Statistical Inference for Panel Dynamic Simultaneous Equations Models*

Cheng Hsiao† Qiankun Zhou‡

This version May 16, 2014

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 and to demonstrate the sensitivity of statistical inference to the asymptotic bias of an estimator.

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

JEL classification: C01, C30, C32

---

*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.

†Department of Economics, University of Southern California, University Park, Los Angeles, California 90089 chsiao@usc.edu, Department of Quantitative Finance, National Tsing Hua University and WISE, Xiamen University, China.

‡Department of Economics, University of Southern California, University Park, Los Angeles, California 90089 qiankunz@usc.edu.

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 statistical properties of these estimators depend critically on the way N or T goes to infinity. In particular, if $\frac{N}{T} \rightarrow c \neq 0$ as $T \rightarrow \infty$, 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 (2013)). 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 we studied. 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 check 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

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

where $\mathbf{y}_{it} = (y_{1,it}, y_{2,it}, \dots, y_{G,it})'$, $\mathbf{y}_{i,t-1} = (y_{1i,t-1}, y_{2i,t-1}, \dots, y_{Gi,t-1})'$ 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, $\mathbf{y}_{i,0}$ 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): $\{\boldsymbol{\eta}_i : i = 1, 2, \dots N\}$ 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, Pesaran and Tahmiscioglu (2002)), forward demeaning (e.g. Alvarez and Arellano (2003), Arellano and Bover (1995)), or long differencing (e.g. Grassetti (2011), Hahn, Hausman and Kuersteiner (2007)). The efficiency of an estimate 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 depends 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

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

or reduced form

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

System (3.1) or (2.2) 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. Equation (2.2) implies that

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

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. Under

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

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

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, Pesaran and Tahmiscioglu (2002) show that, we can write[1]

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

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 substracting $\mathbf{y}_{i0}$ from $\mathbf{y}_{i1}$ yields

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

Premultiplying (3.6) by $\mathbf{B}$ yields the structural form of $\Delta \mathbf{y}_{i1}$ as

$$\mathbf{B}\Delta \mathbf{y}_{i1} + \mathbf{C}\mathbf{x}_{i1} + \mathbf{A}^* \bar{\mathbf{x}}_i = \boldsymbol{\xi}_i^* + \mathbf{u}_{i1}, \; i = 1, 2, \ldots, N, \tag{3.7}$$

where $\mathbf{A}^* = -\left(\mathbf{B} + \Gamma\right) \mathbf{A}$ and $\boldsymbol{\xi}_i^* = \mathbf{B}\mathbf{w}_i - \left(\mathbf{B} + \Gamma\right) \left[\mathbf{w}_i - \left[I_G - \mathbf{H}_1 L\right]^{-1} \mathbf{v}_{i0}\right]$. A1 and A4 imply that $\boldsymbol{\xi}_i^*$ is independently, identically distributed over i with mean zero and nonsingular constant covariance matrix $\Omega_{\boldsymbol{\xi}^*}$.

Thus, neither the structural form consisting of (3.1) and (3.7), nor the reduced form consisting of (3.2) and (3.6) contains the time-invariant individual-specific effects, $\boldsymbol{\eta}_i$ or $\boldsymbol{\alpha}_i$. It was shown by Binder, Hsiao and Pesaran (2005) that $\mathbf{H}_1$ and $\mathbf{H}_2$ can be consistently estimated by either the (quasi) maximum likelihood method or the GMM method. Using the relations between the structural parameters and reduced form parameters, it can be shown that a necessary and sufficient condition for the identification of the g-th equation (e.g. Hsiao (1983)) is

$$\text{rank}\left(\mathbf{B}_g, \boldsymbol{\Gamma}_g, \mathbf{C}_g\right) = G - 1, \tag{3.8}$$

[1] For a stationary invertible MA process, $\mathbf{x}_t$ can be equivalently written $\mathbf{x}_{it} = \mathbf{A}\left(F\right) \mathbf{x}_{i,t+1} + \boldsymbol{\varepsilon}_{it}$, $\sum_{j=1}^{\infty} |\mathbf{A}_j| < \infty$ 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, $\left[I_G - \mathbf{H}_1\right] \left[I_G - \mathbf{H}_1 L\right]^{-1} \mathbf{H}_2 \mathbf{x}_{i0} = \left[I_G - \mathbf{H}_1\right] \left[I_G + \sum_{v=1} \mathbf{H}_1^v L^v\right] \mathbf{H}_2 \left[\mathbf{A}\left(F\right) \left(\frac{\mathbf{A}(F)^{-1}}{F^v}\right)_+\right] \mathbf{x}_{i1} + \boldsymbol{\varepsilon}_{i0}$. 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 $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots$.

