Bias in Dynamic Panel Models under Time Series Misspecification

Yoonseok Lee*

September 2009

Abstract

This paper considers higher-order autoregressive panel models with exogenous regressors and fixed effects, where the lag order is possibly misspecified. The within-group estimator is considered and its asymptotic biases are studied. Specifically, we extend the N-asymptotic bias formula in Nickell (1981) to the case that the dynamics follow general autoregressive forms. We also develop NT-asymptotic distribution of the within-group estimator that allows for lag order misspecification. Besides the bias incurred by the within-group transformation in dynamic fixed-effects models, additional bias under misspecification is obtained, which is of the same asymptotic order as the fixed-effect bias. We suggest some bias reduction methods under the possible misspecification.

*Keywords***:** Bias, dynamic panel, fixed effects, misspecification, bias reduction.
*JEL Classifications***:** C23, C33

1 Introduction

Since the influential papers by Nerlove (1967) and Nickell (1981), finite sample autoregressive bias in fixed-effect dynamic panel models has been well understood and many bias reduction methods were proposed in the context of within-group (or least squares dummy variables; LSDV) estimators (e.g., Kiviet, 1995; Hahn and Kuersteiner, 2002; Alvarez and Arellano,

*First draft: March 2005. An earlier version of this paper is the part of my dissertation at Yale University and it was circulated as "A General Approach to Bias Correction in Dynamic Panels under Time Series Misspecification." I thank Peter Phillips, Donald Andrews, Yuichi Kitamura, and seminar participants at Cornell, MSU, Texas A&M, SUNY-Binghamton, Yale, 2005 Inter University Conference at Princeton, 2007 Midwest Econometrics Group Meeting, and Conference in Honor of Peter C. B. Phillips. All errors are solely mine. *Address:* Department of Economics, University of Michigan, 611 Tappan Street, 365C Lorch Hall, Ann Arbor, MI 48109-1220. *E-mail:* yoolee@umich.edu.

2003; Bun and Carree, 2005; and Phillips and Sul, 2007 to name a few). Dynamic panel studies, however, have long relied on the first-order autoregressive structure, which is indeed unavoidable especially when the length of the time series (T) is small. As longer panel data become available, it is more natural to consider higher order dynamics and thus first-order models are most likely misspecified. This paper evaluates the effects of this type of misspecification particularly on the fixed-effect bias in the dynamic panel regressions.

Specifically, we extend the bias formula of Nickell (1981) to the case that the dynamics follow general autoregressive forms with exogenous regressors, $ARX(p)$, but the lag order p could be misspecified. We also develop $\sqrt{NT}$-normalized limit distribution for the within-group estimator that allows for lag order misspecification, when both the cross section sample size N and T get large at the same rate. Besides the order misspecification bias and the fixed-effect bias, the analytical results reveal an additional bias, which is generated from combining order misspecification and incidental parameters problem (Neyman and Scott, 1948). The additional bias is, however, still of the same order of magnitude as the standard fixed-effect bias (i.e., $O(1/T)$) and thus it is possible to develop a bias reduction method. Though attempts to adjust for the bias using formulae that correct for first-order dynamics could be wrong and even exacerbate the bias under such misspecification, it is found that we can reduce the fixed-effect bias robustly to the lag order misspecification using the penalized likelihood function approach (e.g., Hahn and Kuersteiner, 2004; Arellano and Hahn, 2006; Bester and Hansen, 2007) or the model selection based approach (e.g., Lee, 2006).

2 The Model

We consider a panel process $\{y_{i,t}\}$ generated from the homogeneous pth-order univariate autoregressive model with exogenous regressors $X_{i,t} \in \mathbb{R}^r$ (i.e., $(ARX\,(p))$) given by

$$y_{i,t} = \mu_i + \sum_{j=1}^{p} \alpha_{pj} y_{i,t-j} + \beta' X_{i,t} + u_{i,t} \quad \text{for } i = 1, \cdots, N \text{ and } t = 1, \cdots, T, \tag{1}$$

where the lag order p is assumed to be finite.[1] We let the initial values $(y_{i,0}, y_{i,-1}, \cdots, y_{i,-p+1})$ be observed for all i. We first assume the following conditions.

[1] As noted in Bhansali (1981) and Kunitomo and Yamamoto (1985), we can consider the infinite p case given suitable definitions of the infinite dimensional autoregressive coefficient matrix A and the choice matrices e_p and J_q. This is the case of estimating an approximate $ARX\,(p_T)$ model with $p_T \to \infty$, where $p_T^3/T \to 0$ and $T^{1/2} \sum_{j=p+1}^{\infty} |\alpha_{pj}| \to 0$ as $T \to \infty$ (e.g., Bhansali, 1978). See Lee (2006) for further discussions.

Assumption E *(i)* $\{u_{i,t}\}$ *is i.i.d. across* i *and* t*; (ii)* $\mathbb{E}(u_{i,t}|X_{i,1},\cdots,X_{i,T},\mu_i) = 0$, $\mathbb{E}u_{i,t}^2 = \sigma^2$ *with* $0 < \sigma^2 < \infty$ *and* $\mathbb{E}u_{i,t}^8 < \infty$ *for all* i *and* t*; (iii)* $\{X_{i,t}\}$ *is i.i.d. across* i *and strictly stationary in* t *for each* i*; (iv)* $||\mathbb{E}(X_{i,t}X_{i,t}')|| < \infty$.

Assumption S $\sum_{j=1}^{p} |\alpha_{pj}| < \infty$ *and all roots of the characteristic equation* $1-\sum_{j=1}^{p} \alpha_{pj} z^j = 0$ *lie outside the unit circle.*

Assumption E implies weak (or sequential) exogeneity in $y_{i,t-1}$ but strict exogeneity in $X_{i,t}$. Note that $X_{i,t}$ could be serially correlated; but we assume that the higher order lags of $y_{i,t}$ and $X_{i,t}$ capture all the persistence and thus the error term does not have any serial correlation. Assuming independence over t makes the expressions simple. We also exclude cross sectional dependence in $u_{i,t}$. Finally note that the eighth moment condition is required to derive joint asymptotic CLT in the later section. We allow for non-zero correlation between the unobserved individual effect μ_i and $X_{i,t}$ (i.e., fixed effects).

