ARTICLE IN PRESS

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

Contents lists available at ScienceDirect

Journal of Econometrics

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

Statistical inference for panel dynamic simultaneous equations models[☆]

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

[a] Department of Economics, University of Southern California, University Park, Los Angeles, CA 90089, United States

[b] Department of Quantitative Finance, National Tsing Hua University and WISE, Xiamen University, China

[c] Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902, United States

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C01
C30
C32

Keywords:
Panel dynamic simultaneous equations
Maximum likelihood
Instrumental variable
Generalized method of moments
Multi-dimensional asymptotics

ABSTRACT

We study the identification and estimation of panel dynamic simultaneous equations models. We show that the presence of time-persistent individual-specific effects does not lead to changes in the identification conditions of traditional Cowles Commission dynamic simultaneous equations models. However, the limiting properties of the estimators depend on the way the cross-section dimension, N, or the time series dimension, T, goes to infinity. We propose three limited information estimator: panel simple instrumental variables (PIV), panel generalized two stage least squares (PG2SLS), and panel limited information maximum likelihood estimation (PLIML). We show that they are all asymptotically unbiased independent of the way of how N or T tends to infinity. Monte Carlo studies are conducted to compare the performance of the PLIML, PIV, PG2SLS, the Arellano–Bond type generalized method of moments and the Akashi–Kunitomo least variance ratio estimator. We demonstrate that the reliability of statistical inference depends critically on whether an estimator is asymptotically unbiased or not.

1. Introduction

This paper considers statistical inference for panel dynamic simultaneous equations models. There are three unique features in the analysis of panel dynamic simultaneous equations models that are different from that of conventional Cowles Commission dynamic simultaneous equations models (e.g. Hood and Koopmans (1953)): (i) the presence of time-invariant individual specific effects raises the issue of incidental parameters, be the specific effects are considered random or fixed; (ii) the formulation of initial observations; and (iii) the multi-dimensional nature of panel data.

Statistical inference can only be made in terms of observed data. The joint dependence of observed variables raises the possibility that many observational equivalent structures could generate the same observed phenomena (e.g. Hood and Koopmans (1953)). Moreover, given the inertia in human behavior and the institutional and technological rigidities, many people believe that "all interesting economic behaviors is inherently dynamic, dynamic model are the only relevant models" (e.g. Nerlove (2000)). However, the presence of time-invariant individual-specific effects creates correlations between the unobserved individual-specific effects and all current and past realized endogenous variables. Whether the presence of this time-invariant effects affects conditions for identification of a dynamic simultaneous equations model needs to be explored.

Current outcomes depend on past outcomes also raises the issue of how to treat the initial observations. In a time series framework, this is a moot issue when the time dimension, T, goes to infinity because the relevance of the initial observations becomes negligible. However, in a panel framework, there is also a cross-sectional dimension, the impact of initial observation is magnified by the dimension of cross-section, N, even T is large. It turns out that the statistical properties of different simultaneous equations model estimators could depend critically on how initial observation is formulated and the way N or T goes to infinity.

Akashi and Kunitomo (2012) consider several estimators for a dynamic simultaneous equations model, the within group, the generalized methods of moments estimator (GMM), the panel limited information maximum likelihood estimators. They show that the

☆ This work was supported in part by China NSF grant #71131008. We would like to thank M.H. Pesaran and K. Hayakawa for calling our attention to the Grassetti (2011) paper and three referees for helpful comments.

* Corresponding author at: Department of Economics, University of Southern California, University Park, Los Angeles, California 90089, United States.

E-mail addresses: chsiao@usc.edu (C. Hsiao), qiankunz@usc.edu (Q. Zhou).

http://dx.doi.org/10.1016/j.jeconom.2015.03.031
0304-4076/© 2015 Elsevier B.V. All rights reserved.

statistical properties of these estimators depend critically on the way N or T goes to infinity. In particular, if $\frac{N}{T} \to c \neq 0$ as $(N, T) \to \infty$, all these estimators are asymptotically biased. Whether a consistent estimator is asymptotically biased or not plays a pivotal role in the validity of statistical inference (e.g. Hsiao and Zhang (forthcoming)). In this paper, we propose three limited information estimators that are independent of the way N or T or both go to infinity: panel simple instrumental variable estimator (PIV); panel generalized two stage estimator (PG2SLS) and the panel limited information (quasi) maximum likelihood estimator (MLE). We show that the likelihood approach possesses desirable properties independent of the way N or T goes to infinity provided the initial observation is properly formulated. However, if the initial value is mistreated as fixed constants, the likelihood approach is asymptotically biased of order $\sqrt{\frac{N}{T}}$ when $\frac{N}{T} \to c \neq 0$ and $c < \infty$ as $T \to \infty$.

This paper is organized as follows. Section 2 describes the model. Section 3 discusses identification and related transformation of the model. Section 4 discusses MLE and its asymptotic properties for the over-identified model. Section 5 discusses methods of moments and several other related estimators for dynamic system. Section 6 provides two simulations to examine the performance of various estimators. Concluding remarks are at Section 7. All mathematical proofs are provided in the Appendix.