if the prior restrictions on (2.1) is in the form of excluding certain variables from the g-th equation in (2.1), where $(\mathbf{B}_g, \mathbf{\Gamma}_g, \mathbf{C}_g)$ is the matrix formed from the column of $(\mathbf{B}, \mathbf{\Gamma}, \mathbf{C})$ that are zeros on the g-th row. In other words, the presence of individual-specific effects $\boldsymbol{\eta}_i$ doesn't change the conditions for the identification of an equation in (2.1). A necessary condition for the satisfaction of the rank condition (3.8) is that the number of excluded variables from the system (2.1) is no less than $G - 1$.

4 Limited information Quasi-Maximum likelihood (LIML) Estimation of the Transformed System

4.1 The model

Following Anderson and Rubin (1949), we just consider the estimation of a structual equation ignoring the prior restrictions on the other equations of the system (2.1). To illustrate the impact of lag dependent variables on the asymptotic distribution, there is no loss of generality to consider the estimation of the first equation of the following system

$$\begin{aligned} y_{1,it} &= \boldsymbol{\beta}'\mathbf{y}_{2,it} + \gamma_1 y_{1,it-1} + \mathbf{c}_1'\mathbf{x}_{1,it} + \eta_{1i} + u_{1,it} \\ \mathbf{y}_{2,it} &= \Pi_{21}y_{1,it-1} + \Pi_{22}\mathbf{y}_{2,it-1} + \Pi_{23}\mathbf{x}_{1,it} + \Pi_{24}\mathbf{x}_{2,it} + \boldsymbol{\eta}_{2i} + \mathbf{u}_{2,it}, \end{aligned} \tag{4.1}$$

where $\mathbf{y}_{2,it}$ is a $(G-1)\times 1$ vector of the included joint dependent variables, $\mathbf{x}_{1,it}$ and $\mathbf{x}_{2,it}$ are $k_1 \times 1$ and $k_2 \times 1$ vectors of included and excluded exogenous variables. Premultiplying the system (4.1) by

$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & -\boldsymbol{\beta}' \\ 0 & I_{G-1} \end{pmatrix}^{-1},$$

yields the reduced form as

$$\begin{pmatrix} y_{1,it} \\ \mathbf{y}_{2,it} \end{pmatrix} = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & \Pi_{14} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} & \Pi_{24} \end{pmatrix} \begin{pmatrix} y_{1,it-1} \\ \mathbf{y}_{2,it-1} \\ \mathbf{x}_{1,it} \\ \mathbf{x}_{2,it} \end{pmatrix} + \begin{pmatrix} \alpha_{1i} \\ \boldsymbol{\alpha}_{2i} \end{pmatrix} + \begin{pmatrix} v_{1,it} \\ \mathbf{v}_{2,it} \end{pmatrix}, \tag{4.2}$$

where

$$\begin{aligned} \Pi_{11} + \boldsymbol{\beta}'\Pi_{21} &= \gamma_1, \Pi_{13} + \boldsymbol{\beta}'\Pi_{23} = \mathbf{c}_1', \Pi_{12} + \boldsymbol{\beta}'\Pi_{22} = \mathbf{0}', \\ \Pi_{14} + \boldsymbol{\beta}'\Pi_{24} &= \mathbf{0}', \alpha_{1i} + \boldsymbol{\beta}'\boldsymbol{\alpha}_{2i} = \eta_{1i}, v_{1,it} + \boldsymbol{\beta}'\mathbf{v}_{2i} = u_{1,it}. \end{aligned} \tag{4.3}$$

We shall assume that the first equation of (4.1) is identified, i.e.,

$$\text{rank}\begin{pmatrix} \Pi_{12} & \Pi_{14} \\ \Pi_{22} & \Pi_{24} \end{pmatrix} = G - 1.$$

