A Non-Invariance Problem in Panel GMM Estimators When Level Instruments Are Used for Differenced Equations

Seung C. Ahn[a]
Sogang University

Yoshitsugu Kitazawa[b]
Kyushu Sangyo University

First Draft: August, 2010
This Version: February, 2014

Abstract

The most popular way to handle unobservable individual-specific effects in panel data models is to remove the effects by first-differencing or quasi-differencing regression equations. Then, the differenced equations are estimated by the Generalized Method of Moments (GMM). We show that when a constant is not used as an instrument, the GMM estimation results from this method are not invariant to linear transformations of regressors. In contrast, when a constant is used as an instrument, GMM estimators are asymptotically invariant to linear transformations of regressors. The estimators may not be invariant in finite samples. An example is the GMM estimator for count panel data models with endogenous regressors. For the models, two-step GMM estimators are not invariant in finite samples. Nonetheless, our simulation results indicate that the two-step GMM estimator is less sensitive to linear transformation of regressors when it is computed with a constant instrument than when it is not.

***Key Words*: Dynamic Panel Models, Count Panel Models, GMM, Instrumental Variables.**

***JEL Classification*: C23.**

----

[a] Corresponding author, School of Economics, Sogang University, 35 Baekbeom-ro Mapo-gu, Seoul 121 – 742, South Korea. Email: smahn@sogang.ac.kr.

[b] Department of Economics, Kyushu Sangyo University, 3-1 Matsukadai 2-chome, Higashi-ku, Fukuoka, 813-8503, Japan. Email: kitazawa@ip.kyusan-u.ac.jp.

1. Introduction

Panel data regression models have been popularly used for empirical studies. Using panel data, researchers can consistently estimate regression coefficients controlling for unobservable time-invariant individual-specific effects that may be correlated with regressors. The most popular way to handle the individual effects is to remove them by first-differencing or quasi-differencing regression equations. Then, the differenced equations are estimated by the Generalized Method of Moments (GMM) of Hansen (1982); for example, see, amongst many, Anderson and Hsiao (1981), Arellano and Bond (1991), and Ahn and Schmidt (1995), for linear dynamic models, and Chamberlain (1992), Wooldridge (1997), and Windmeijer (2000), for count data models. The typical instruments used for the differenced equations are lagged level regressors. We refer to this approach as "level-instruments-for-differenced-equations" (LIDE) approach.

The purpose of this paper is to remind the LIDE users of the importance of including a constant (typically, one) in instrument sets. Consequently, we consider two types of GMM estimators: the first is the estimator using a constant as an instrument in addition to lagged level regressors, and the second, the estimator without a constant in the instrument set. We refer to these two estimators respectively as "GMM-C" and "GMM-WC" estimators. For dynamic panel data models, Crépon, Kramarz and Trognon (1997) have shown that the GMM-C estimator is asymptotically more efficient than the GMM-WC estimator. In this paper, we address a more serious problem related to the GMM-WC estimator. That is, we show that the GMM-WC estimation results are not invariant to the means of regressors. That is, when some constant numbers are added to regressors, the estimation results can change.

The asymptotic variance of the GMM-WC estimator could increase or decrease as the mean of a regressor increases. Use of linearly transformed regressors can lead to different statistic inferences. For example, suppose a researcher wishes to estimate dynamics of earnings of individual workers using panel data. The dependent variable is logarithm of earnings and a regressor is its one-period lagged value. The mean of the logarithmic earnings depends on what unit is used to measure earnings. The mean is greater if earnings are measured in dollars instead of thousand dollars. For this case, the asymptotic variance matrix of the GMM-WC estimator depends on what units of earnings are used. In contrast, the asymptotic variance matrix of the GMM-C estimator is invariant to the units of earnings used. This problem is not limited to estimation of dynamic panel models.

Differenced equations contain intercepts if the model to be estimated contains time-specific effects common to all individuals. To estimate the models with time effects, researchers naturally use the GMM-C estimator to estimate the time effects. On the other hand, differenced equations of the models without time effects do not contain intercepts. The GMM-WC estimator is often used to estimate such models. In particular, theoretical papers often consider the models without time effects for analytical convenience. Simulations comparing finite-sample properties of GMM estimators are also done without using a constant as an instrument. This tradition might have given empirical researchers a perception that use of GMM-C estimators may be desirable for the estimation of the models without time effects. It is well known that GMM estimators using too many moment conditions (instrumental variables) often have poor finite sample properties.[1] A large number of moment conditions are available for panel data models. It is often even infeasible to implement all of the available moment conditions in GMM. Accordingly, researchers are often forced to use only a subset of moment conditions. For such cases, researchers may prefer the GMM-C estimator to the GMM-WC estimator because the latter uses a smaller number of moment conditions.

The main message of this paper is that the GMM-C estimator should be used for panel models not only because it is more efficient than the GMM-WC estimator, but also because it is often invariant to linear transformation of regressors. Although the GMM-C estimator is asymptotically invariant, it is not always invariant in finite samples. For example, count panel data models can be estimated by the quasi-differencing method of Wooldridge (1997) when regressors are endogenous. While the two-step GMM-C estimator using the Wooldridge method is invariant to linear transformation of regressors asymptotically, it is not invariant in finite sample. However, our simulation results indicate that even for such cases, the GMM-C estimator is less sensitive to linear transformations than the GMM-WC estimator is.

This paper is organized as follows. Section 2 considers three different LIDE methods. One is the Arellano-Bond (1991) method for dynamic panel data models. The other two methods are the quasi-differencing methods of Chamberlain (1992) and Wooldridge (1997, endnote 2) for count panel data models. We show that both the asymptotic and finite-sample distributions of the GMM-WC estimators from these methods are not invariant to the means of regressors. We also consider the GMM-C estimators from the three methods. Section 3 reports

[1] See Altonji and Segal (1996) and Burnside and Eichenbaum (1996), amongst many others.

the results from our small-scale simulation exercises. Some concluding remarks follow in section 4.

2. Asymptotic Distributions of GMM Estimators

This section investigates the asymptotic distributions of GMM-WC and GMM-C estimators for simple dynamic and count data models. We consider the models with a large number of cross-section observations (*N*) and a small number of time series observations (*T*). Accordingly, asymptotics apply as $N \to \infty$ with fixed T. We assume that data are cross-sectionally independently and identically distributed (*i.i.d.*) with finite fourth moments. Under this assumption, usual GMM asymptotic theory apples. To save space, we only consider the cases with $T = 2$ and a single regressor. Our results can be easily generalized to the cases with more regressors and larger *T*.

Throughout this paper we use the following notation. First, subscripts "*i*" and "*t*" index individuals and time, respectively ($i = 1, \ldots, N$, $t = 1, \ldots, T$). Second, β denotes a regression coefficient. Third, we use notation $m_{wc,i}(\beta)$ and $m_i(\beta) = (m_{wc,i}(\beta)', m_{c,i}(\beta)')'$ to denote the moment functions used for the GMM-WC and GMM-C estimators, respectively. The moment function $m_{wc,i}(\beta)$ has the form of a lagged level regressor (level instrument) times a differenced (or quasi-differenced) error term, while $m_{c,i}(\beta)$ is simply the differenced error (*i.e.*, a constant times differenced error). Fourth, we use notation β_o to denote the true value of β. Fifth, we use

$$M_{wc} = E\left(\frac{\partial m_{wc,i}(\beta_o)}{\partial \beta'}\right); M_c = E\left(\frac{\partial m_{c,i}(\beta_o)}{\partial \beta'}\right); M = \begin{pmatrix} M_{wc} \\ M_c \end{pmatrix};$$

$$V_{j,j'} = E[m_{j,i}(\beta_o) m_{j',i}(\beta_o)']; \; V = \left[V_{j.j'}\right], \text{ where } j, j' = wc, c.$$

Sixth, for any variable x_{it}, $\Delta x_{it} = x_{it} - x_{i,t-1}$.