We eliminate fixed effects μ_i by subtracting the individual sample average over time (i.e., within-group transformation) from equation (1) to obtain

$$y_{i,t}^0 = \sum_{j=1}^{p} \alpha_{pj} y_{i,t-j}^0 + \beta' X_{i,t}^0 + u_{i,t}^0, \tag{2}$$

where for any variable $w_{i,t}$ we define $w_{i,t-j}^0 = w_{i,t-j} - \overline{w}_{i,\cdot-j}$ and $\overline{w}_{i,\cdot-j} = (1/T)\sum_{s=1}^{T} w_{i,s-j}$ for $j = 0, 1, \cdots, p$. The formula (2) readily transforms into the first-order p-dimensional vector autoregressive process with exogenous regressors given by

$$Y_{i,t}^0 = AY_{i,t-1}^0 + BX_{i,t}^0 + U_{i,t}^0, \tag{3}$$

where $Y_{i,t-j}^0 = (y_{i,t-j}^0, y_{i,t-j-1}^0, \cdots, y_{i,t-j-p+1}^0)'$ for $j = 0, 1$, and

$$A = \begin{bmatrix} \alpha(p)' \\ I_{p-1},\ 0 \end{bmatrix}$$

with $\alpha(p) = (\alpha_{p1}, \alpha_{p2}, \cdots, \alpha_{pp})'$ and I_k as the identity matrix of rank k. We also let $U_{i,t}^0 = e_p u_{i,t}^0$ and $B = e_p \otimes \beta'$, where e_p is the $p \times 1$ column vector with one in the first element and zeros elsewhere. Note that, from (3), Assumption S is equivalent to $\det[I_p - Az] \neq 0$ for all $|z| \leq 1$, or that each eigenvalue of A has modulus less than one. It thus guarantees that the sequence $\{A^j\}$ is absolutely summable and $\sum_{j=0}^{\infty} A^j = (I_p - A)^{-1}$ exists. Hence, if we define vector linear processes $V_{i,t} = \sum_{j=0}^{\infty} A^j U_{i,t-j}$ and $Z_{i,t} = \sum_{j=0}^{\infty} A^j BX_{i,t-j}$ with $U_{i,t} = e_p u_{i,t}$, then both $V_{i,t}$ and $Z_{i,t}$ exist in the mean square sense and we can rewrite

(3) as $Y_{i,t}^0 = Z_{i,t}^0 + V_{i,t}^0$. Moreover, if we let $\Gamma_j = \mathbb{E}(V_{i,t}V_{i,t+j}')$ for $j = 0, 1, \cdots$, we have $\Gamma_j = \Gamma_0 A'^j$, where $\Gamma_0 = \sigma^2 \sum_{j=0}^{\infty} A^j e_p e_p' A'^j$. For later convenience we also introduce the $p \times p$ long-run covariance matrix of $V_{i,t}$ as $\Omega = \sum_{j=-\infty}^{\infty} \Gamma_j = \Lambda' + \Gamma_0 + \Lambda$, where $\Lambda = \sum_{j=1}^{\infty} \Gamma_j$ and $\Lambda' = \sum_{j=1}^{\infty} \Gamma_j' = \sum_{j=-1}^{-\infty} \Gamma_j$. Assumptions E and S guarantee that Ω exists.

In most cases, the true lag order p of the underlying autoregressive process $\{y_{i,t}\}$ in (1) is unknown. Hence we consider the situation that $\{y_{i,t}\}$ is fitted to an $ARX\,(q)$ process instead of $ARX\,(p)$, where $1 \leq q \leq p$.[2] By stacking cross section observations first and then time series observations, (3) can be rewritten as $Y^0 = Y_{-1}^0 A' + X^0 B' + U^0$, from which the within-group estimators in this case are defined as

$$\widehat{\alpha}\,(p,q) = \left(J_q' Y_{-1}^{0\prime} Q_{X^0} Y_{-1}^0 J_q\right)^{-1} J_q' Y_{-1}^{0\prime} Q_{X^0} Y^0 e_p \tag{4}$$
$$\widehat{\beta} = \left(X^{0\prime} X^0\right)^{-1} X^{0\prime} \left(Y^0 e_p - Y_{-1}^0 J_q \widehat{\alpha}\,(p,q)\right), \tag{5}$$

where $Q_{X^0} = I_{NT} - X^0(X^{0\prime}X^0)^{-1}X^{0\prime}$, $\widehat{\alpha}\,(p,q) = (\widehat{\alpha}_{p,q1}, \widehat{\alpha}_{p,q2}, \cdots, \widehat{\alpha}_{p,qq})'$ and $\widehat{\alpha}_{p,qr}$ for $r = 1, 2, \cdots, q$ is the within-group estimator for the coefficient of $y_{i,t-r}$ when $ARX\,(p)$ process is fitted to $ARX\,(q)$. The $p \times q$ choice matrix J_q is defined as

$$J_q = \begin{cases} [I_q, 0]' & \text{if } q < p; \\ I_p & \text{if } q = p. \end{cases}$$

We further assume the invertibility of $J_p' Y_{-1}^{0\prime} Q_{X^0} Y_{-1}^0 J_p$ and $X^{0\prime}X^0$, which is implied by the following condition.

Assumption R *(i) For given p, $Y_{-1}^{0\prime} Q_{X^0} Y_{-1}^0$ is a full column rank matrix; (ii) X is a full column rank matrix and it does not include time-invariant variables; (iii) $1 \leq q \leq p < \infty$.*

3 Bias Formulae

When the number of lags p is not correctly chosen, and especially when the data is fitted to a lower order ARX model, the within-group estimators in (4) and (5) are expected to have variable omission bias on top of the standard fixed-effect bias. Specifically, we let $\widehat{\Gamma}_0^{Y|X} = Y_{-1}^{0\prime} Q_{X^0} Y_{-1}^0 / NT$, $\widehat{\Gamma}_1^{Y|X} = Y_{-1}^{0\prime} Q_{X^0} Y^0 / NT$, $\Gamma_0^{Y|X} = \text{plim}_{N\to\infty} Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0 / NT + \Gamma_0$ and $\Gamma_1^{Y|X} = \text{plim}_{N\to\infty} Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0 A' / NT + \Gamma_1$. Then, using $C^{-1} = \left(I - C^{-1}\left(C - C_0\right)\right) C_0^{-1}$

[2] The case of $q > p$ is less interesting since there is no variable omission bias. In this case, we simply let $\alpha(p,q) = (\alpha(p)', 0_{q-p}')'$. An interesting example is when $(p,q) = (0,1)$ without exogenous regressors $X_{i,t}$ (i.e., $y_{i,t} = \mu_i + u_{i,t}$ is the data generating model). In this case, for the within-group estimator $\widehat{\alpha}(0,1) = \sum_{i=1}^N \sum_{t=1}^T u_{i,t-1}^0 u_{i,t}^0 / \sum_{i=1}^N \sum_{t=1}^T \left(u_{i,t-1}^0\right)^2$, we can derive that $\text{plim}_{N\to\infty} \widehat{\alpha}(0,1) = -T^{-1}$.