We can take the first difference to eliminate $\boldsymbol{\eta}_i$. Under A4, it yields a system of (3.1) and (3.7). However, the first difference yields a first order moving average error term. The inversion of a $T \times T$ covariance matrix of a moving average error is quite complicated (e.g. Hsiao and Zhang (2013)). Since it is well known that the efficiency of the MLE is invariant to the nonsingular linear transformation, we shall follow Gressetti (2011) to take the long difference, $\tilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \mathbf{y}_{i0}$. Under A4, it yields a system of the form,

$$\mathbf{B}\tilde{\mathbf{y}}_{it} + \Gamma\tilde{\mathbf{y}}_{it-1} + \mathbf{C}\mathbf{x}_{it} + \mathbf{A}^*\bar{\mathbf{x}}_i = \mathbf{u}_{it} + \boldsymbol{\xi}_i^*, \;\; i = 1, 2, \ldots, N; t = 1, \ldots, T, \tag{4.4}$$

That is, the long difference transformed system becomes

$$\begin{aligned} \tilde{y}_{1,it} &= \boldsymbol{\beta}'\tilde{\mathbf{y}}_{2,it} + \gamma_{11}\tilde{y}_{1,it-1} + \mathbf{c}_1'\mathbf{x}_{1,it} + \mathbf{a}_1^{*\prime}\bar{\mathbf{x}}_i + \xi_{1i}^* + u_{1,it} \\ \tilde{\mathbf{y}}_{2,it} &= \Pi_{21}\tilde{y}_{1,it-1} + \Pi_{22}\tilde{\mathbf{y}}_{2,it-1} + \Pi_{23}\mathbf{x}_{1,it} + \Pi_{24}\mathbf{x}_{2,it} + \mathbf{A}_2^*\bar{\mathbf{x}}_i + \boldsymbol{\xi}_{2i}^* + \mathbf{u}_{2,it}, \end{aligned} \tag{4.5}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

4.2 Log-likelihood function

In this section, we will study the limited information quasi-maximum likelihood (LIML) estimation of the system (4.1). For ease of exposition, we assume $G = 2$ and $k = 2$, because the asymptotic distribution of the QMLE essentially takes the same form for models with more than two variables but algebraically much more tedious. Specifically, we will consider the following model

$$\begin{aligned} y_{1,it} &= \beta y_{2,it} + \gamma_{11}y_{1,it-1} + c_{11}x_{1,it} + \eta_{1i} + u_{1,it} \\ y_{2,it} &= \gamma_{21}y_{1,it-1} + \gamma_{22}y_{2,it-1} + c_{21}x_{1,it} + c_{22}x_{2,it} + \eta_{2i} + u_{2,it}, \end{aligned} \tag{4.6}$$

where $(\gamma_{22}, c_{22}) \neq 0$. After taking long difference, we have

$$\begin{aligned} \tilde{y}_{1,it} &= \beta\tilde{y}_{2,it} + \gamma_{11}\tilde{y}_{1,it-1} + c_{11}x_{1,it} + \mathbf{a}_1^{*\prime}\bar{\mathbf{x}}_i + \xi_{1i}^* + u_{1,it} \\ \tilde{y}_{2,it} &= \gamma_{21}\tilde{y}_{1,it-1} + \gamma_{22}\tilde{y}_{2,it-1} + c_{21}x_{1,it} + c_{22}x_{2,it} + \mathbf{a}_2^{*\prime}\bar{\mathbf{x}}_i + \xi_{2i}^* + u_{2,it}, \end{aligned} \tag{4.7}$$

where $\bar{\mathbf{x}}_i = (\bar{x}_{1i}, \bar{x}_{2i})'$. The reduced form of (4.7) is

$$\begin{aligned} \begin{pmatrix} \tilde{y}_{1,it} \\ \tilde{y}_{2,it} \end{pmatrix} &= \mathbf{B}^{-1}\Gamma\begin{pmatrix} \tilde{y}_{1,it-1} \\ \tilde{y}_{2,it-1} \end{pmatrix} + \mathbf{B}^{-1}\mathbf{C}\mathbf{x}_{it} + \mathbf{B}^{-1}\mathbf{A}^*\bar{\mathbf{x}}_i + \mathbf{B}^{-1}\begin{pmatrix} \xi_{1i}^* \\ \xi_{2i}^* \end{pmatrix} + \mathbf{B}^{-1}\begin{pmatrix} u_{1it} \\ u_{2it} \end{pmatrix} \\ &= \Pi\tilde{\mathbf{y}}_{i,t-1} + \Psi\mathbf{x}_{it} + \Theta\bar{\mathbf{x}}_i + \boldsymbol{\xi}_i + \mathbf{v}_{it}, \end{aligned} \tag{4.8}$$

