Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models[☆]

Cheng Hsiao [a,b], Qiankun Zhou [c,*]

[a] Department of Economics, University of Southern California, University Park, Los Angeles, CA 90089, United States
[b] Department of Quantitative Finance, NTHU and WISE, Xiamen University, China
[c] Department of Economics, Louisiana State University, Baton Rouge LA 70803, United States

ARTICLE INFO

Article history:
Received 17 September 2016
Received in revised form 31 August 2017
Accepted 24 April 2018
Available online xxxx

JEL classification:
C01
C13
C23

Keywords:
Dynamic panel models
Individual effects
Initial values
Projection method
Conditional or unconditional likelihood approach

ABSTRACT

We use a quasi-likelihood function approach to clarify the role of initial values and the relative sample size of the cross-section dimension N and the time series dimension T on the asymptotic properties of estimators for dynamic panel data models with the presence of individual-specific effects. We show that a properly specified quasi-likelihood estimator (QMLE) that uses the Mundlak–Chamberlain approach to condition the unobserved effects and initial values on the observed strictly exogenous covariates is asymptotically unbiased if N goes to infinity whether T is fixed or goes to infinity. Monte Carlo studies are conducted to demonstrate the importance of properly treating initial values in getting valid statistical inference. The simulation results also suggest that to deal with the incidental parameters issues arising from the presence of individual-specific effects or initial values, following the Mundlak's (1978) suggestion to condition on the time series average of individual's observed regressors performs better than conditioning on each observed variable at all different time periods.

1. Introduction

In the estimation of dynamic panel data models with the presence of time-invariant individual effects, three issues have arisen (e.g., Hsiao (2014)): (i) whether the unobserved individual-specific effects should be treated as fixed or random? (2) whether the initial values should be treated as fixed constants or random? (iii) does the relative size of cross-sectional dimension N and time series dimension T matter? We argue in this paper that all three issues matter in obtaining consistent estimation of unknown parameters and obtaining valid statistical inference. We illustrate our points using a quasi-likelihood function approach because it allows us to synthesize all these issues, also because many panel estimators such as the within estimator (e.g., Hsiao (2014)), the Bai (2013a, b) factor estimator or the Phillips (2010, 2015) control function estimator can also be put in this framework.

Because the impact of the presence of time-invariant individual specific effects on the limiting distribution differs between a panel time series model and a model involving other explanatory variables, we consider these issues first in a panel time series setting, then for a general dynamic panel model containing exogenous explanatory variables in Sections 2 and 3, respectively. Section 4 discusses the implication of Chamberlain (1980)–Mundlak (1978) approach to deal with the

☆ This paper was stimulated by the private communication with Jushan Bai. We would like to thank him for pointing out the terms that cancel the impact of individual-specific effects and the impact of correlation between the errors of the equation and the lagged dependent variables. We would also like to thank the editor Oliver Linton, an associate editor, two anonymous referees and Elie Tamer for helpful comments. Partial research support by China NSF #71103004 and #71631004 to the first author is also gratefully acknowledged.

* Corresponding author.
E-mail addresses: chsiao@usc.edu (C. Hsiao), qzhou@lsu.edu (Q. Zhou).

https://doi.org/10.1016/j.jeconom.2018.04.005

issue of incidental parameters. Section 5 considers the case of heteroscedastic errors. Section 6 provides a small scale Monte Carlo study to highlight the issues involved. Concluding remarks are in Section 7. All proofs are in the Appendix.

Throughout this paper, we use $(N, T) \to \infty$ to denote that both N and T jointly go to infinity, "$\to_p$" and "$\to_d$" to denote convergence in probability and in distribution, respectively.

2. A panel time series model

In this section, we discuss the asymptotic properties of the QMLE of a simple panel time series model. We distinguish two cases: inference based on fixed initial and random initial observations.