for symmetric invertible matrices C and C_0, we can rewrite (4) as

$$\begin{aligned} \widehat{\alpha}(p,q) - J_q'\alpha(p) &= \left\{\alpha(p,q) - J_q'\alpha(p)\right\} + \left\{\widehat{\alpha}(p,q) - \alpha(p,q)\right\} \\ &= \left\{\alpha(p,q) - J_q'\alpha(p)\right\} + \Big\{(J_q'\Gamma_0^{Y|X}J_q)^{-1}J_q'(\widehat{\Gamma}_1^{Y|X} - \Gamma_1^{Y|X})e_p \\ &\quad -(J_q'\widehat{\Gamma}_0^{Y|X}J_q)^{-1}[J_q'(\widehat{\Gamma}_0^{Y|X} - \Gamma_0^{Y|X})J_q](J_q'\Gamma_0^{Y|X}J_q)^{-1}J_q'\widehat{\Gamma}_1^{Y|X}e_p\Big\} \\ &\equiv b_1(\widehat{\alpha}(p,q)) + b_2(\widehat{\alpha}(p,q)), \end{aligned} \tag{6}$$

where $\alpha(p,q) = (J_q'\Gamma_0^{Y|X}J_q)^{-1}J_q'\Gamma_1^{Y|X}e_p$ is the theoretical parameter value from the $ARX(q)$ fitting. In addition, since we can write $ARX(q)$ regression under the data generating process (3) as

$$J_q'Y_{i,t}^0 = J_q'AJ_qJ_q'Y_{i,t-1}^0 + J_q'BX_{i,t}^0 + J_q'U_{i,t}^0,$$

(5) can be rewritten as

$$\begin{aligned} \widehat{\beta} - \beta &= \left\{\left(X^{0\prime}X^0\right)^{-1}X^{0\prime}Y_{-1}^0J_q\left(J_q'\alpha(p) - \alpha(p,q)\right)\right\} \\ &\quad + \left\{\left(X^{0\prime}X^0\right)^{-1}X^{0\prime}Y_{-1}^0J_q\left(\alpha(p,q) - \widehat{\alpha}(p,q)\right)\right\} + o_p(1) \\ &\equiv b_1(\widehat{\beta}) + b_2(\widehat{\beta}) + o_p(1). \end{aligned} \tag{7}$$

by letting $\left(X^{0\prime}X^0\right)^{-1}X^{0\prime}U^0e_p = o_p(1)$ from the strict exogeneity. Note that in these two bias expressions (6) and (7), $b_1(\widehat{\alpha}(p,q))$ and $b_1(\widehat{\beta})$ are the variable omission biases, which cannot be eliminated unless the model is correctly specified. This part of bias, therefore, should be took care of by proper lag order selection methods (e.g., Lee, 2006, 2009a). On the other hand, $b_2(\widehat{\alpha}(p,q))$ and $b_2(\widehat{\beta})$ are the biases from the within-group transformation. In this section, we will show that the second part of the biases have different expressions from the standard fixed-effect bias formula and they indeed have additional terms polluted by the misspecification. $b_2(\widehat{\alpha}(p,q))$ and $b_2(\widehat{\beta})$ are, however, shown to be still $O_p(1/T)$, which will be disappear as $T \to \infty$, whereas the misspecification biases $b_1(\widehat{\alpha}(p,q))$ and $b_1(\widehat{\beta})$ do not vanish even when $N, T \to \infty$.

The main interest here is, therefore, in the biases $b_2(\widehat{\alpha}(p,q))$ and $b_2(\widehat{\beta})$, which are manageable, instead of the entire biases $\widehat{\alpha}(p,q) - J_q'\alpha(p)$ and $\widehat{\beta} - \beta$.[3] This is the case when the researcher believes $ARX(q)$ is true and she tries to find the bias formula from the pa-

[3] It seems more reasonable to derive the bias from the true parameter values if we consider the parameters that are not model specific. From (7), however, it is apparent that even the parameter estimators that are not model specific have bias components from misspecification. Note that the bias from the theoretical parameter value is also studied in the standard time series literature under misspecification (e.g., Bhansali, 1981; Kunitomo and Yamamoto, 1985).

rameters in $ARX(q)$ in order to correct the fixed-effect bias. A leading example is fitting $AR(\infty)$ process using finite number of lags. The main lesson should be that the standard bias formula for the within-group estimator is no longer valid even around the theoretical parameter values (i.e., even we disregard the misspecification bias) and thus the standard bias correction methods would not work properly in this case.

3.1 Nickell bias

We first examine the asymptotic bias of the within-group estimator when N tends to infinity but T is fixed. Even in the case of correct specification, it is well known that the standard within-group estimator for $AR(1)$ fixed-effects models is not consistent for large N (e.g., Nerlove, 1967; Nickell, 1981). As well described in Phillips and Sul (2007), such autoregressive bias arises from the correlation of the error and the lagged dependent variables after the unknown mean is estimated to be removed. Not surprisingly, such bias becomes more complicated when the lag order is not correctly specified. The first theorem generalizes the Nickell bias to the case with time series misspecification.

Theorem 1 *Let $\{y_{i,t}\}$ be generated from (1) and $S_X = \text{plim}_{N\to\infty} Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^{0}/NT$ exists. Under Assumptions E, R and S,*

$$\underset{N\to\infty}{plim}\left(\widehat{\alpha}(p,q)-\alpha(p,q)\right) = -\left(I_q - R_{q,01}\right)^{-1} R_{q,12} - \left(I_q - R_{q,01}\right)^{-1}\left\{R_{q,11} - R_{q,01}(D_q^X)^{-1}J_q'(S_X A' + \Gamma_1)e_p\right\}, \quad (8)$$

where $D_q^X = (J_q'(\Gamma_0 + S_X)J_q)^{-1}$, $R_{q,01} = D_q^X J_q' G_{01} J_q$, $R_{q,11} = D_q^X J_q' G_{11} e_p$ and $R_{q,12} = D_q^X J_q' G_{12} e_p$. Expressions of G_{01}, G_{12} and G_{12} are given as (A.1), (A.2) and (A.3) in Appendix.