where

$$\begin{aligned}
\mathbf{B} &= \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \end{pmatrix}, \mathbf{C} = \begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix}, \mathbf{A}^{*}=\begin{pmatrix} \mathbf{a}_1^{*\prime} \\ \mathbf{a}_2^{*\prime} \end{pmatrix}, \\
\Pi &= \mathbf{B}^{-1}\Gamma, \Psi = \mathbf{B}^{-1}\mathbf{C}, \Theta = \mathbf{B}^{-1}\mathbf{A}^{*}, \\
\boldsymbol{\xi}_i &= \mathbf{B}^{-1}\boldsymbol{\xi}_i^{*} = \mathbf{B}^{-1}\begin{pmatrix} \xi_{1i}^{*} \\ \xi_{2i}^{*} \end{pmatrix}, \mathbf{v}_{it} = \begin{pmatrix} v_{1,it} \\ v_{2,it} \end{pmatrix} = \mathbf{B}^{-1}\begin{pmatrix} u_{1,it} \\ u_{2,it} \end{pmatrix},
\end{aligned}$$

where $\boldsymbol{\xi}_i$ and $\mathbf{v}_{it}$ are independent by construction.

Stacking the ith individual's T time series observations in the vector form yields

$$\begin{aligned}
\underset{T\times 1}{\tilde{\mathbf{y}}_{1i}} &= \tilde{\mathbf{y}}_{2i}\beta + \tilde{\mathbf{y}}_{1i,-1}\gamma_{11} + \mathbf{x}_{1i}c_{11} + \mathbf{a}_1^{*\prime}\bar{\mathbf{x}}_i 1_T + \xi_{1i}^{*} \otimes 1_T + \mathbf{u}_{1i} \\
\underset{T\times 1}{\tilde{\mathbf{y}}_{2i}} &= \tilde{\mathbf{Y}}_{i,-1}\boldsymbol{\gamma}_2 + \mathbf{X}_i\mathbf{c}_2 + \mathbf{a}_2^{*\prime}\bar{\mathbf{x}}_i 1_T + \xi_{2i}^{*} \otimes 1_T + \mathbf{u}_{2i},
\end{aligned} \tag{4.9}$$

where $\tilde{\mathbf{Y}}_{i,-1} = (\tilde{\mathbf{y}}_{1i,-1}, \tilde{\mathbf{y}}_{2i,-1})$, $\boldsymbol{\gamma}_2 = (\gamma_{21}, \gamma_{22})'$, $\mathbf{X}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i})$ and $\mathbf{c}_2 = (c_{21}, c_{22})'$.

Let $\mathbf{Z}_i = (\tilde{\mathbf{y}}_{2i}, \tilde{\mathbf{y}}_{1i,-1}, \mathbf{x}_{1i}, 1_T\bar{\mathbf{x}}_i')$ and $\mathbf{W}_i = \left(\tilde{\mathbf{Y}}_{i,-1}, \mathbf{X}_i, 1_T\bar{\mathbf{x}}_i'\right) = (\tilde{\mathbf{y}}_{1i,-1}, \tilde{\mathbf{y}}_{2i,-1}, \mathbf{x}_{1i}, \mathbf{x}_{2i}, 1_T\bar{\mathbf{x}}_i')$, rewrite (4.9) as

$$\begin{pmatrix} \tilde{\mathbf{y}}_{1i} \\ \tilde{\mathbf{y}}_{2i} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_i & 0 \\ 0 & \mathbf{W}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} + \bar{\mathbf{V}}_i, \tag{4.10}$$

where $\boldsymbol{\delta}_1 = (\beta, \gamma_{11}, c_{11}, \mathbf{a}_1^{*\prime})'$ and $\boldsymbol{\delta}_2 = (\boldsymbol{\gamma}_2', \mathbf{c}_2', \mathbf{a}_2^{*\prime})'$, and $\bar{\mathbf{V}}_i = \boldsymbol{\xi}_i^{*} \otimes 1_T + \mathbf{U}_i$ and $\mathbf{U}_i = (\mathbf{u}_{1i}', \mathbf{u}_{2i}')'$, and $1_T = (1, \ldots, 1)'$. Then

$$\Omega_{\bar{V}} = E\left(\bar{\mathbf{V}}_i\bar{\mathbf{V}}_i'\right) = \Omega_{\xi^*} \otimes 1_T 1_T' + \Omega_u \otimes I_T,$$

where

$$\Omega_u = E\left(\mathbf{u}_{it}\mathbf{u}_{it}'\right) = \begin{pmatrix} \sigma_{u,11} & \sigma_{u,12} \\ \sigma_{u,21} & \sigma_{u,22} \end{pmatrix}, \Omega_{\xi^*} = E\left(\boldsymbol{\xi}_i^{*}\boldsymbol{\xi}_i^{*\prime}\right) = \begin{pmatrix} \sigma_{\xi^*,11} & \sigma_{\xi^*,12} \\ \sigma_{\xi^*,21} & \sigma_{\xi^*,22} \end{pmatrix}.$$

