Journal of Econometrics 187 (2015) 312–322

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

IV, GMM or likelihood approach to estimate dynamic panel models when either N or T or both are large

Cheng Hsiao [a,b,c,*], Junwei Zhang [a]

[a] Department of Economics, University of Southern California, USA
[b] WISE, Xiamen University, China
[c] Department of Quantitative Finance, National Tsinghua University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 9 July 2013
Received in revised form
22 September 2014
Accepted 26 January 2015
Available online 18 March 2015

JEL classification:
C12
C18
C23
C26

Keywords:
IV
MLE
GMM
Asymptotic bias
Large N, T

ABSTRACT

We examine the asymptotic properties of IV, GMM or MLE to estimate dynamic panel data models when either N *or* T or both are large. We show that the Anderson and Hsiao (1981, 1982) simple instrumental variable estimator (IV) or maximizing the likelihood function with initial value distribution properly treated (quasi-maximum likelihood estimator) is asymptotically unbiased when either N or T or both tend to infinity. On the other hand, the QMLE mistreating the initial value as fixed is asymptotically unbiased only if N is fixed and T is large. If both N and T are large and $\frac{N}{T} \to c\,(c \neq 0, c < \infty)$ as $T \to \infty$, it is asymptotically biased of order $\sqrt{\frac{N}{T}}$. We also explore the source of the bias of the Arellano and Bond (1991) type GMM estimator. We show that it is asymptotically biased of order $\sqrt{\frac{T}{N}}$ if $\frac{T}{N} \to c\,(c \neq 0, c < \infty)$ as $N \to \infty$ even if we restrict the number of instruments used. Monte Carlo studies show that whether an estimator is asymptotically biased or not has important implications on the actual size of the conventional t-test.

1. Introduction

Panel data involves at least two dimensions: a cross-sectional dimension of size N and a time series dimension of size T. The multi-dimensional asymptotics are much more complicated than the traditional one dimension asymptotics. As pointed out by Phillips and Moon (1999), sequentially applying one dimensional asymptotics can be misleading when both N and T increase at the same or arbitrary rate. For instance, in a dynamic panel data model of the form

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + u_{it}, \quad |\gamma| < 1,\ i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{1}$$

it is well known that the maximum likelihood estimator treating the individual-specific effects α_i and initial values y_{i0} as fixed constants is the covariance estimator (e.g. Hsiao (2003)),

$$\hat{\gamma}_{cv} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \bar{y}_{i,-1})^2}, \tag{2}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$. The estimator $\hat{\gamma}_{cv}$ is consistent and $\sqrt{T}(\hat{\gamma}_{cv} - \gamma)$ is asymptotically normally distributed with mean 0 if N is fixed and T is large. However, Hahn and Kuersteiner (2002) show that the maximum likelihood estimator is asymptotically biased of order $\sqrt{\frac{N}{T}}$ if both the cross-sectional dimension N and the time series dimension T go to infinity and $\frac{T}{N} \to c\,(c \neq 0, c < \infty)$. On the other hand, the Arellano and Bond (1991) type generalized method of moments estimators (GMM) is consistent and asymptotically unbiased if T is fixed and $N \to \infty$. If $\frac{T}{N} \to c^*\,(c^* \neq 0, c^* < \infty)$ as $N \to \infty$, Alvarez and Arellano (2003) show that the GMM estimator of a dynamic panel data model is asymptotically biased of order $\sqrt{c^*}$.

* Corresponding author.
E-mail addresses: chsiao@usc.edu (C. Hsiao), junweizh@usc.edu (J. Zhang).

http://dx.doi.org/10.1016/j.jeconom.2015.01.008
0304-4076/

Two issues arise in the statistical inference for dynamic panel data models: the presence of individual-specific effects and the treatment of initial observations. The process of removing individual-specific effects creates correlations between the lagged regressors and the transformed errors. Standard approach to purge the correlations between the regressors and errors of the equation is to use instrumental variables. However, in the multi-dimensional case, the way the sample moments are constructed to approximate the population moments could have important implications to the asymptotic distribution of the resulting estimators, for instance, the Anderson and Hsiao (1981, 1982) simple instrumental variable estimator (IV) vs. the Arellano–Bond type GMM estimator. So is the treatment of initial values, for instance, the (quasi) maximum likelihood estimator (MLE) treating initial values as fixed constants vs. random variables.

Whether an estimator is asymptotically biased or not has important implications in statistical inference. In this paper, we wish to explore the source of asymptotic bias following the joint limit or sequential limit approach of Phillips and Moon (1999) as well as to find robust estimators that are asymptotically unbiased independent of the way that N or T goes to infinity. We provide the model and estimators in Section 2. Sections 3–5 discuss the asymptotic properties of simple IV, GMM, and (quasi) MLE when either N or T or both go to infinity, respectively. Monte Carlo simulation results of the properties of different estimators with different combination of N and T are in Section 6. Concluding remarks are in Section 7. We use $\to_p$ to denote convergence in probability, $\to_d$ to denote convergence in distribution when sample size goes to infinity, and $N, T \to \infty$ to indicate both N and T go to infinity. Proofs of the asymptotic results are in the Appendix.

2. Model setup and estimators

For ease of exposition, we consider a dynamic panel data model of the form

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + u_{it}, \quad |\gamma| < 1. \tag{3}$$

Assumption 1. u_{it} is independent of α_i and is independently, identically distributed (i.i.d) across i and t with mean 0, variance σ^2, and finite fourth moment.