The first term of the bias (8) is mainly from the nonzero correlation between $Y_{i,t-1}^0$ and $U_{i,t}^0$, and the second term is mainly from order misspecification (more precisely, from the combination of endogeneity and order misspecification). It can be easily verified that the asymptotic bias of $\widehat{\alpha}(p,q)$ around $\alpha(p,q)$ becomes negligible as T grows. The bias expression of (5) follows easily from (7) and (8). If there is no exogenous regressors, the expression for $\text{plim}_{N\to\infty}\left(\widehat{\alpha}(p,q)-\alpha(p,q)\right)$ remains the same without S_X term, which can be approximated as

$$-\frac{\sigma^2}{T-p} D_q J_q' \left(I_p - A\right)^{-1} e_p - \frac{1}{T-p} D_q J_q' \Omega \left(I_p - J_q D_q J_q' \Gamma_0\right) A' e_p + O\left(\frac{1}{T^2}\right) \quad (9)$$

for large T, where $D_q = (J_q' \Gamma_0 J_q)^{-1}$.

Remark When $q = 1$, (9) could be further reduced to

$$-\frac{\sigma^2 \ell}{(T-p)\gamma_0} - \frac{1}{(T-p)\gamma_0}\left(\sum\nolimits_{k=1}^{p} \alpha_{pk}\omega_{1k} - \frac{\gamma_1}{\gamma_0}\omega_{11}\right) + O\left(\frac{1}{T^2}\right), \tag{10}$$

where $\ell = 1/\left(1 - \sum_{k=1}^{p} \alpha_{2k}\right)$ and $\omega_{1k} = \sum_{j=-\infty}^{\infty} \gamma_{j+k-1}$ with γ_j being the autocovariances of $y_{i,t}$. As a particular example with $(p, q) = (2, 1)$, it can be further simplified as

$$-\frac{1}{T-2}\left((1-\alpha_{22})(1+\rho_1) + \frac{2\alpha_{22}}{1-\rho_1}\right) - \frac{1}{T-2}\left(\alpha_{22}\left(\frac{1+\alpha_{22}}{1-\alpha_{22}}\right)(1+\rho_1)\right) + O\left(\frac{1}{T^2}\right) \tag{11}$$

using the Yule-Walker equation, where $\rho_1 = \alpha_{21}/(1-\alpha_{22})$ is the autocorrelation of $y_{i,t}$ in $AR(2)$, which corresponds to the theoretical coefficient $\alpha(2,1) = \alpha_{2,11}$ in $AR(1)$.[4] Note that, if $p = q = 1$, then $\alpha_{22} = 0$ and (11) is equivalent to the bias expression of Nickell (1981), which could be also seen in (10) by letting $\alpha_{1k} = 0$ for $k \geq 2$ since $\alpha_{11} = \gamma_1/\gamma_0$. An interesting finding is that the absolute value of the bias increases with positive α_{22}. In particular, the second part of the bias, which is mainly from the misspecification, explodes as α_{22} gets close to unity and thus the process becomes the unit root. Similar phenomenon was also discussed by Kunitomo and Yamamoto (1985) in the time series context.

3.2 Noncentrality in the asymptotic distribution

We now consider the asymptotic distribution of the within-group estimator when both N and T are large. Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), and Lee (2008) consider a similar case, where the asymptotic ratio of N to T is a finite constant. More precisely, we assume the following condition.

Assumption NT $\lim_{N,T\to\infty} N/T = \kappa$, *where* $0 < \kappa < \infty$.

Under Assumption NT, it is well known that the standard $\sqrt{NT}$-normalized within-group estimator has nondegenerating asymptotic bias, which is proportional to the limiting sample size ratio, κ. As in Theorem 1, however, the asymptotic bias could increase when the dynamic specification is incorrect. We also assume the following initial condition.

[4] In an independent work by Okui (2008), it is also shown that the within-group estimator of the panel $AR(1)$ coefficient converges to the first order autocorrelation coefficient as $N \to \infty$ and $T \to \infty$ even under dynamic misspecification. This finding is a particular example of Theorem 1 since $\widehat{\alpha}(p,1) - \alpha(p,1) \to_p 0$ as $N \to \infty$ and $T \to \infty$, where $\alpha(p,1)$ should be the first order autocorrelation coefficient for any p by construction.

Assumption I *(i)* $Y_{i,0} \sim iid(\mu_i (I_p - A)^{-1} e_p + \mathbb{E}(Z_{i,t}), var(Z_{i,t}) + \Gamma_0)$ *for each* i*, where* $var(Z_{i,t}) < \infty$*;* *(ii)* $(1/N)\sum_{i=1}^{N}\mu_i^2 = O_p(1)$.

The initial condition in Assumption I is standard in time series models when studying stationary autoregressive process. Since we assume large T, we can use this initial condition without loss of generality as the initial values become less important for longer stationary panels.

Theorem 2 *We assume that* $\lim_{T\to\infty} S_X = \text{plim}_{N,T\to\infty} Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0/NT$ *and it exists. Under Assumptions E, R, S, I and NT,*

$$\sqrt{NT}\left(\widehat{\alpha}\left(p,q\right)-\alpha\left(p,q\right)\right)\to_d \mathcal{N}\left(-\sqrt{\kappa}\Phi_q^X, \sigma^2 D_q J_q'(\Gamma_0+\lim_{T\to\infty} S_X)J_q D_q\right) \qquad (12)$$

as $N,T\to\infty$ *jointly, where* $\Phi_q^X = D_q J_q'(\sigma^2 (I_p - A)^{-1} M) + \Omega A')e_p - D_q(J_q'\Omega J_q)(J_q'(\Gamma_0 + \lim_{T\to\infty} S_X)J_q)^{-1} J_q'(\Gamma_1 + \lim_{T\to\infty} S_X)e_p$ *and* $D_q = (J_q'\Gamma_0 J_q)^{-1}$.