Seventh, $\hat{\beta}_{wc}$ and $\hat{\beta}_c$ denote the GMM-WC and GMM-C estimators, respectively. These estimators are obtained by minimizing $\bar{m}_{wc}(\beta)'(\hat{V}_{wc,wc})^{-1}\bar{m}_{wc}(\beta)$ and $\bar{m}(\beta)'\hat{V}^{-1}\bar{m}(\beta)$, respectively, where $\bar{m}_{wc}(\beta) = N^{-1}\Sigma_{i=1}^{N} m_{wc,i}(\beta)$, $\bar{m}(\beta) = N^{-1}\Sigma_{i=1}^{N} m_i(\beta)$, and $\hat{V}_{wc,wc}$ and $\hat{V}$ are any consistent estimators of $V_{wc,wc}$ and V. Eighth, $\Xi_{wc} = \lim_{N\to\infty} \mathrm{var}\left[\sqrt{N}(\hat{\beta}_{wc} - \beta_o)\right]$ is the asymptotic variance of $\sqrt{N}(\hat{\beta}_{wc}-\beta_o)$, and Ξ_c is similarly defined. Since data are *i.i.d.* over different *i*, $\Xi_{wc}=[M'_{wc}V_{wc}^{-1}M_{wc}]^{-1}$ and $\Xi_c=[M'V^{-1}M]^{-1}$.

Finally, for any variable x_{it}, $x_{it}^b = x_{it}-b$, where b is a constant. The constant b could be replaced by a random variable b_N (which depends on *N*, such as $\bar{x}=(NT)^{-1}\Sigma_{i=1}^N\Sigma_{t=1}^T x_{it}$), so long as $p\lim_{N\to\infty}|b_N|=|b|<\infty$. We will use superscript "*b*" for the terms related to the GMM estimators using x_{it}^b instead of x_{it}; *e.g.*, M^b for M and V^b for V.

2.1. Simple Dynamic Panel Model

Consider the following simple dynamic panel model:

$$y_{it}=\beta_o y_{i,t-1}+\alpha_i+\varepsilon_{it}, \tag{1}$$

where $-1<\beta_0<1$, y_{it} is the dependent variable, α_i is the unobservable time-invariant individual effect, and ε_{it} is a usual regression error term. We assume that the initial values y_{i0} are observed. Following Ahn and Schmidt (1995), we assume:

$$\begin{pmatrix} y_{i0} \\ \alpha_i \\ \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_0 \\ \mu_\alpha \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{\alpha 0} & 0 & 0 \\ \sigma_{\alpha 0} & \sigma_\alpha^2 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}\right). \tag{2}$$

As is well known in the literature, the ordinary least squares (OLS) or the within estimators of β are inconsistent (see Nickell (1981)). A consistent estimator can be obtained by the GMM methods of Anderson and Hsiao (1981) or Arellano and Bond (1991).[2] To use their methods, we first-difference the equation (1) and then estimate β_o by GMM using the instruments $(y_{i0},1)'$.

Let $m_{wc,i}(\beta)=y_{i0}(\Delta y_{i2}-\beta\Delta y_{i1})$ and $m_{c,i}(\beta)=\Delta y_{i2}-\beta\Delta y_{i1}$.[3] Then, under (2), we can

[2] When $T>2$ or some restrictions are imposed on $\alpha_{\alpha 0}$, some additional moment conditions are available that are not of the form of level lagged regressors times differenced errors. See Arellano and Bover (1995), Ahn and Schmidt (1995), and Blundell and Bond (1999), for more details. We do not consider these additional moment conditions here. However, our result here can be easily extended to their estimators.

[3] The original moment conditions proposed by Crépon, Kramarz and Trognon (1997) are not of the form of $m_{c,i}(\beta)$.

easily show that

$$E\left(m_i(\beta_o)\right) = E\begin{pmatrix} m_{wc,i}(\beta_o) \\ m_{c,i}(\beta_o) \end{pmatrix} = E\begin{pmatrix} y_{i0}(\Delta y_{i2} - \beta_o \Delta y_{i1}) \\ \Delta y_{i2} - \beta_o \Delta y_{i1} \end{pmatrix} = E\begin{pmatrix} y_{i0}\Delta\varepsilon_{i2} \\ \Delta\varepsilon_{i2} \end{pmatrix} = 0.$$[4]

Thus, the GMM-WC ($\hat{\beta}_{wc}$) and GMM-C ($\hat{\beta}_c$) estimators respectively using the moment functions $m_{wc,i}(\beta)$ and $m_i(\beta)$ are consistent.

Since $\Delta y_{i1} = (\beta_o - 1)y_{i0} + \alpha_i + \varepsilon_{i1}$, we can easily show that under (2),

$$M = -\begin{pmatrix} E(y_{io}\Delta y_{i1}) \\ E(\Delta y_{i1}) \end{pmatrix} = -\begin{pmatrix} (\beta_o - 1)\sigma_0^2 + \sigma_{\alpha 0} + \mu_0\{(\beta_o - 1)\mu_0 + \mu_\alpha\} \\ (\beta_o - 1)\mu_0 + \mu_\alpha \end{pmatrix};$$

$$V = E\begin{pmatrix} y_{i0}^2(\Delta\varepsilon_{i2})^2 & y_{i0}(\Delta\varepsilon_{i2})^2 \\ y_{i0}(\Delta\varepsilon_{i2})^2 & (\Delta\varepsilon_{i2})^2 \end{pmatrix} = 2\sigma_\varepsilon^2 \begin{pmatrix} \mu_0^2 + \sigma_0^2 & \mu_0 \\ \mu_0 & 1 \end{pmatrix}.$$

where $(\beta_o - 1)\mu_o + \mu_\alpha = E(\Delta y_{i1})$. With these results, a straightforward algebra shows

$$\Xi_{wc} = \left[M'_{wc}(V_{wc,wc})^{-1}M_{wc}\right]^{-1} = \frac{2\sigma_\varepsilon^2(\mu_0^2 + \sigma_0^2)}{\left[(\beta_o - 1)\sigma_0^2 + \sigma_{\alpha 0} + \mu_0\{(\beta_o - 1)\mu_0 + \mu_\alpha\}\right]^2}; \quad (3)$$

$$\Xi_c = \left[M'V^{-1}M\right]^{-1} = \frac{2\sigma_\varepsilon^2\sigma_0^2}{[(\beta_o - 1)\sigma_0^2 + \sigma_{\alpha 0}]^2 + \sigma_0^2[(\beta_o - 1)\mu_0 + \mu_\alpha]^2}. \quad (4)$$

It can be also shown

$$\Xi_{wc} - \Xi_c = \frac{2\sigma_\varepsilon^2\left(\mu_0\sigma_{\alpha 0} - \sigma_0^2\mu_\alpha\right)^2}{\left[(\beta_o-1)\sigma_0^2+\sigma_{\alpha 0}+\mu_0\{(\beta_o-1)\mu_0+\mu_\alpha\}\right]^2\left[\{(\beta_o-1)\sigma_0^2+\sigma_{\alpha 0}\}^2+\sigma_0^2\{(\beta_o-1)\mu_0+\mu_\alpha\}^2\right]} \geq 0,$$

where the equality holds only if $\mu_0 = \mu_\alpha = 0$. That is, the GMM-C estimator is strictly more efficient that the GMM-WC estimator unless $\mu_0 = \mu_\alpha = 0$, as Crépon, Kramarz and Trognon (1997) projected.