Following Avery (1977), we can express $\Omega_{\bar{V}}$ in terms of eigenvalues, Ω_u, and $\Omega^{*} = \Omega_u + T\Omega_{\xi^*}$, and the product of eigenvectors of $\Omega_{\bar{V}}$ as

$$\Omega_{\bar{V}} = \Omega_u \otimes Q + \Omega^{*} \otimes J,$$

where $Q = I_T - \frac{1}{T}1_T 1_T'$ and $J = \frac{1}{T}1_T 1_T'$. It follows that (e.g. Hsiao (2003))

$$\Omega_{\bar{V}}^{-1} = \Omega_u^{-1} \otimes Q + \Omega^{*-1} \otimes J.$$

Under A1-A4 and the assumption that $\mathbf{u}_{it}$ and $\boldsymbol{\xi}_i^*$ are normally distributed[2], the log-likelihood function (4.10) is given by

$$\begin{aligned}
\log L &= -\frac{NT}{2}\log|\Omega_{\bar{V}}| - \frac{1}{2}\sum_{i=1}^{N}\left\{\left[\tilde{\mathbf{y}}_{1i}' - \boldsymbol{\delta}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \boldsymbol{\delta}_2'\mathbf{W}_i'\right]\Omega_{\bar{V}}^{-1}\left[\tilde{\mathbf{y}}_{1i}' - \boldsymbol{\delta}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \boldsymbol{\delta}_2'\mathbf{W}_i'\right]'\right\} \\
&= -\frac{N(T-1)}{2}\log|\Omega_u| - \frac{N}{2}\log|\Omega^*| \\
&\quad -\frac{1}{2}\sum_{i=1}^{N}\left\{\left[\tilde{\mathbf{y}}_{1i}' - \boldsymbol{\delta}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \boldsymbol{\delta}_2'\mathbf{W}_i'\right]\left[\Omega_u^{-1}\otimes Q\right]\left[\tilde{\mathbf{y}}_{1i}' - \boldsymbol{\delta}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \boldsymbol{\delta}_2'\mathbf{W}_i'\right]'\right\} \\
&\quad -\frac{T}{2}\sum_{i=1}^{N}\left\{\left[\bar{\tilde{y}}_{1i}' - \boldsymbol{\delta}_1'\bar{\mathbf{Z}}_i', \bar{\tilde{y}}_{2i}' - \boldsymbol{\delta}_2'\bar{\mathbf{W}}_i'\right]\Omega^{*-1}\left[\bar{\tilde{y}}_{1i}' - \boldsymbol{\delta}_1'\overline{\mathbf{Z}}_i', \bar{\tilde{y}}_{2i}' - \boldsymbol{\delta}_2'\overline{\mathbf{W}}_i'\right]'\right\}, \qquad (4.11)
\end{aligned}$$

where

$$\bar{\tilde{\mathbf{y}}}_i' = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\tilde{y}_{1,it}\\ \tilde{y}_{2,it}\end{pmatrix} = \begin{pmatrix}\bar{\tilde{y}}_{1i}\\ \bar{\tilde{y}}_{2i}\end{pmatrix}, \overline{\mathbf{Z}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_{it} \text{ and } \overline{\mathbf{W}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{w}_{it},$$

with $\mathbf{z}_{it}$ and $\mathbf{w}_{it}$ being the t-th element of $\mathbf{Z}_i$ and $\mathbf{W}_i$, respectively.