The asymptotic distribution of (5) follows easily from (7) and (12). Theorem 2 shows that the asymptotic bias depends on the sample size ratio, κ, and thus large T does not attenuate the bias unless κ is zero. If there is no exogenous regressors, the expression reduces to

$$\sqrt{NT}\left(\widehat{\alpha}\left(p,q\right)-\alpha\left(p,q\right)\right)\to_d \mathcal{N}\left(-\sqrt{\kappa}\Phi_q, \sigma^2 D_q\right) \qquad (13)$$

with $-\sqrt{\kappa}\Phi_q = -\sqrt{\kappa}\sigma^2 D_q J_q'\left(I_p - A\right)^{-1} e_p - \sqrt{\kappa} D_q\left\{J_q'\Omega A' - J_q'\Omega J_q D_q J_q'\Gamma_1\right\} e_p$, where the first component corresponds to the well-known negative bias from the within-group transformation, which always underestimates $\alpha\left(p,q\right)$ even when the number of lags p is correctly chosen. Meanwhile, the direction of the second component of the bias is ambiguous since it depends on the nuisance parameters, and thus the overall direction of the asymptotic bias is not clear. As discussed in Alvarez and Arellano (2003), we can further derive that $\sqrt{NT}\left(\widehat{\alpha}\left(p,q\right)-\left[\alpha\left(p,q\right)-\left(1/T\right)\Phi_q\right]\right)\to_d \mathcal{N}\left(0,\sigma^2 D_q\right)$ using a higher order expansion of the bias term (e.g., Lee, 2006) provided that $N/T^3 \to 0$. Note that, under the correct time series specification, we have

$$\sqrt{NT}\left(\widehat{\alpha}\left(p\right)-\alpha\left(p\right)\right)\to_d \mathcal{N}\left(-\sqrt{\kappa}\sigma^2\left(\Gamma_0-\Gamma_1'\right)^{-1} e_p, \sigma^2\Gamma_0^{-1}\right)$$

as Hahn and Kuersteiner (2002). In particular, when $p=1$, we have $\sqrt{NT}\left(\widehat{\alpha}\left(1\right)-\alpha\left(1\right)\right)\to_d \mathcal{N}\left(-\sqrt{\kappa}\left(1+\alpha(1)\right), 1-\alpha(1)^2\right)$.

4 Bias Reductions

Theorems 1 and 2 show that the within-group estimators have additional bias even around the theoretical parameter values when the lag order is not correctly chosen. Therefore, most existing bias corrections would not work properly because the correction formulae assume correct model specification. In fact, attempts to adjust for the bias using formulae that correct for $AR(1)$ dynamics would be wrong and may even exacerbate the bias when the true lag order is larger than one. For example, we consider a panel $AR(p)$ model with $p \geq 2$ and we define a bias-corrected estimator for $\alpha(p, 1)$ as $\widetilde{\alpha}(p, 1) = ((T + 1)/T)\widehat{\alpha}(p, 1) + (1/T)$ following Hahn and Kuersteiner (2002), where $\widehat{\alpha}(p, 1)$ is the within-group estimator. From Theorem 2, $\sqrt{NT}(\widetilde{\alpha}(p, 1) - \alpha(p, 1)) = \sqrt{NT}(\widehat{\alpha}(p, 1) - \alpha(p, 1)) + \sqrt{N/T}(1 + \widehat{\alpha}(p, 1)) \rightarrow_d \mathcal{N}(-\sqrt{\kappa}\phi_c, \sigma^2/\gamma_0)$ as $N, T \rightarrow \infty$, where the non-zero bias after correction is given by $\phi_c = \phi - (1 + \alpha(p, 1))$, since $\sqrt{N/T}(1 + \widehat{\alpha}(p, 1)) \rightarrow_p \sqrt{\kappa}(1 + \alpha(p, 1))$ as $N, T \rightarrow \infty$. Therefore, the bias correction in $\widetilde{\alpha}(p, 1)$, which is originally developed for the correctly specified $AR(1)$ process, does not work properly. It should be emphasized that if $\phi < 0$ then $|\phi_c| > |\phi|$ because $(1 + \alpha(p, 1)) > 0$ from the stationarity condition. Therefore, the bias correction in $\widetilde{\alpha}(p, 1)$, which ignores model misspecification, can even exacerbate the bias especially when the overall asymptotic bias is positive (i.e., $-\kappa\phi > 0$).

Apparently, a precise dynamic specification is important particularly when we attempt to correct the fixed-effect bias. A reasonable approach is to conduct model selection before any bias corrections as suggested in Lee (2006).[5] By doing so, the additional bias terms from the misspecification will disappear asymptotically and we can focus on eliminating the pure fixed-effect bias using the bias formulae derived.

Alternatively, we can address a bias reduction, even under possible lag order misspecification, using the penalized likelihood function approach (e.g., Hahn and Kuersteiner, 2004; Arellano and Hahn, 2006; Bester and Hansen, 2007). In fact, the original formulae were developed under the correct model specification. However, since they are formulated using the HAC estimator for the variance of the scores, we can use this bias correction approach robustly to the dynamic order misspecification. Note that any lag order misspecification yields erroneous serial correlation in the error term (or in the scores in general). More precisely, we can derive the penalty estimators in this context as

[5] Also see Lee (2009a) for a modified lag order selection criterion developed for the $AR(p)$ dynamic panel models with fixed effects. It is important to note that we cannot use the standard information-based lag order selection criteria (e.g., AIC, BIC) in this case. Intuitively, this is because the MLE could not be consistent for small T in the presence of incidental parameters (e.g., the fixed effects).

$$\frac{1}{2T\sigma^2}\sum\nolimits_{\ell=-m}^{m}\sum\nolimits_{t=\max\{1,\ell+1\}}^{\max\{T,T+\ell\}} u_{i,t}u_{i,t-\ell} \tag{14}$$

for some truncation parameter m satisfying $m/T^{1/2} \to 0$, which is based on the HAC estimator for the long-run variance of $u_{i,t}$ and thus indeed allows for general forms of serial correlations in $u_{i,t}$. Note that Bester and Hansen (2007) suggest to use small m (e.g., $m = 1$ for $AR(1)$ model) but we need to use large m as in the standard HAC estimation to cover general forms of serial correlations from possible order misspecifications.

Tables I and II summarizes some Monte Carlo simulation results showing how the bias reductions work even under the lag order misspecification. The true model is $AR(3)$ without exogenous regressors but fitted to $AR(1)$. The true coefficients satisfy $\alpha_{31} = \alpha_{32} = \alpha_{33}$. The theoretical parameter value $\alpha_{3,11}$ is the first order autocorrelation coefficient in the given $AR(3)$ process so that it can be calculated from the true parameter values. μ_i are randomly drawn from $\mathcal{U}(-0.5, 0.5)$ and $u_{i,t}$ from $\mathcal{N}(0,1)$. $\widehat{\alpha}_{3,11}$ is the within-group estimator before bias correction; $\widetilde{\alpha}_{3,11}^{AR(1)}$ is the Hahn-Kuersteiner (2002) type bias-corrected estimator assuming $AR(1)$ is correct; $\widetilde{\alpha}_{3,11}^{R}$ is the bias-corrected estimator from the penalized likelihood using (14), which is supposed to be robust to order misspecification; $\widetilde{\alpha}_{p^*1}$ is the bias-corrected estimator assuming the selected lag p^* is true (e.g., Lee, 2006); $\widetilde{\alpha}_{p^*,11}$ is the bias-corrected estimator using (8) also with the selected p^*. Table I shows the biases from the true parameter values, which thus include the misspecification biases (i.e., $b_1(\widehat{\alpha}(p,q)) + b_2(\widehat{\alpha}(p,q))$ in (6)), whereas Table II gives the biases from the theoretical parameter values so that they only cover $b_2(\widehat{\alpha}(p,q))$ part in (6). The values are obtained from 1000 replications.