2. The model

We will show that the presence of lagged dependent variables is the source that a consistent estimator could be asymptotically biased when both N and T are large. Therefore, there is no loss of generality to consider a panel dynamic simultaneous equations model of the form

$$\mathbf{B}\mathbf{y}_{it} + \Gamma\mathbf{y}_{i,t-1} + \mathbf{C}\mathbf{x}_{it} = \boldsymbol{\eta}_i + \mathbf{u}_{it},\quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{2.1}$$

where $\mathbf{y}_{it} = (y_{1,it}, y_{2,it}, \ldots, y_{G,it})'$, $\mathbf{y}_{i,t-1} = (y_{1,it-1}, y_{2,it-1}, \ldots, y_{G,it-1})'$ are $G \times 1$ contemporaneous and lagged joint dependent variables, $\mathbf{x}_{it}$ is a $k \times 1$ vector of strictly exogenous variables, $\boldsymbol{\eta}_i$ is a $G \times 1$ vector of time-invariant individual-specific effects. For ease of notation, $\mathbf{y}_{i0}$ are observed. We assume that

Assumption 1 (A1): $\mathbf{u}_{it}$ is independent, identically distributed over i and t with zero mean, and nonsingular covariance matrix Ω_u, and finite eighth moment, and are independent of $\mathbf{x}_{it}$.

Assumption 2 (A2): $\{\boldsymbol{\eta}_i : i = 1, 2, \ldots, N\}$ are iid across individuals with finite fourth moment.

The distinct feature of panel dynamic simultaneous equations models are the joint dependence of $\mathbf{y}_{it}$ and the presence of time persistent effects $\boldsymbol{\eta}_i$ in the ith individual's time series observations. The joint dependence of $\mathbf{y}_{it}$ makes $\mathbf{B} \neq I_G$.

Assumption 3 (A3): $|\mathbf{B}| \neq 0$ and all the roots of $|\mathbf{B} - \lambda\Gamma| = 0$ lie outside the unit circle.

Premultiplying $\mathbf{B}^{-1}$ to (2.1) yields the reduced form specification

$$\mathbf{y}_{it} = \mathbf{H}_1\mathbf{y}_{i,t-1} + \mathbf{H}_2\mathbf{x}_{it} + \boldsymbol{\alpha}_i + \mathbf{v}_{it}, \tag{2.2}$$

where $\mathbf{H}_1 = -\mathbf{B}^{-1}\Gamma$, $\mathbf{H}_2 = -\mathbf{B}^{-1}\mathbf{C}$, $\boldsymbol{\alpha}_i = \mathbf{B}^{-1}\boldsymbol{\eta}_i$ and $\mathbf{v}_{it} = \mathbf{B}^{-1}\mathbf{u}_{it}$. The presence of time-persistent $\boldsymbol{\alpha}_i$ creates correlation between $\mathbf{y}_{it}$, $\mathbf{y}_{i,t-j}$ and $\boldsymbol{\alpha}_i$ for all j. Under A3, $\mathbf{H}_1^n \to 0$ as $n \to \infty$.

3. Identification and methods to remove the individual specific effects

The time-invariant specific effects enter the system (2.1) (or (2.2)) linearly, it can be removed by taking linear difference of an individual's time series observation. The three popular approaches are first differencing (e.g. Anderson and Hsiao (1981, 1982), Hsiao et al. (2002)), forward demeaning (e.g. Alvarez and Arellano (2003), Arellano and Bover (1995)), or long differencing (e.g. Grassetti (2011), Hahn et al. (2007)). The efficiency of an estimator could depend on which way $\boldsymbol{\eta}_i$ is removed and the relevant moment conditions used. However, the goal of this paper is to study if a particular type of estimator is asymptotically biased, or if it is, what is the order of the asymptotic bias, not the exact formula for the bias, we shall freely use either form depending on the ease of demonstration because the order of the asymptotic bias of the estimators to be studied in this paper are not affected by which of these three methods are used.

The first difference considers the system in terms of $\Delta\mathbf{y}_{it} = \mathbf{y}_{it} - \mathbf{y}_{i,t-1}$. The long difference considers the system in terms of $\tilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \mathbf{y}_{i0}$. Taking the first difference yields the system in structural form as

$$\mathbf{B}\Delta\mathbf{y}_{it} + \Gamma\Delta\mathbf{y}_{i,t-1} + \mathbf{C}\Delta\mathbf{x}_{it} = \Delta\mathbf{u}_{it},\quad i = 1, \ldots, N;\ t = 2, \ldots, T, \tag{3.1}$$

or reduced form

$$\Delta\mathbf{y}_{it} = \mathbf{H}_1\Delta\mathbf{y}_{i,t-1} + \mathbf{H}_2\Delta\mathbf{x}_{it} + \Delta\mathbf{v}_{it},\quad i = 1, \ldots, N;\ t = 2, \ldots, T. \tag{3.2}$$