We now investigate the asymptotic variances of the GMM-WC and GMM-C estimators

---

Their moment conditions are $m_{ckt,i}(\beta) = [(y_{i1} - \beta y_{i0}), (y_{i2} - \beta y_{i1})]'$. However, these moment functions are valid to use only if $\mu_\alpha = 0$. When $\mu_\alpha \neq 0$, the moment function $m_{wc,i}(\beta)$ should replace $m_{ckt,i}(\beta)$.

[4] When the data y_{it} are mean-stationary (i.e., $E(y_{it})$ is the same for all t), the moment function $m_{c,i}(\beta)$ alone cannot identify β_o because $E(m_{c,i}(\beta)) = 0$ for all β. However, even for this case, the GMM estimator using $m_{c,i}(\beta)$ in addition to $m_{we,i}(\beta)$ as moment functions is asymptotically more efficient than the GMM estimation using $m_{wc,i}(\beta)$ alone.

using y_{it}^b (= $y_{it} - b$) instead of y_{it} . We could consider a more general case in which $y_{it}^* = ay_{it} - b$ is used for y_{it} . However, the constant a does not influence the statistical inferences from the GMM-WC and GMM-C estimators.

Rewriting equation (1) with y_{it} replaced by y_{it}^b, we have

$$y_{it}^b = \beta_o y_{i,t-1}^b + \alpha_i^b + \varepsilon_{it}, \tag{5}$$

where $\alpha_i^b = \alpha_i + (\beta_o - 1)b$. We can still consistently estimate β by the LIDE approach using y_{i0}^b or $(1, y_{i0}^b)'$ as instruments. Observe that $\mathrm{var}(y_{it}^b)$ = σ_0^2 , $\mathrm{cov}(y_{i0}^b, \alpha_i^b) = \sigma_{\alpha 0}$, and $\mathrm{var}(\alpha_i^b) = \sigma_\alpha^2$. Accordingly, the two equations (1) and (5) are the same except that $E(y_{i0}^b) = \mu_0 - b$ and $E(\alpha_i^b) = \mu_\alpha + (\beta_o - 1)b$ in (5). Therefore, the asymptotic variances of the GMM-WC and GMM-C estimators using the y_{it}^b instead of the y_{it} can be easily obtained by replacing μ_0 and μ_α in (3) and (4) by $E(y_{i0}^b) = \mu_0 - b$ and $E(\alpha_i^b) = \mu_\alpha + (\beta_o - 1)b$. That is,

$$\Xi_{wc}^b = \frac{2\sigma_\varepsilon^2[(\mu_0 - b)^2 + \sigma_0^2]}{\left[(\beta_o - 1)\{(\mu_0 - b)^2 + \sigma_0^2\} + \sigma_{\alpha 0} + (\mu_0 - b)\{\mu_\alpha + b(\beta_o - 1)\}\right]^2} \neq \Xi_{wc};$$

$$\Xi_c^b = \frac{2\sigma_\varepsilon^2 \sigma_0^2}{[(\beta_o - 1)\sigma_0^2 + \sigma_{\alpha 0}]^2 + \sigma_0^2[(\beta_o - 1)\mu_0 + \mu_\alpha]^2} = \Xi_c,$$

if $b \neq 0$. Observe that Ξ_{wc}^b depends on b. In contrast, Ξ_c^b is invariant to the choice of b.

We are unable to track down the general relationship between Ξ_{wc}^b and b . The asymptotic variance could increase or decrease with b depending on the values of the other parameters in Ξ_{wc}^b. However, a clear relationship emerges if we impose some restrictions on the parameters in (2). For example, suppose that the dynamics of the y_{it} had begun from a far past although they are observed from time zero. Under this assumption, Ahn and Schmidt (1997) have shown that

$$y_{i0} = \frac{\alpha_i}{1 - \beta_o} + v_{i0} , \tag{6}$$

where $E(v_{i0}) = 0$, $\mathrm{var}(v_{i0}) = \sigma_v^2$, and v_{i0} is uncorrelated with ε_{i1}, ε_{i2}, and α_i.

Many previous studies have used (1), (2) and (6) for their Monte Carlo experiments (see, for example, Arellano and Bond (1991), Blundell and Bond (1999), and Hahn (1999)). The restriction (6) implies that

$$\mu_0 = \frac{\mu_\alpha}{1-\beta_o}; \sigma_0^2 = \frac{\sigma_\alpha^2}{(1-\beta_o)^2} + \sigma_v^2; \sigma_{\alpha 0} = \frac{\sigma_\alpha^2}{1-\beta_o}. \tag{7}$$

Substituting these restrictions into Ξ_{wc} and with some algebra, we have

$$\Xi_{wc} = \frac{2\sigma_\varepsilon^2}{(1-\beta_o)^4} \frac{(\mu_\alpha^2 + \sigma_\alpha^2) + (1-\beta_o)^2 \sigma_v^2}{\sigma_v^4}; \tag{8}$$

$$\Xi_c = \frac{2\sigma_\varepsilon^2}{(1-\beta_o)^4} \frac{\sigma_\alpha^2 + (1-\beta_o)^2 \sigma_v^2}{\sigma_v^4}. \tag{9}$$

Observe that given σ_ε^2, σ_α^2, σ_v^2 and β_o, Ξ_{wc} monotonically increases with $|\mu_\alpha|$. In fact, the variance Ξ_{wc} is minimized at Ξ_c if $\mu_\alpha = 0$ (and therefore, $E(y_{i0}) = 0$). Holding other things equal, the efficiency of the GMM-WC estimator has a negative relationship with the absolute value of the mean of y_{it}, because $E(y_{it}) = \mu_\alpha / (1-\beta_o)$ for all *t* if (6) holds.

This result also implies that the GMM-WC estimator is not invariant to *b*. Observe that $y_{i0}^b = \alpha_i^b / (1-\beta_o) + v_{i0}$. Thus, Ξ_{wc}^b can be obtained by replacing μ_α in (8) with $E(\alpha_i^b)$ = $[\mu_\alpha + (\beta_o - 1)b]$. That is,

$$\Xi_{wc}^b = \frac{2\sigma_\varepsilon^2}{(1-\beta_o)^4} \frac{[\{\mu_\alpha + (\beta_o - 1)b\}^2 + \sigma_\alpha^2] + (1-\beta_o)^2 \sigma_v^2}{\sigma_v^4}. \tag{10}$$

The non-invariance of the GMM-WC estimator can be analyzed in a more systematic way that can be also used in the later subsections. Observe that

$$\begin{aligned} m_i^b(\beta) &= \begin{pmatrix} y_{i0}^b(\Delta y_{i2}^b - \beta_o \Delta y_{i1}^b) \\ \Delta y_{i2}^b - \beta_o \Delta y_{i1}^b \end{pmatrix} = \begin{pmatrix} 1 & -b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{i0}(\Delta y_{i2} - \beta_o \Delta y_{i1}) \\ \Delta y_{i2} - \beta_o \Delta y_{i1} \end{pmatrix} \\ &\equiv \begin{pmatrix} c_1(b)' \\ c_2' \end{pmatrix} m_i(\beta) \equiv C(b)' m_i(\beta) \end{aligned}. \tag{11}$$

