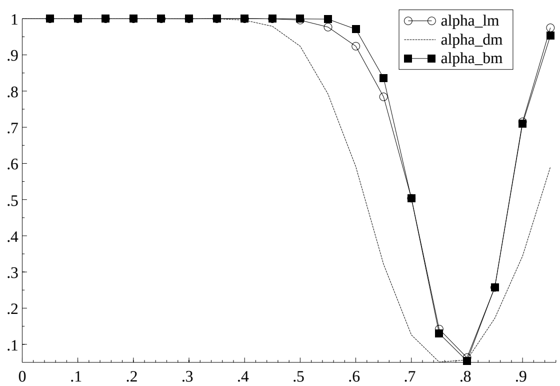

Figure 10: $\sigma_\eta^2/\sigma_v^2 = 0.2$

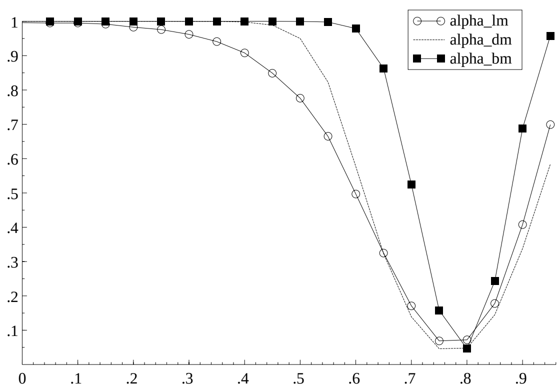

Figure 11: $\sigma_\eta^2/\sigma_v^2 = 1$

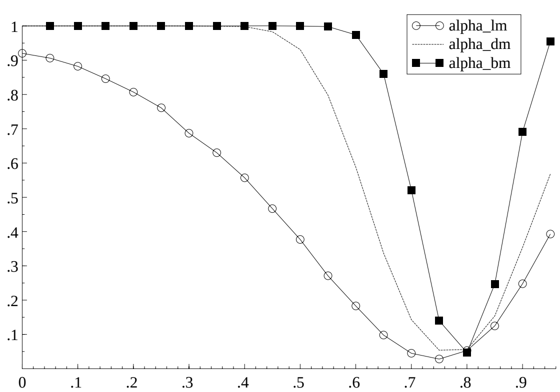

Figure 12: $\sigma_\eta^2/\sigma_v^2 = 10$