System (3.1) or (3.2) is a complete system if $(\mathbf{y}_{i1} - \mathbf{y}_{i0})$ are fixed constants. However, if the data generating process of $\mathbf{y}_{i0}$ is not different from $\mathbf{y}_{it}$, then $\mathbf{y}_{i0}$ or $\Delta\mathbf{y}_{i1} = \mathbf{y}_{i1} - \mathbf{y}_{i0}$ cannot be treated as fixed constants. Eq. (2.2) implies that

$$\begin{aligned}\mathbf{y}_{i0} &= \mathbf{H}_1\mathbf{y}_{i,-1} + \mathbf{H}_2\mathbf{x}_{i0} + \boldsymbol{\alpha}_i + \mathbf{v}_{i0}\\ &= [I_G - \mathbf{H}_1L]^{-1}\mathbf{H}_2\mathbf{x}_{i0} + [I_G - \mathbf{H}_1L]^{-1}\boldsymbol{\alpha}_i\\ &\quad + [I_G - \mathbf{H}_1L]^{-1}\mathbf{v}_{i0},\end{aligned} \tag{3.3}$$

where L denotes the lag operator, $L\mathbf{y}_{it} = \mathbf{y}_{i,t-1}$. However, $\mathbf{x}_{i0}$, $\mathbf{x}_{i,-1}, \ldots$ are unobservable. Following the approach of Bhargava and Sargan (1983), we assume that

Assumption 4 (A4): $\mathbf{x}_{it}$ is generated by

$$\mathbf{x}_{it} = \boldsymbol{\mu} + \sum_{j=1}^{\infty}\mathbf{b}_j\boldsymbol{\nu}_{i,t-j},\qquad \sum_{j=1}^{\infty}\left|\mathbf{b}_j\right| < \infty, \tag{3.4}$$

where $\boldsymbol{\mu}$ is a $G \times 1$ vector of constants, $\mathbf{b}_j$ is a $G \times G$ matrix of constants, and $\boldsymbol{\nu}_{it}$ is i.i.d over i and t with nonsingular covariance matrix, Hsiao et al. (2002) show that, we can write[1]

$$\begin{aligned}&[I_G - \mathbf{H}_1][I_G - \mathbf{H}_1L]^{-1}\mathbf{H}_2\mathbf{x}_{i0}\\ &= E\left[[I_G - \mathbf{H}_1][I_G - \mathbf{H}_1L]^{-1}\mathbf{H}_2\mathbf{x}_{i0}|\bar{\mathbf{x}}_i\right] + \mathbf{w}_i\\ &= \mathbf{A}\bar{\mathbf{x}}_i + \mathbf{w}_i,\quad i = 1, 2, \ldots, N,\end{aligned} \tag{3.5}$$

where $\mathbf{A}$ is a $k \times k$ constant matrix, $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}$ and $\mathbf{w}_i$ is i.i.d across i with nonsingular covariance matrix. Substituting (3.5) into (3.3) and subtracting $\mathbf{y}_{i0}$ from $\mathbf{y}_{i1}$ yields

$$\begin{aligned}\Delta\mathbf{y}_{i1} &= \mathbf{H}_2\Delta\mathbf{x}_{i1} - [I_G - \mathbf{H}_1]\mathbf{A}\bar{\mathbf{x}}_i + \mathbf{v}_{i1} - [I_G - \mathbf{H}_1L]\\ &\quad\times\left[\mathbf{w}_i - [I_G - \mathbf{H}_1L]^{-1}\mathbf{v}_{i0}\right].\end{aligned} \tag{3.6}$$

[1] For a stationary invertible MA process, $\mathbf{x}_t$ can be equivalently written $\mathbf{x}_{it} = \mathbf{A}(F)\mathbf{x}_{i,t+1} + \boldsymbol{\varepsilon}_{it}$, $\sum_{j=1}^{\infty}\left|\mathbf{A}_j\right| < \infty$ and F denotes the forward operator. (Box and Jenkins (1970), ch. 6). The minimum mean square predictor of $\mathbf{x}_{-j}$, $E\left(\mathbf{x}_{-j}|\mathbf{x}_{i1}, \ldots\right)$ is of the same form across i, (Box and Jenkins (1970), ch. 6). Thus, $[I_G - \mathbf{H}_1][I_G - \mathbf{H}_1L]^{-1}\mathbf{H}_2\mathbf{x}_{i0} = [I_G - \mathbf{H}_1]\left[I_G + \sum_{v=1}\mathbf{H}_1^vL^v\right]\mathbf{H}_2\left[\mathbf{A}(F)\left(\frac{\mathbf{A}(F)^{-1}}{F^v}\right)_+\right]\mathbf{x}_{i1} + \boldsymbol{\varepsilon}_{i0}$ where $\left(\frac{\mathbf{A}(F)^{-1}}{F^v}\right)_+ = \sum_{j=0}^{\infty}\mathbf{b}_{v+j}F^j$. Utilizing the result that $\mathbf{A}_j \to 0$ and $\mathbf{H}_1^j \to 0$ as j increases, the minimum mean square predictor can be approximated arbitrarily well by a finite order forward predictor. For ease of notation, we use $\bar{\mathbf{x}}_i$ in stead of $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots$.