Thus, we have $M^b = C(b)'M$ and $V^b = C(b)'VC(b)$. Therefore,

$$\Xi_{wc}^b = [M'c_1(b)(c_1(b)'Vc_1(b))^{-1}c_1(b)'M]^{-1} = \frac{c_1(b)'Vc_1(b)}{(c_1(b)'M)^2}; \tag{12}$$

$$\Xi_c^b = \left[M'C(b)(C(b)'VC(b))^{-1}C(b)'M\right]^{-1} = [M'V^{-1}M]^{-1} = \Xi_c. \tag{13}$$

From (12), we can clearly see that why Ξ_{wc}^b depends on *b* through $c_1(b)' = (1, -b)$. Because

$c_1(b)$ is a vector, not an invertible matrix, it does not cancel out from (12). In contract, $C(b)$ is an invertible square matrix, and therefore, it cancel out from Ξ_c^b.

In fact the GMM-C estimator is invariant to *b* even in finite samples. This is because the matrix $C(b)$ in (11) does not depend on data or the parameter β. Let $Z_i^b = (y_{i0}^b, 1)'$ and $r_i^b(\beta) = \Delta y_{i2}^b - \beta \Delta y_{1i}^b$ for any scalar *b* including zero. Using this notation, the moment function of the GMM-C estimator is given by $m_i^b(\beta) = Z_i^b r_i^b(\beta)$. Then, following Arellano and Bond (1991), we can obtain a one-step GMM-C estimator $\tilde{\beta}_c^b$ by minimizing

$$Q_{1,N}^b(\beta) \equiv \left(\Sigma_{i=1}^N Z_i^b r_i^b(\beta)\right)' \left[\Sigma_{i=1}^N Z_i^b H Z_i^{b\prime}\right]^{-1} \left(\Sigma_{i=1}^N Z_i^b r_i^b(\beta)\right), \tag{14}$$

where $H = 2$.[5] However, $Q_{1,N}^b(\beta)$ is the same for all *b*, because

$$\begin{aligned} Q_{1,N}^b(\beta) &= \left(\Sigma_{i=1}^N C(b)' Z_i^0 r_i^0(\beta)\right)' \left[\Sigma_{i=1}^N C(b)' Z_i^0 H Z_i^{0\prime} C(b)\right]^{-1} \left(\Sigma_{i=1}^N C(b)' Z_i^0 r_i^0(\beta)\right) \\ &= \left(\Sigma_{i=1}^N Z_i^0 r_i^0(\beta)\right)' \left[\Sigma_{i=1}^N Z_i^0 H Z_i^0\right]^{-1} \left(\Sigma_{i=1}^N Z_i^0 r_i^0(\beta)\right) \\ &\equiv Q_{1,N}^0(\beta) \end{aligned}$$

Similarly, the minimand of the two-step GMM-C estimator using the y_{it}^b, say $Q_{2,N}^b(\beta)$, is the same for all *b*, because

$$\begin{aligned} Q_{2,N}^b(\beta) &\equiv \left(\Sigma_{i=1}^N Z_i^b r_i^b(\beta)\right)' \left[\Sigma_{i=1}^N Z_i^b r_i^b(\tilde{\beta}_c^b) r_i^b(\tilde{\beta}_c^b)' Z_i^{b\prime}\right]^{-1} \left(\Sigma_{i=1}^N Z_i^b r_i^b(\beta)\right) \\ &\equiv \left(\Sigma_{i=1}^N C(b)' Z_i^0 r_i^0(\beta)\right)' \left[\Sigma_{i=1}^N C(b)' Z_i^0 r_i^0(\tilde{\beta}_c^0) r_i^0(\tilde{\beta}_c^0)' Z_i^{0\prime} C(b)\right]^{-1} \left(\Sigma_{i=1}^N C(b)' Z_i^0 r_i^0(\beta)\right) \\ &= \left(\Sigma_{i=1}^N Z_i^0 r_i^0(\beta)\right)' \left[\Sigma_{i=1}^N Z_i^0 r_i^0(\tilde{\beta}_c^0) r_i^0(\tilde{\beta}_c^0)' Z_i^{0\prime}\right]^{-1} \left(\Sigma_{i=1}^N Z_i^0 r_i^0(\beta)\right) \\ &\equiv Q_{2,N}^0(\beta). \end{aligned} \tag{15}$$

Thus, both the one-step and two-step GMM-C estimators are invariant to *b* in finite samples.

2.2. Simple Count Data Models

We now consider a simple multiplicative count data model where the dependent variable takes nonnegative integer numbers only. Following Chamberlain (1992), Wooldridge (1997, endnote

[5] The general form of H for the cases with $T > 2$ can be found from Arellano and Bond (1991) or Section 3 of this paper.

2) and Windmeijer (2000), amongst many others, we consider the model:

$$y_{it} = \exp(x_{it}\beta_o + \alpha_i + \varepsilon_{it}) = \exp(x_{it}\beta_o)\eta_i v_{it}, \quad (16)$$

where $\eta_i = \exp(\alpha_i)$ and $v_{it} = \exp(\varepsilon_{it})$. We assume that the regressor x_{it} is only weakly exogenous to the error v_{it} conditional on η_i. That is,

$$E(v_{i1} \mid \eta_i, x_{i1}) = 1;\ E(v_{i2} \mid \eta_i, x_{i1}, x_{i2}) = 1. \quad (17)$$

There are two differencing methods that can be used to control for the individual effects. The first is the one by Chamberlain (1992). His method utilizes the following quasi-differenced error function:

$$p_i^0(\beta) = y_{i1} - \exp(-\Delta x_{i2}\beta) y_{i2}.$$

Under (16), $p_i^0(\beta_o) = \exp(x_{i1}\beta_o)\eta_i(v_{i1} - v_{i2})$. Thus, we can easily show that $E[m_i(\beta_o)] = 0$, where $S_i^0 = x_{i1}$, $Z_i^0 = (S_i^0, 1)'$, and

$$m_i(\beta) = \begin{pmatrix} m_{wc,i}(\beta) \\ m_{c,i}(\beta) \end{pmatrix} = Z_i^0 p_i^0(\beta). \quad (18)$$

Notice that $\partial p_i^0(\beta)/\partial\beta = \exp(-\Delta x_{i2}\beta)\Delta x_{i2} y_{i2}$. Thus, the asymptotic variance of the GMM-C estimator using the moment functions (18) equals $\Xi_c = [M'V^{-1}M]^{-1}$, where

$$M = E\left(Z_i^0 \exp(-\Delta x_{i2}\beta_o) y_{i2}\Delta x_{i2}\right) = E\left(Z_i^0 \exp(x_{i1}\beta_o)\eta_i \Delta x_{i2}\right); \quad (19)$$

$$V = E[\exp(2x_{i1}\beta_o)\eta_i^2 (v_{i1} - v_{i2})^2 Z_i^0 Z_i^{0\prime}]. \quad (20)$$

We now consider the GMM-WC and GMM-C estimators using x_{it}^b instead of x_{it} (t = 1, 2). Let $S_i^b = x_{i1}^b$, $Z_i^b = (S_i^b, 1)'$ and $p_i^b(\beta) = y_{i1} - \exp(-\Delta x_{i2}^b\beta) y_{i2}$. Because $\Delta x_{i2}^b = \Delta x_{i2}$, $p_i^b(\beta)$ = $p_i^0(\beta)$. Thus, we can easily show that $m_i^b(\beta) = C(b)' m_i(\beta)$, where $C(b)$ is defined as in (11). Thus, we obtain the results (12) and (13); that is, Ξ_{wc}^b depends on the choice of b, while Ξ_c does not.