The result tells several important points. First, for the original patterns of the fixed-effect bias, Table II shows that it decreases as T gets large, which corresponds to the analytical findings in the main theorems, whereas Table I shows it is not the case when the misspecification bias is also considered. Second, the standard bias correction ($\widetilde{\alpha}_{3,11}^{AR(1)}$) could exacerbate the overall bias (Table I) but it reduces $b_2(\widehat{\alpha}(p,q))$ to some degree (Table II). Interestingly, in Table I, the penalized likelihood based approach ($\widetilde{\alpha}_{3,11}^{R}$) outperforms only for small T case and when the process is very stable (i.e., $\alpha_{31} = 0.1$). Recall that the correction term (14), which is based on the HAC estimator, is $O_p(1/T)$ so that the correction effect decreases as T gets large and thus the robustness to serial-correlation diminishes. In comparison, we can see that the correction method involving model selection ($\widetilde{\alpha}_{p^*1}$ and $\widetilde{\alpha}_{p^*,11}$) outstands for all cases and the performance improves as T increases, which would be mainly because that the correct selection probability of the lag order selection procedure improves with T.

TABLE I: Bias from the true parameter and RMSE of bias corrected estimates

α_{31}	N	T	$\widehat{\alpha}_{3,11}-\alpha_{31}$		$\widetilde{\alpha}_{3,11}^{AR(1)}-\alpha_{31}$		$\widetilde{\alpha}_{3,11}^{R}-\alpha_{31}$		$\widetilde{\alpha}_{p^*1}-\alpha_{31}$	
			bias	*rmse*	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*
0.10	250	25	-0.0450	0.0471	-0.0028	0.0147	-0.0001	0.0145	-0.0113	0.0185
	250	100	0.0083	0.0105	0.0194	0.0204	0.0233	0.0242	-0.0036	0.0065
	250	250	0.0187	0.0192	0.0231	0.0236	0.0251	0.0255	-0.0010	0.0041
0.30	250	25	0.1842	0.1850	0.2457	0.2463	0.2449	0.2454	-0.0634	0.0647
	250	100	0.3963	0.3964	0.4136	0.4137	0.4165	0.4165	-0.0111	0.0123
	250	250	0.4306	0.4306	0.4376	0.4376	0.4408	0.4409	-0.0036	0.0052

TABLE II: Bias from the theoretical parameter and RMSE of bias corrected estimates

$\alpha_{3,11}$	N	T	$\widehat{\alpha}_{3,11}-\alpha_{3,11}$		$\widetilde{\alpha}_{3,11}^{AR(1)}-\alpha_{3,11}$		$\widetilde{\alpha}_{3,11}^{R}-\alpha_{3,11}$		$\widetilde{\alpha}_{p^*,11}-\alpha_{3,11}$	
			bias	*rmse*	*bias*	*rmse*	*bias*	*rmse*	*bias*	*rmse*
0.125	250	25	-0.0700	0.0714	-0.0278	0.0313	-0.0251	0.0290	-0.0361	0.0389
	250	100	-0.0167	0.0179	-0.0056	0.0085	-0.0017	0.0067	-0.0026	0.0070
	250	250	-0.0063	0.0078	-0.0019	0.0050	0.0001	0.0047	-0.0004	0.0047
0.750	250	25	-0.2658	0.2663	-0.2043	0.2050	-0.2051	0.2057	-0.1725	0.1737
	250	100	-0.0537	0.0544	-0.0364	0.0374	-0.0335	0.0345	-0.0252	0.0267
	250	250	-0.0194	0.0204	-0.0124	0.0139	-0.0092	0.0110	-0.0077	0.0100

5 Concluding Remarks

This paper calls into question the simple $ARX\,(1)$ structure in dynamic panels with fixed effects. When the lag orders are unknown, the first-order models are most likely misspecified. In such cases, attempts to adjust for the bias using formulae that correct for $ARX\,(1)$ models would be wrong and may even exacerbate the bias. To address these concerns, we undertake an in-depth investigation of the asymptotic bias of the within-group estimator, where the bias formulae are derived under possible lag order misspecification. It should be noted that, therefore, the main focus of this paper is different from developing bias correction methods, which is robust to the serial correlations in the error (e.g., Hahn and Kuersteiner, 2004). It is closely related with estimating autocorrelations of the error term (e.g., Solon, 1984) but deriving explicit bias formulae of the autoregressive parameters using the nonlinear relations with autocorrelations is limited.

For dynamic panel regression, instrumental variables estimation after first differencing (e.g., Anderson and Hsiao, 1981; Holtz-Eakin, Newey and Rosen, 1988; Arellano and Bond, 1991) is an alternative approach, which does not require any bias correction. However, the instruments are found from the lagged values of the dependent variable with presuming that the error does not have serial correlations. With lag order misspecification, however, the regression error could impose serial correlation and thus the instruments could be no longer valid, which still incurs inconsistency of the estimator.

This paper emphasizes the lag order misspecification in the linear model. When the linearity is in doubt, however, we could consider the nonparametric approach as Lee (2008).

Appendix: Mathematical proofs

For $V_{i,t} = \sum_{j=0}^{\infty} A^j U_{i,t-j}$, we first derive the following lemmas. See Lee (2009b) for the proof.

Lemma A1 *Under Assumptions E and S,*

$$\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} V_{i,t-1}^{0} U_{i,t}^{0\prime} = -\frac{\sigma^2}{T} \left(I_p - A\right)^{-1} \left(I_p - H_T\right) M,$$

$$\operatorname*{plim}_{N\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} V_{i,t-1}^{0} V_{i,t-1}^{0\prime} = \Gamma_0 - \frac{1}{T} \left(I_p - A\right)^{-1} \left(I_p - H_T A\right) \Gamma_0 - \frac{1}{T} \{ \left(I_p - A\right)^{-1} \left(I_p - H_T\right) A \Gamma_0 \}^{\prime},$$

where $M = e_p e_p^{\prime}$ *and* $H_T = \left(I_p - A\right)^{-1} \left(I_p - A^T\right) / T$.