Assumption 2. α_i are i.i.d. across individuals with $E[\alpha_i] = 0$, $var(\alpha_i) = \sigma_\alpha^2$ and finite fourth order moments.

Let the observed sample be composed of y_{it}, $i = 1, \dots, N$ and $t = 0, 1, \dots, T$. Stacking the $T \times 1$ observed value of y_{it} as $y_i = (y_{i1}, \dots, y_{iT})$ yields

$$y_i = y_{i,-1}\gamma + \tau\alpha_i + u_i, \quad i = 1, \dots, N, \tag{4}$$

where $y_{i,-1} = (y_{i0}, \dots, y_{i,T-1})'$, $u_i = (u_{i1}, \dots, u_{iT})'$ and τ is a $T \times 1$ vector of ones, $\tau = (1, \dots, 1)'$.

The two popular approaches of estimating the common parameter γ are the method of moment approach and the likelihood approach. The simple IV estimator proposed by Anderson and Hsiao (1981, 1982) is to first difference (3) to yield[1]

$$\triangle y_{it} = \triangle y_{i,t-1}\gamma + \triangle u_{it}, \quad i = 1, 2, \dots, N, \; t = 2, \dots, T, \tag{5}$$

where $\triangle = (1-L)$, L denotes the lag operator, $Ly_{it} = y_{i,t-1}$, then use either $y_{i,t-2}$ or $\triangle y_{i,t-2}$ as the instrument for the moment condition

$$E[y_{i,t-2} \triangle u_{it}] = 0 \quad \text{or} \quad E[\triangle y_{i,t-2} \triangle u_{it}] = 0. \tag{6}$$

Then

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} \triangle y_{it} y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=2}^{T} \triangle y_{i,t-1} y_{i,t-2}} = \gamma + \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} \triangle u_{it} y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=2}^{T} \triangle y_{i,t-1} y_{i,t-2}}, \tag{7}$$

or

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \triangle y_{it} \triangle y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=3}^{T} \triangle y_{i,t-1} \triangle y_{i,t-2}} = \gamma + \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} \triangle u_{it} \triangle y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=3}^{T} \triangle y_{i,t-1} \triangle y_{i,t-2}}. \tag{8}$$

The Arellano and Bond (1991) (or Arellano and Bover (1995)) GMM approach first eliminates the individual specific effect α_i through a $(T-1) \times T$ deviation operator A that satisfies the condition that $A\tau = 0$, where τ is a $T \times 1$ vector of ones, then use the instruments that satisfy the condition that

$$E[Z_i A u_i] = 0. \tag{9}$$

For instance, multiplying (4) by the operator

$$A = \begin{bmatrix} -1 & 1 & \dots & 0 & 0 \\ 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix} \tag{10}$$

yields the $T-1$ first difference equations

$$\triangle y_i = \triangle y_{i,-1}\gamma + \triangle u_i, \quad i = 1, \dots, N, \tag{11}$$

where $\triangle y_i = (\triangle y_{i2}, \dots, \triangle y_{iT})'$, $\triangle y_{i,-1} = (\triangle y_{i1}, \dots, \triangle y_{i,T-1})'$. Then

$$E[Au_i u_i' A'] = \sigma^2 \begin{bmatrix} 2 & -1 & \dots & 0 & 0 \\ -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix} = \sigma^2 \Omega_0 = \Omega. \tag{12}$$

Arellano and Bover (1995) suggest using an upper triangular forward orthogonal operator that satisfies $A\tau = 0$, $A'A = Q = I_T - \frac{1}{T}\tau\tau'$ and $AA' = I_{T-1}$, then $u_i^* = Au_i$ with

$$u_{it}^* = c_t \left[u_{it} - \frac{1}{T-t}(u_{i,t+1} + \dots + u_{iT}) \right], \quad t = 1, \dots, T-1, \tag{13}$$

where $c_t^2 = \frac{T-t}{T-t+1}$, and $E[u_i^* u_i^*] = \sigma^2 I_{T-1}$.

Let Z_i be the block diagonal matrix which satisfies condition (9). When A takes the form (11), the instrument Z_i takes the form $Z_i = (q_{it})$, where $q_{it} = (y_{i0}, \dots, y_{i,t-2})'$. When A takes the form of forward deviation, the instrument q_{it} takes the form $(y_{i0}, \dots, y_{i,t-1})'$. The Arellano and Bond (1991) generalized method of moments (GMM) estimator solves γ by minimizing

$$\left(\frac{1}{N}\sum_{i=1}^{N} Z_i A u_i \right)' \left(\frac{1}{N^2}\sum_{i=1}^{N} Z_i A u_i u_i' A' Z_i' \right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' A' u_i \right). \tag{14}$$

The likelihood approach notes that the initial observation y_{i0} is a random variable and considers the joint distribution of $(y_{i0}, y_{i1}, \dots, y_{iT})$, $i = 1, \dots, N$ (e.g. Anderson and Hsiao (1981, 1982)). Assuming the data generating process is the same for all y_{it}, then

$$y_{i0} = \gamma y_{i,-1} + \alpha_i + u_{i0} = (1 + \gamma + \cdots)\alpha_i + \sum_{j=0} u_{i,-j}\gamma^j. \tag{15}$$

[1] Note that the $\triangle y_{i1}$ equation is undefined since $y_{i,-1}$ is unobserved.