The GMM-C estimator is invariant to b even in finite samples. To see this, set H = 1. Then, the minimands of the one-step and two-step GMM-C estimators have the same forms as (14) and (15), respectively. Thus, both the one-step and two-step GMM-C estimators are invariant to b.

The second quasi-differencing method is the one by Wooldridge (1997, endnote 2). The differenced error function he considered is

$$q_i^0(\beta) = \exp(-x_{i1}\beta) y_{i1} - \exp(-x_{i2}\beta) y_{i2}.$$

Because $q_i^0(\beta_o) = \eta_i(v_{i1} - v_{i2})$ under (16), $E[m_i(\beta_o)] = 0$, where $S_i^0 = x_{i1}$, $Z_i^0 = (S_i^0, 1)'$, and

$$m_i(\beta) = \begin{pmatrix} m_{wc,i}(\beta) \\ m_{c,i}(\beta) \end{pmatrix} = Z_i^0 q_i^0(\beta). \tag{21}$$

The asymptotic variance of the GMM-C estimator using the moment function (21) is of the form $[M'V^{-1}M]^{-1}$, where

$$M = \begin{pmatrix} M_{wc} \\ M_c \end{pmatrix} = -E\left(Z_i^{0\prime}[-\eta_i x_{i1} v_{i1} + \eta_i x_{i2} v_{i2}]\right) = E\left(Z_i^{0\prime} \eta_i \Delta x_{i2}\right); \tag{22}$$

$$V = \begin{pmatrix} V_{wc,wc} & V_{wc,c} \\ V_{wc,c} & V_{c,c} \end{pmatrix} = E\left(\eta_i^2 (v_{i1} - v_{i2})^2 Z_i^0 Z_i^{0\prime}\right). \tag{23}$$

From now on, we refer to the GMM estimators using the moment functions (18) and (21) as the Chamberlain and Wooldridge GMM-C estimators, respectively.

Comparing (19) and (20) with (22) and (23), we can easily see that the asymptotic distribution of the Chamberlain GMM-C estimator is different from that of the Wooldridge GMM-C estimator unless $\beta_o = 0$. When $\beta_o \neq 0$, we are unable to determine which of the two GMM-C estimators is more efficient unless the data generating process is fully specified.

This paper does not attempt to determine which of the Chamberlain and Wooldridge GMM-C estimators would be preferred for actual data analysis, although it should be an important research agenda. However, we note that the Chamberlain GMM-C estimator would be inconsistent for the cases in which the regressors x_{it} are endogenous and contemporaneously correlated with the errors v_{it} (that is, $\mathrm{cov}(x_{it}, v_{it}) \neq 0$). In contrast, when the Wooldridge transformation is used in GMM, the parameter β can be consistently estimated by using higher order lagged regressors as instruments. For example, if x_{i0} is observed, then $(x_{i0}, 1)'$ is used for Z_i^0 in (22).

As Wooldridge (1997, endnote 2) pointed out, a problem of using the moment functions (21), which, is that $\lim_{\beta \to \infty} m_i(\beta) = 0$ for all i if the support of x_{it} contains only non-negative numbers. A similar problem arises when the support of x_{it} contains only non-positive numbers.

This implies that the GMM estimation using the moment functions (21) may fail to obtain an interior solution. As a treatment to this problem, Windmeijer (2000) proposes using the demeaned regressor $x_{it} - \bar{x}$ instead of the level regressor x_{it}, where $\bar{x} = (NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}x_{it}$. When the demeaned regressor is used in GMM, the moment functions (21) have zero expectations only at the interior point $\beta = \beta_o$ because the support of the demeaned regressor contains both positive and negative numbers.

Demeaning is not the only solution for this computational problem. Use of x_{it}^{b} could be a solution, so long as b is chosen such that the support of the x_{it}^{b} contains both negative and positive values. The Windmeijer estimator is a GMM-C estimator obtained using x_{it}^{b} for x_{it} with $b = \bar{x}$.

Differently from its GMM-WC counterpart, the Woodridge GMM-C estimator is asymptotically invariant to b. To see why, let $q_i^b(\beta)$ = $\exp(-x_{i1}^{b}\beta)y_{i1}$ - $\exp(-x_{i2}^{b}\beta)y_{i2}$; and $m_i^b(\beta) = Z_i^b q_i^b(\beta)$. Then, it can be shown that $m_i^b(\beta) = \exp(b\beta)C(b)'m_i(\beta)$, where $C(b)$ is defined in (11) and $m_i(\beta)$, in (21). This implies that

$$M^{b} = \exp(b\beta_o)C(b)'M \; ; \; V^{b} = \exp(2b\beta_o)C(b)'VC(b),$$

where M and V are defined in (22) and (23). Thus, $[M^{b\prime}(V^{b})^{-1}M^{b}]^{-1} = [M'V^{-1}M]^{-1}$, which indicates that the asymptotic variance of the Woodridge GMM-C estimator is invariant to b.

The one-step and two-step GMM-C estimators using the Wooldridge moment function are however not invariant to b in small samples. To see why, notice that the minimand of the one-step GMM estimator $\tilde{\beta}_c^b$ is given:

$$\begin{aligned}&\left(\Sigma_{i=1}^{N}Z_i^b q_i^b(\beta)\right)'\left[\Sigma_{i=1}^{N}Z_i^b Z_i^{b\prime}\right]^{-1}\left(\Sigma_{i=1}^{N}Z_i^b q_i^b(\beta)\right)\\ &= \exp(2b\beta)\left(\Sigma_{i=1}^{N}Z_i^0 r_i^0(\beta)\right)'\left[\Sigma_{i=1}^{N}Z_i^0 Z_i^{0\prime}\right]^{-1}\left(\Sigma_{i=1}^{N}Z_i^0 r_i^0(\beta)\right).\end{aligned} \tag{24}$$

Clearly, the minimand (24) depends on b and thus, the one-step GMM estimator depends on b. The minimand of the two-step GMM-C estimator $\hat{\beta}_c^b$ is

$$
\begin{aligned}
&\left(\Sigma_{i=1}^{N} Z_i^b q_i^b(\beta)\right)' \left[\Sigma_{i=1}^{N} Z_i^b q_i^b(\tilde{\beta}_c^b) q_i^b(\tilde{\beta}_c^b)' Z_i^{b\prime}\right]^{-1} \left(\Sigma_{i=1}^{N} Z_i^b q_i^b(\beta)\right) \\
&\quad = \frac{\exp(2b\beta)}{\exp(2b\tilde{\beta}_c^b)} \left(\Sigma_{i=1}^{N} Z_i^0 q_i^0(\beta)\right)' \left[\Sigma_{i=1}^{N} Z_i^0 q_i^0(\tilde{\beta}_c^b) q_i^0(\tilde{\beta}_c^b)' Z_i^{0\prime}\right]^{-1} \left(\Sigma_{i=1}^{N} Z_i^0 q_i^0(\beta)\right).
\end{aligned} \tag{25}
$$

The two-step GMM-C estimator depends on b for two reasons. First, the one-step estimator $\tilde{\beta}_c^b$ is not invariant. Second, even if the one-step estimator were the same for all b, the minimand of the two-step GMM estimator still would depend on b.