Lemma A2 *Under Assumptions E, S, I and NT,*

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} vec\left(V_{i,t-1}^{0} U_{i,t}^{0\prime}\right) \to_d \mathcal{N}\left(-\sqrt{\kappa}\sigma^2 vec(\left(I_p - A\right)^{-1} M), \sigma^2 \left(M \otimes \Gamma_0\right)\right)$$

as $N, T \to \infty$ *jointly, where* $\Gamma_0 = E V_{i,t} V_{i,t}^{\prime} = \sigma^2 \sum_{j=0}^{\infty} A^j M A^{\prime j}$.

Proof of Theorem 1 Recall that $Y^0 = Y_{-1}^0 A^{\prime} + X^0 B^{\prime} + U^0 = Z^0 + V^0$. Lemma A1 implies that

$$\operatorname*{plim}_{N\to\infty} \widehat{\Gamma}_0^{Y|X} = \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^{0} + \operatorname*{plim}_{N\to\infty} \frac{1}{NT} V_{-1}^{0\prime} V_{-1}^{0} = S_X + \Gamma_0 - G_{01},$$

$$\operatorname*{plim}_{N\to\infty} \widehat{\Gamma}_1^{Y|X} = \operatorname*{plim}_{N\to\infty} \frac{1}{NT} Y_{-1}^{0\prime} Q_{X^0} Y_{-1}^{0} A^{\prime} + \operatorname*{plim}_{N\to\infty} \frac{1}{NT} V_{-1}^{0\prime} U^{0} = S_X A^{\prime} + \Gamma_1 - G_{11} - G_{12},$$

from the strict exogeneity of $X_{i,t}$, where

$$G_{01} = \frac{1}{T} \left[\left(I_p - A\right)^{-1} \left(I_p - H_T A\right) \Gamma_0 + \{ \left(I_p - A\right)^{-1} \left(I_p - H_T\right) A \Gamma_0 \}^{\prime} \right], \tag{A.1}$$

$$G_{11} = \frac{1}{T} \left[\left(I_p - A\right)^{-1} \left(I_p - H_T A\right) \Gamma_1 + \Gamma_1 \{ A \left(I_p - A\right)^{-1} \left(I_p - H_T\right) \}^{\prime} \right], \tag{A.2}$$

$$G_{12} = \frac{\sigma^2}{T} \left(I_p - A\right)^{-1} \left(I_p - H_T\right) M. \tag{A.3}$$

Note that $\Gamma_1 = \Gamma_0 A'$. Then, from (6), the bias expression for $\widehat{\alpha}(p,q)$ follows as

$$\begin{aligned}
&\underset{N\to\infty}{\text{plim}}\ (\widehat{\alpha}(p,q) - \alpha(p,q)) \\
&= -\left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'(G_{11}+G_{12})\,e_p \\
&\quad + \left(I_q - \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q' G_{01} J_q\right)^{-1} \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q' G_{01} J_q \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} \\
&\qquad \times J_q'\left(S_X A' + \Gamma_1 - G_{11} - G_{12}\right) e_p \\
&= -\left(R_{q,11}+R_{q,12}\right) + \left(I_q - R_{q,01}\right)^{-1} R_{q,01}\left(\left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'(S_X+\Gamma_0)A'e - R_{q,11} - R_{q,12}\right) \\
&= -\left(I_q - R_{q,01}\right)^{-1} R_{q,12} - \left(I_q - R_{q,01}\right)^{-1}\left\{R_{q,11} - R_{q,01}\left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'(S_X+\Gamma_0)A'e\right\}
\end{aligned}$$

since $\left(J_q'(S_X+\Gamma_0)J_q - J_q'G_{01}J_q\right)^{-1} = (I_q - \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'G_{01}J_q)^{-1}\left(J_q'(S_X+\Gamma_0)J_q\right)^{-1}$ and by letting $R_{q,01} = \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'G_{01}J_q$, $R_{q,11} = \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'G_{11}e$ and $R_{q,12} = \left(J_q'(S_X+\Gamma_0)J_q\right)^{-1} J_q'G_{12}e$. $\square$

Proof of Theorem 2 First note that

$$\sqrt{NT}(\widehat{\Gamma}_1^{Y|X} - \Gamma_1^{Y|X}) = \sqrt{\frac{N}{T}} \cdot T(\widehat{\Gamma}_0^{Y|X} - \Gamma_0^{Y|X})A' + \frac{1}{\sqrt{NT}} Y_{-1}^{0\prime} Q_{X^0} U^0,$$

in which the first term satisfies $\sqrt{N/T}\cdot T(\widehat{\Gamma}_0^{Y|X} - \Gamma_0^{Y|X})A' \to_p -\sqrt{\kappa}\Omega A'$ as $N,T\to\infty$ from Lemma A1.[6] For the second term,

$$\frac{1}{\sqrt{NT}} Y_{-1}^{0\prime} Q_{X^0} U^0 = \frac{1}{\sqrt{NT}} Z_{-1}^{0\prime} Q_{X^0} U^0 + \frac{1}{\sqrt{NT}} V_{-1}^{0\prime} Q_{X^0} U^0 \equiv C_{N,T}^1 + C_{N,T}^2.$$

It is easy to verify that $vec(C_{N,T}^1) \to_d \mathcal{N}\left(0, \sigma^2\left(M \otimes \text{plim}_{N,T\to\infty}(Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0/NT)\right)\right)$ as $N,T\to\infty$ provided $\text{plim}_{N,T\to\infty}(Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0/NT)$ exists. However, since $X^{0\prime}X^0/NT = O_p(1)$, $X^{0\prime}U^0/\sqrt{NT} = O_p(1)$ and $V_{-1}^{0\prime}X^0/NT = o_p(1)$ from the strict exogeneity, we have

$$C_{N,T}^2 = \frac{1}{\sqrt{NT}} V_{-1}^{0\prime} U^0 - \left(\frac{V_{-1}^{0\prime} X^0}{NT}\right)\left(\frac{X^{0\prime}X^0}{NT}\right)^{-1}\left(\frac{X^{0\prime}U^0}{\sqrt{NT}}\right) = \frac{1}{\sqrt{NT}} V_{-1}^{0\prime} U^0 + o_p(1).$$