2.1. The model

There is no loss of generality to consider the following simple model,

$$y_{it} = \rho y_{it-1} + \eta_i + u_{it}, i = 1, \dots, N; t = 1, \dots, T, \tag{2.1}$$

where $|\rho| < 1$ and the initial value y_{i0} is available for $i = 1, \dots, N$. We make the following assumptions:

Assumption A1(a): The errors u_{it} are independent of η_i and are independently and identically distributed (i.i.d.) over i and t with mean zero and constant variance σ_u^2. For ease of notation, we let $\sigma_u^2 = 1$.

Assumption A2: The individual-specific effects η_i are i.i.d. over i with mean zero and variance σ_η^2.

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, $\mathbf{y}_{i,-1} = \left(y_{i0}, \dots, y_{i,T-1}\right)'$, $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$ and 1_T be a $T \times 1$ vector of ones, model (2.1) can be rewritten as a T-equation system of the form,

$$\mathbf{y}_i = \mathbf{y}_{i,-1}\rho + 1_T\eta_i + \mathbf{u}_i,\ \ i = 1, \dots, N. \tag{2.2}$$

2.2. Fixed initial observation

Under the assumption y_{i0} are fixed constants, the quasi-likelihood function takes the form

$$L = \prod_{i=1}^{N} (2\pi)^{-\frac{T}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y}_i - \rho\mathbf{y}_{i,-1}\right)'\mathbf{V}^{-1}\left(\mathbf{y}_i - \rho\mathbf{y}_{i,-1}\right)\right\}, \tag{2.3}$$

where

$$\mathbf{V} = \mathbf{I}_T + \sigma_\eta^2 1_T 1_T',\ \ \mathbf{V}^{-1} = \mathbf{I}_T - \frac{\sigma_\eta^2}{1 + T\sigma_\eta^2} 1_T 1_T'. \tag{2.4}$$

The quasi-maximum likelihood estimator (QMLE) is obtained by maximizing the logarithm of (2.3). When σ_u^2 and σ_η^2 are known, the QMLE is the (naive) generalized least squares (GLS) estimator,

$$\hat{\rho}_{QMLE,f} = \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}'\mathbf{V}^{-1}\mathbf{y}_{i,-1}\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}'\mathbf{V}^{-1}\mathbf{y}_i\right). \tag{2.5}$$

where $_{QMLE,f}$ refers to QMLE treating y_{i0} as fixed constants.

Remark 2.1. Bai (2013a, b) derives (2.5) from the factor analytic framework by minimizing[1,2]

$$\log|\Sigma_N(\boldsymbol{\theta})| + tr\left(\Sigma_N(\boldsymbol{\theta})^{-1}S_N\right), \tag{2.6}$$

where $\boldsymbol{\theta} = \left(\rho, \sigma_\eta^2, \sigma_u^2\right)'$, $\Sigma_N(\boldsymbol{\theta}) = \Gamma\left(\sigma_u^2\mathbf{I}_T + \left(\sigma_\eta^2 + \rho\frac{1}{N}\sum_{i=1}^{N}y_{i0}^2\right)1_T 1_T^{\text{âĂÍ}}\right)\Gamma'$ and $S_N = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$ with $\bar{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i$,[3]

$$\Gamma_{T\times T} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \end{pmatrix}.$$

---

[1] Bai (2013a, b) derived (2.5) under the assumption that $y_{i0} = 0$. However, one may view $y_{i0} = 0$ as a special case of y_{i0} being a constant.

[2] Bai (2013a, b) actually considers a model involving both the individual- and time-specific effects. However, taking the deviation of individual observation from the cross-section mean at time t, $(y_{it} - \bar{y}_t)$, removes the time-specific effects, where $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N}y_{it}$. The transformed model no longer involves time-specific effects. The asymptotic distributions for Bai (2013a, b) model or (2.1) are identical. So for ease of exposition, we just consider (2.1).

[3] For simplicity of exposition, we do not include an intercept term in (2.1). Thus, under our framework, S_N should be just $\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i\mathbf{y}_i'$.

Please cite this article in press as: Hsiao C., Zhou Q., Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. Journal of Econometrics (2018), https://doi.org/10.1016/j.jeconom.2018.04.005.