An invariant estimator that has the same asymptotic distribution as the two-step Wooldridge estimator is the continuous-updating GMM estimator proposed by Hansen, Heaton and Yaron (1996). This estimator minimizes

$$
\begin{aligned}
&\left(\Sigma_{i=1}^{N} Z_i^b q_i^b(\beta)\right)' \left[\Sigma_{i=1}^{N} Z_i^b q_i^b(\beta) q_i^b(\beta)' Z_i^{b\prime}\right]^{-1} \left(\Sigma_{i=1}^{N} Z_i^b q_i^b(\beta)\right) \\
&\quad = \left(\Sigma_{i=1}^{N} Z_i^0 q_i^0(\beta)\right)' \left[\Sigma_{i=1}^{N} Z_i^0 q_i^0(\beta) q_i^0(\beta)' Z_i^{0\prime}\right]^{-1} \left(\Sigma_{i=1}^{N} Z_i^0 q_i^0(\beta)\right)
\end{aligned} . \tag{26}
$$

This minimand is the same for all b. Thus, the continuous-updating GMM estimator should be invariant to b.

3. Simulations

This section investigates the finite-sample properties of the two-step GMM-C and GMM-WC estimators. We also illustrate the invariance property of the GMM-C estimator. Our experiments are carried out with a simple dynamic panel model and two count panel models. The econometric software TSP 4.5 (Hall and Cummins, 2006) is used. When we conduct the simulations, the data are demeaned with their overall means before subtracting b from them.

3.1. Dynamic Panel Data Model

Our first simulation experiment is conducted with a simple dynamic panel model. Data are generated by (1) and (6) with $\beta_o = 0.5$, $\sigma_\alpha^2 = \sigma_\varepsilon^2 = 1$, and $\sigma_v^2 = \sigma_\varepsilon^2 / (1-\beta_o^2)$. The α_i, ε_{it}, and v_{i0} are independently drawn from $N(0,\sigma_\alpha^2)$, $N(0,\sigma_\varepsilon^2)$, and $N(0,\sigma_v^2)$, respectively. Because μ_α = 0, under this setup, the asymptotic variance (or standard deviation) of the GMM-WC estimator using the y_{it}^b instead of y_{it} increases with $|b|$, as shown in (10). When $b=0$, the GMM-WC and GMM-C estimators have the same asymptotic variances.

When the y_{it}^b are used instead of the y_{it}, the moment functions used for the GMM-WC and GMM-C estimators are $m_{wc,i}^b(\beta) = S_i^b r_i^b(\beta)$ and $m_i^b(\beta) = Z_i^b r_i^b(\beta)$, respectively, where

$$S_i^{b\prime} = \text{diag}(y_{i0}^b, (y_{i0}^b, y_{i1}^b), \ldots, (y_{i0}^b, \ldots, y_{i,T-2}^b))\,;\; Z_i^{b\prime} = (S_i^{b\prime}, I_{T-1})\,;$$

$$u_{it}^b(\beta) = y_{it}^b - \beta y_{i,t-1}^b\,;\; r_i^b(\beta) = (\Delta u_{i2}^b(\beta), \ldots, \Delta u_{iT}^b(\beta))'\,.$$

The one-step and two-step GMM-C estimators are obtained by minimizing (14) and (15), respectively, replacing H by the $(T-1)\times(\mathrm{T}-1)$ square matrix $H = [H_{jk}]$ whose $(s,s)^{th}$ elements are two, $(s,s+1)^{th}$ and $(s,s\text{-}1)^{th}$ elements are minus one, and other elements are all zero. The one-step and two-step GMM-WC estimators are obtained by minimizing (14) and (15), respectively, replacing Z_i^b by S_i^b. We generate 1,000 different data sets to find the finite-sample distribution of the GMM-WC and GMM-C estimators.

We can easily show that $Z_i^b = AZ_i^0$ and $r_i^b = r_i^0$, where A is an invertible matrix of the form

$$A = \begin{pmatrix} I_{T(T-1)/2} & -bB \\ 0 & I_{T-1} \end{pmatrix},$$

where $B' = diag(1,(1,1), \ldots ,(1,\ldots,1))$. Thus, both the one-step and two-step GMM-C estimators are invariant to any selection of b as we discussed in Section 2.

Table 1 presents the simulation results from the two-step GMM estimation of the simple dynamic model. For each of the GMM-WC and GMM-C estimators, the reported statistics are *bias*, *rmse* (root mean squared error), *mcsd* (Monte Carlo standard deviation) and *mcmse* (Monte Carlo mean of standard error)). As predicted, the distribution of the GMM-WC estimator changes depending on *b*. For each *N*, both the *mcsd* and *rmse* of the GMM-WC estimator increases as $|b|$ increases, although only in a small margin. In contrast, for given *N*, the GMM-C estimation results are the same for any choice of *b*.

3.2. Count Panel Data Model with Predetermined Regressor

Our second experiment is done with the count panel model with a predetermined explanatory variable. We generate data following Windmeijer (2008):

$$y_{it} \sim \text{Poisson}[\exp(x_{it}\beta + \alpha_i + \varepsilon_{it} - (1/2)\sigma_\varepsilon^2)]\,;$$

$$x_{i1} = \frac{\delta}{1-\rho}\eta_i + \frac{1}{\sqrt{1-\rho^2}}(\theta v_i + w_{i1});$$

$$x_{it} = \rho x_{i,t-1} + \delta\eta_i + \theta\varepsilon_{i,t-1} + w_{it},$$

where t = 1, ... , T. The random variables, α_i, ε_{it}, v_i, and w_{it}, are independently drawn from $N(0,\sigma_\alpha^2)$, $N(0,\sigma_\varepsilon^2)$, $N(0,\sigma_v^2)$, and $N(0,\sigma_w^2)$, respectively. Under this setup, the regressor x_{it} is weakly exogenous to the error term ε_{it}. Thus, both the Chamberlain and Wooldridge GMM estimators can be used to consistently estimate β. We here only consider the Chamberlain GMM estimators.

Define:

$$p_{it}^b(\beta) = y_{it} - \exp(-\Delta x_{i,t+1}^b \beta) y_{i,t+1};$$

$$S_i^{b\prime} = \text{diag}(x_{i1}^b, (x_{i1}^b, x_{i2}^b), ..., (x_{i1}^b, ..., x_{i,T-1}^b)); \; Z_i^{b\prime} = (S_i^{b\prime}, I_{T-1});$$

$$r_i^b(\beta) = (p_{i1}^b(\beta), ..., p_{i,T-1}^b(\beta))'.$$

With this notation, the one-step and two-step Chamberlain GMM-C estimators are computed by minimizing (14) and (15), respectively, with $H = I_{T-1}$. The one-step and two-step Chamberlain GMM-WC estimators are obtained by the same methods but with Z_i^b replaced by S_i^b. The one-step and two-step GMM-C estimators are invariant to b as discussed in Section 2.2. The finite-sample properties of the Wooldridge GMM estimators are examined in the next subsection.