Lemma A2 implies that, therefore, $vec(C_{N,T}^2) \to_d \mathcal{N}(-\sqrt{\kappa}\sigma^2 vec((I_p - A)^{-1} M), \sigma^2 (M \otimes \Gamma_0))$ as $N,T\to\infty$. In sum, we have

$$\sqrt{NT} vec(\widehat{\Gamma}_1^{Y|X} - \Gamma_1^{Y|X}) \to_d \mathcal{N}\left(-\sqrt{\kappa}\Psi, \sigma^2\left(M \otimes (\Gamma_0 + \text{plim}_{N,T\to\infty}(Z_{-1}^{0\prime} Q_{X^0} Z_{-1}^0/NT))\right)\right),$$

[6] See Lee (2009b) for the equivalence between the sequential and the joint probability limits in this case based on Phillips and Moon (1999).

where $\Psi = vec(\sigma^2 (I_p - A)^{-1} M) + \Omega A')$ since $C^1_{N,T}$ and $C^2_{N,T}$ are uncorrelated. Therefore, from (6), for a random variable

$$W \sim \mathcal{N}(-\sqrt{\kappa}(\sigma^2 (I_p - A)^{-1} M) + \Omega A')e_p, \sigma^2(\Gamma_0 + \text{plim}_{N,T\to\infty}(Z^{0\prime}_{-1} Q_{X^0} Z^0_{-1}/NT)),$$

we can conclude that

$$\begin{aligned}
&\sqrt{NT}\,(\widehat{\alpha}\,(p,q) - \alpha\,(p,q)) \\
&= (J_q'\Gamma_0^{Y|X} J_q)^{-1} J_q' \{\sqrt{NT}(\widehat{\Gamma}_1^{Y|X} - \Gamma_1^{Y|X})e_p\} \\
&\quad - (J_q'\widehat{\Gamma}_0^{Y|X} J_q)^{-1} [J_q'\sqrt{N/T} \cdot T(\widehat{\Gamma}_0^{Y|X} - \Gamma_0^{Y|X})J_q](J_q'\Gamma_0^{Y|X} J_q)^{-1} J_q'\widehat{\Gamma}_1^{Y|X} e_p \\
&\to_d (J_q'\Gamma_0 J_q)^{-1} J_q' W + (J_q'\Gamma_0 J_q)^{-1}[J_q'\sqrt{\kappa}\Omega J_q](J_q'(\Gamma_0 + \lim_{T\to\infty} S_X)J_q)^{-1} J_q'(\Gamma_1 + \lim_{T\to\infty} S_X)e_p \\
&=_d \mathcal{N}\left(-\sqrt{\kappa}\Phi_q^X, \sigma^2 (J_q'\Gamma_0 J_q)^{-1} J_q'(\Gamma_0 + \lim_{T\to\infty} S_X)J_q(J_q'\Gamma_0 J_q)^{-1}\right),
\end{aligned}$$

where the mean is $\Phi_q^X = (J_q'\Gamma_0 J_q)^{-1} J_q'(\sigma^2 (I_p - A)^{-1} M) + \Omega A')e_p - (J_q'\Gamma_0 J_q)^{-1}(J_q'\Omega J_q)(J_q'(\Gamma_0 + \lim_{T\to\infty} S_X)J_q)^{-1} J_q'(\Gamma_1 + \lim_{T\to\infty} S_X)e_p$ provided $\lim_{T\to\infty} S_X = \text{plim}_{N,T\to\infty} Z^{0\prime}_{-1} Q_{X^0} Z^0_{-1}/NT$ (i.e., equivalence between the sequential and the joint probability limits). □

References

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

[2] Anderson, T.W. and C. Hsiao (1981). Estimation of dynamic models with error components, *Journal of the American Statistical Association*, 76, 598-606.

[3] Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application of employment equations, *Review of Economics Studies*, 58, 277-297.

[4] Arellano, M. and J. Hahn (2006). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects, *CEMFI Working Paper*: No. 0613.

[5] Bester, C.A. and C. Hansen (2007). A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects, *Journal of Business and Economic Statistics*, forthcoming.

[6] Bhansali, R.J. (1978). Linear prediction by autoregressive model fitting in the time domain, *Annals of Statistics*, 6, 224-231.

[7] Bhansali, R.J. (1981). Effects of not knowing the order of an autoregressive process on the mean squared error of prediction−I, *Journal of the American Statistical Association*, 76, 588-597.

[8] Bun, M.J.G. and M.A. Carree (2005). Bias-corrected Estimation in Dynamic Panel Data Models, *Journal of Business and Economic Statistics*, 23, 200-210.

[9] Hahn, J. and G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects, *Econometrica*, 70, 1639-1657.

[10] Hahn, J. and G. Kuersteiner (2004). Bias reduction for dynamic nonlinear panel models with fixed effects, unpublished manuscript, UCLA.

[11] Holtz-Eakin, D., W. Newey, and H. Rosen (1988). Estimating vector autoregressions with panel data, *Econometrica*, 56, 1371-1395.

[12] Kiviet, J.F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel models, *Journal of Econometrics*, 68, 53-78.

[13] Kunitomo, N. and T. Yamamoto (1985). Properties of predictors in misspecified autoregressive time series models, *Journal of the American Statistical Association*, 80, 941-950.

[14] Lee, Y. (2006). *Nonparametric Approaches to Dynamic Panel Modelling and Bias Correction*, Ph.D. dissertation, Yale University.

[15] Lee, Y. (2008). Nonparametric Estimation of dynamic panel models with fixed effects, unpublished manuscript, University of Michigan.

[16] Lee, Y. (2009a). Likelihood based model selection in the presence of incidental parameters, unpublished manuscript, University of Michigan.

[17] Lee, Y. (2009b). Supplementary Appendix to "Bias in Dynamic Panel Models under Time Series Misspecification", available on www-personal.umich.edu/~yoolee /research.html.

[18] Nerlove, M. (1967). Experimental evidence on the estimation of dynamic economic relations from a time series of cross-sections, *Economic Studies Quarterly*, 18, 42-74.

[19] Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations, *Econometrica*, 16, 1-32.

[20] Nickell, S. (1981). Biases in dynamic models with fixed effects, *Econometrica*, 49, 1417-1425.

[21] Okui, R. (2008). Panel AR(1) estimators under misspecification, *Economics Letters*, 101, 210-213.

[22] Phillips, P.C.B. and H.R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057-1111.

[23] Phillips, P.C.B. and D. Sul (2007). Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence, *Journal of Econometrics*, 137, 162-188.

[24] Solon, G. (1984). Estimating autocorrelations in fixed-effects models, NBER Technical Working Paper No.32.