Table 2 presents the Monte Carlo results from the two-step Chamberlain GMM estimation. As predicted, the distribution of the two-step GMM-WC estimator changes as b changes. Interestingly, for given N, the *mcsd* and *rmse* of the GMM-WC estimator are smallest when b = 0.5 (not when $b = 0$) and they increase as b decreases from 0.5. When $N = 100$ and $b = 0.5$, the *mcsd* and *rmse* of the GMM-WC estimator are almost identical to those of the GMM-WC estimator. However, when $N = 100$ and $b = -1$, the *rmse* of the GMM-WC estimator is greater than that of the GMM-C estimator by 80 (= (0.54-0.30)/0.30×100) percent. The mcsd of the GMM-WC estimator is also much larger than that of the GMM-C estimator. The performance of the GMM-WC estimator appears to be quite sensitive to b, especially when the sample size N is small.

3.3. Count Panel Data Model with Endogenous Regressor

Our third and final experiment is based on the count panel model with an endogenous regressor. Following Windmeijer (2000), we generate data by

$$y_{it} \sim \text{Poisson}(\exp(x_{it}\beta + \alpha_i + \varepsilon_{it} - (1/2)\sigma_\varepsilon^2))\,;$$

$$x_{it} = \rho x_{i,t-1} + \delta\eta_i + \theta\varepsilon_{it} + w_{it}\,,$$

where $t = 1, 2, \ldots, T$, and

$$x_{i0} = \frac{\delta}{1-\rho}\alpha_i + \frac{1}{\sqrt{1-\rho^2}}(\theta\varepsilon_{i0} + w_{i0})\,.$$

The random variables, α_i, ε_{it}, and w_{it} are generated independently drawn from $N(0,\sigma_\alpha^2)$, $N(0,\sigma_\varepsilon^2)$, and $N(0,\sigma_w^2)$, respectively. Under this setup, the regressor x_{it} is endogenous because it is contemporaneously correlated with the error term ε_{it}. The Chamberlain GMM estimation is not appropriate. Thus, we estimate β by the Wooldridge GMM. As we discussed in Section 2, both the Wooldridge GMM-WC and GMM-C estimators are not invariant to linear transformation of the regressor x_{it} in finite sample, although the later one is invariant asymptotically.

Let

$$q_{it}^b(\beta) = \exp(-x_{it}^b\beta)y_{it} - \exp(-x_{i,t+1}^b\beta)y_{i,t+1}\,;$$

$$S_i^{b\prime} = \text{diag}(\{x_{i0}^b\},\{x_{i0}^b, x_{i1}^b\},\ldots,\{x_{i0}^b,\ldots,x_{i,T-2}^b\})\,;\; \mathrm{Z}_i^{b\prime} = (S_i^{b\prime}, I_{T-1})\,;$$

$$r_i^b(\beta) = (p_{i1}^b(\beta),\ldots,p_{i,T-1}^b(\beta))'\,.$$

With these, the two-step GMM-WC and GMM-C estimators are computed by minimizing (14) and (15), respectively, with $H = I_{T-1}$.

Table 3 presents the Monte Carlo results for the two-step Wooldridge GMM estimators. The finite sample performances of both the GMM-WC and GMM-C estimators deteriorate as $|b|$ deviates from zero, especially when N = 100. It appears that the performance of the GMM-WC estimator depends on the ratio of positive and negative values of the regressor x_{it}. As the distribution of x_{it} becomes more skewed to the positive or negative sides, the *mcsd* and *rmse* of the GMM-WC estimator get larger. The GMM-C estimator shows a similar pattern. However, the finite-sample performance of the GMM-C estimator is less sensitive to *b* than that of the GMM-WC estimator.

The results reported in Table 3 indicate that both the Wooldridge GMM-C and GMM-WC estimators need to be used with some caution, especially when the sample size *N* is small. When a regressor's realized values in data are too often positive or negative, the two-step GMM-C estimator, as well as its GMM-WC counterpart, can have large bias and large standard error. Windmeijer (2000) reported that use of demeaned regressors could improve the finite-sample properties of the Wooldridge GMM estimators. Under our simulation setting, the x_{it}^{b} with $b = 0$ is the same as the demeaned regressor $x_{it} - \overline{x}$. Thus, the results in Table 3 are consistent with his findings.

The continuous-updating GMM-C estimator could be a viable alternative to the two-step counterpart because it should be invariant to *b*, although we here do not investigate the finite-sample property of the estimator. We leave the analysis of the finite-sample properties of the continuous-updating estimator to a future study.

4. Concluding Remark

In this paper, we have investigated a non-invariance problem in the panel GMM estimation based on the level-instruments-for-differenced-equations (LIDE) approach. Panel studies often first-difference or quasi-difference regression equations to remove the unobservable individual effects. Then, the differenced equations are estimated by GMM using lagged level regressors as instruments. We have shown that when a constant is not used as instrument, the asymptotic and finite-sample distributions of the GMM estimators depend on overall means of the regressors used. When the sample size is small, estimation results could change dramatically when linearly transformed regressors are used. For many cases, this non-invariance problem can be solved simply by including a constant into the instrument set. Using a constant as an instrument often serves as a ballast for stabilizing the LIDE estimators.

The GMM estimators using a constant as an additional instrument have the asymptotic distributions that are invariant to linear transformation of regressors. However, their finite sample distributions may depend on overall means of regressors. One example is the quasi-differencing method of Woodridge (1997, footnote 2). Even if a constant is used as an instrument, the finite-sample distribution of the GMM estimator depends on overall means of regressors, although its asymptotic distribution does not. However, consistent with Windmeijer (2000), our simulation results support the notion that the GMM-C estimator computed with demeaned regressors produces quite reliable inferences.

References

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

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

Altonji, J.G., and L.M. Segal, 1996, Small-sample bias in GMM estimation of covariance structures, *Journal of Business & Economic Statistics*, 14, 353-366.

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

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

Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-component models, *Journal of Econometrics*, 68, 29-51.

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

Burnside, C., and M.S. Eichenbaum, 1996, Small-sample properties of GMM-based Wald tests, *Journal of Business & Economic Statistics*, 14, 294-308.

Chamberlain, G., 1992, Comment: Sequential moment restrictions in panel data, *Journal of Business & Economic Statistics*, 10, 20 – 26.

Crépon, B., F. Kramarz, and A. Trognon, 1997, Parameters of interest, nuisance parameters and orthogonality conditions: An application to autoregressive error component models, *Journal of Econometrics*, 82, 135 – 156.

Hahn, J., 1999, How informative is the initial condition is the dynamic panel data with fixed effects?, *Journal of Econometrics*, 93, 309 – 326.

Hall, B.H., and C. Cummins, 2006, TSP 5.0 user's guide, TSP international.

Hansen, L., 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029 – 1054.

Hansen, L., J. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, 14, 262 – 280.

Nickell, S., 1981, Biases in dynamic models with fixed effects, *Econometrica*, 49, 1417-1426.

Windmeijer, F., 2000, Moment conditions for fixed effects count data models with endogenous regressors, *Economics Letters*, 68, 21 – 24.

Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of *Econometrics*, 126, 25 – 51.

Windmeijer, F., 2008, GMM for panel count data models. *In: László Mátyás and Patrick Sevestre (eds.), The econometrics of panel data. Fundamentals and recent developments in theory and practice,* Springer.

Wooldridge, J., 1997, Multiplicative panel data models without the strict exogeneity assumption, *Econometric Theory*, 13, 667 – 678.

Table 1: Monte Carlo Results for a Dynamic Panel Data Model

($T = 5$, $\beta_o = 0.5$, $\sigma_\varepsilon^2 = \sigma_\alpha^2 = 1$)

		N = 100		N = 500		N = 1000	
		bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcsd)
b = 20	GMM-WC	-0.08	0.18	-0.01	0.08	-0.01	0.05
		(0.17)	(0.14)	(0.08)	(0.07)	(0.05)	(0.05)
	GMM-C	-0.06	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.10)	(0.06)	(0.05)	(0.04)	(0.04)
b = 2	GMM-WC	-0.06	0.16	-0.01	0.06	-0.01	0.05
		(0.15)	(0.12)	(0.06)	(0.06)	(0.04)	(0.04)
	GMM-C	-0.06	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.10)	(0.06)	(0.05)	(0.04)	(0.04)
b = 0	GMM-WC	-0.04	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.11)	(0.06)	(0.05)	(0.04)	(0.04)
	GMM-C	-0.06	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.10)	(0.06)	(0.05)	(0.04)	(0.04)
b = -2	GMM-WC	-0.06	0.16	-0.01	0.07	-0.01	0.05
		(0.15)	(0.12)	(0.06)	(0.06)	(0.05)	(0.04)
	GMM-C	-0.06	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.10)	(0.06)	(0.05)	(0.04)	(0.04)
b = -20	GMM-WC	-0.08	0.18	-0.01	0.08	-0.01	0.05
		(0.17)	(0.14)	(0.08)	(0.07)	(0.05)	(0.05)
	GMM-C	-0.06	0.14	-0.01	0.06	-0.01	0.04
		(0.13)	(0.10)	(0.06)	(0.05)	(0.04)	(0.04)

Notes: The number of replications is 1,000. The standard errors of the two-step GMM estimators are computed by the usual GMM formulas. The finite variance correction proposed by Windmeijer (2005, 2008) is not used.

Table 2: Monte Carlo Results from Count Panel Data Models with Predetermined Regressors
($T = 5$, $\beta_o = 0.5$, $\delta = 0.1$, $\rho = 0.8$, $\theta = 0.3$, $\sigma_\alpha^2 = \sigma_\varepsilon^2 = 0.3$, $\sigma_w^2 = 0.25$)

		$N = 100$		$N = 500$		$N = 1000$	
		bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcsd)
$b = 1$	GMM-WC	-0.09	0.33	-0.03	0.17	-0.01	0.11
		(0.32)	(0.25)	(0.17)	(0.15)	(0.11)	(0.11)
	GMM-C	-0.14	0.30	-0.03	0.13	-0.02	0.09
		(0.27)	(0.17)	(0.13)	(0.11)	(0.09)	(0.08)
$b = 0.5$	GMM-WC	-0.09	0.29	-0.03	0.13	-0.01	0.09
		(0.28)	(0.20)	(0.13)	(0.11)	(0.09)	(0.09)
	GMM-C	-0.14	0.30	-0.03	0.13	-0.02	0.09
		(0.27)	(0.17)	(0.13)	(0.11)	(0.09)	(0.08)
$b = 0$	GMM-WC	-0.14	0.36	-0.03	0.15	-0.02	0.10
		(0.33)	(0.21)	(0.14)	(0.12)	(0.10)	(0.08)
	GMM-C	-0.14	0.30	-0.03	0.13	-0.02	0.09
		(0.27)	(0.17)	(0.13)	(0.11)	(0.09)	(0.08)
$b = -0.5$	GMM-WC	-0.24	0.51	-0.09	0.30	-0.03	0.15
		(0.45)	(0.24)	(0.29)	(0.13)	(0.14)	(0.09)
	GMM-C	-0.14	0.30	-0.03	0.13	-0.02	0.09
		(0.27)	(0.17)	(0.13)	(0.11)	(0.09)	(0.08)
$b = -1$	GMM-WC	-0.27	0.54	-0.13	0.34	-0.05	0.21
		(0.46)	(0.25)	(0.32)	(0.15)	(0.20)	(0.10)
	GMM-C	-0.14	0.30	-0.03	0.13	-0.02	0.09
		(0.27)	(0.17)	(0.13)	(0.11)	(0.09)	(0.08)

Notes: The number of replications is 1,000. The GMM-C estimates of β are the same for the data demeaned with different values of b while the GMM-WC estimates are not. Three different starting values of β were used in nonlinear GMM optimization processes. This table displays the Monte Carlo results from the estimation using the true value of β ($\beta_o = 0.5$) as the starting value. In almost all replications, three different starting values produced the same estimation results. The standard errors of the two-step GMM estimators are computed by the usual GMM formulas. The finite variance correction proposed by Windmeijer (2005, 2008) is not used.

Table 3: Monte Carlo Results from Count Panel Data Models with Predetermined Regressors
($T=5$, $\beta_o=0.5$, $\delta=0.1$, $\rho=0.8$, $\theta=0.3$, $\sigma_\alpha^2=0.3$, $\sigma_\varepsilon^2=0.25$, $\sigma_w^2=0.3$)

		N = 100		N = 500		N = 1000	
		bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcmse)	bias (mcsd)	rmse (mcsd)
b = 1	GMM-WC	-0.48	0.52	-0.22	0.25	-0.13	0.16
		(0.22)	(0.22)	(0.12)	(0.13)	(0.10)	(0.10)
	GMM-C	-0.39	0.44	-0.17	0.21	-0.10	0.13
		(0.19)	(0.18)	(0.11)	(0.11)	(0.08)	(0.08)
b = 0.5	GMM-WC	-0.21	0.27	-0.09	0.14	-0.05	0.10
		(0.17)	(0.19)	(0.10)	(0.12)	(0.08)	(0.09)
	GMM-C	-0.18	0.24	-0.08	0.13	-0.05	0.09
		(0.17)	(0.16)	(0.10)	(0.11)	(0.08)	(0.08)
b = 0	GMM-WC	0.01	0.16	0.00	0.10	0.00	0.08
		(0.16)	(0.19)	(0.10)	(0.12)	(0.08)	(0.09)
	GMM-C	0.01	0.16	0.01	0.10	0.00	0.08
		(0.16)	(0.17)	(0.10)	(0.11)	(0.08)	(0.08)
b = -0.5	GMM-WC	0.26	0.33	0.10	0.15	0.06	0.11
		(0.20)	(0.22)	(0.11)	(0.13)	(0.09)	(0.09)
	GMM-C	0.22	0.29	0.10	0.15	0.06	0.10
		(0.19)	(0.18)	(0.11)	(0.12)	(0.09)	(0.09)
b = -1	GMM-WC	0.67	0.77	0.28	0.33	0.16	0.20
		(0.38)	(0.30)	(0.17)	(0.16)	(0.12)	(0.11)
	GMM-C	0.56	0.64	0.23	0.28	0.13	0.17
		(0.32)	(0.24)	(0.16)	(0.13)	(0.11)	(0.09)

Notes: The number of replications is 1,000. For both the GMM-WC and GMM-C estimation, different estimates of β are obtained from the data demeaned with different values of b. Three different starting values of β were used in nonlinear GMM optimization processes. This table displays the Monte Carlo results from the estimation using the true value of β ($\beta_o=0.5$) as the starting value. In almost all replications, three different starting values produced the same estimation results. The average rates of the positive x_{it}^b in each replication are approximately 15 percent when $b=1$, 30 percent when $b=0.5$, 50 percent when $b=0$, 70 percent when $b=-0.5$, and 85 percent for $b=-1$. The standard errors of the two-step GMM estimators are computed by the usual GMM formulas. The finite variance correction proposed by Windmeijer (2005, 2008) is not used.