4.3 LIML and its asymptotic properties

The MLE are obtained by choosing the values of $\hat{\Omega}_u$, $\hat{\Omega}_\xi$, $\hat{\boldsymbol{\delta}}_1$ and $\hat{\boldsymbol{\delta}}_2$ that simultaneously satisfy the first order conditions

$$\hat{\Omega}_u = \frac{1}{N(T-1)}\sum_{i=1}^{N}\left\{\left[\tilde{\mathbf{y}}_{1i}' - \hat{\boldsymbol{\delta}}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \hat{\boldsymbol{\delta}}_2'\mathbf{W}_i'\right]\left[I_2\otimes Q\right]\left[\tilde{\mathbf{y}}_{1i}' - \hat{\boldsymbol{\delta}}_1'\mathbf{Z}_i', \tilde{\mathbf{y}}_{2i}' - \hat{\boldsymbol{\delta}}_2'\mathbf{W}_i'\right]'\right\} \qquad (4.12)$$

$$\hat{\Omega}_{\xi^*} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix}\left(\bar{\tilde{y}}_{1i} - \bar{\mathbf{Z}}_i\hat{\boldsymbol{\delta}}_1\right)^2 & \left(\bar{\tilde{y}}_{1i} - \bar{\mathbf{Z}}_i\hat{\boldsymbol{\delta}}_1\right)\left(\bar{\tilde{y}}_{2i} - \bar{\mathbf{W}}_i\hat{\boldsymbol{\delta}}_2\right) \\ \left(\bar{\tilde{y}}_{2i} - \bar{\mathbf{W}}_i\hat{\boldsymbol{\delta}}_2\right)\left(\bar{\tilde{y}}_{1i} - \bar{\mathbf{Z}}_i\hat{\boldsymbol{\delta}}_1\right) & \left(\bar{\tilde{y}}_{2i} - \bar{\mathbf{W}}_i\hat{\boldsymbol{\delta}}_2\right)^2\end{bmatrix}, \qquad (4.13)$$

and

$$\begin{pmatrix}\hat{\boldsymbol{\delta}}_1\\ \hat{\boldsymbol{\delta}}_2\end{pmatrix} = \left\{\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\hat{\Omega}_{\bar{V}}^{-1}\begin{pmatrix}\mathbf{Z}_i & 0\\ 0 & \mathbf{W}_i\end{pmatrix}\right\}^{-1}\left\{\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\hat{\Omega}_{\bar{V}}^{-1}\begin{pmatrix}\tilde{\mathbf{y}}_{1i}\\ \tilde{\mathbf{y}}_{2i}\end{pmatrix}\right\}. \qquad (4.14)$$

Equations (4.12)-(4.14) suggest that the LIML can be obtained by iterating between them conditional on the solutions of the other two equations until convergence[3].

[2] If $\mathbf{u}_{it}$ and $\boldsymbol{\xi}_i^*$ are not normally distributed, maximizing (4.11) yields the quasi-MLE (QMLE). The asymptotic distribution of the QMLE remains centered at the true value of the parameters, except that the covariance matrix of the QMLE is no longer the inverse of the information matrix.

[3] Alternatively, we can obtain the MLE through the EM algorithm (e.g. Dempster et. al. (1977), Shah et. al. (1997), Wang and Fan (2010)).

Theorem 4.1 *Let* $\boldsymbol{\delta} = \left(\boldsymbol{\delta}_1', \boldsymbol{\delta}_2'\right)$, $\boldsymbol{\delta}_1 = (\beta, \gamma_{11}, c_{11}, \mathbf{a}_1^{*\prime})'$, $\boldsymbol{\delta}_2 = (\boldsymbol{\gamma}_2', \mathbf{c}_2', \mathbf{a}_2^{*\prime})'$, $\boldsymbol{\theta} = \left(\sigma_{u,11}, \sigma_{u,12}, \sigma_{u,21}, \sigma_{u,22}, \sigma_{\xi^*,11}, \sigma_{\xi^*,12}, \sigma_{\xi^*,21}, \sigma_{\xi^*,22}\right)$, *and let*

$$\begin{aligned} F_{\delta\delta} &= E\left(\frac{1}{NT}\frac{\partial^2 \log L}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right), F_{\delta\theta} = E\left(\frac{1}{NT}\frac{\partial^2 \log L}{\partial\boldsymbol{\delta}\partial\boldsymbol{\theta}'}\right), \\ F_{\theta\theta} &= E\left(\frac{1}{NT}\frac{\partial^2 \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right). \end{aligned}$$

Under A1-A4, the (limited information) maximum likelihood estimator is consistent and

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \xrightarrow{d} N\left(0, \tilde{F}_{\delta\delta}\right), \tag{4.15}$$

where $\tilde{F}_{\delta\delta} = -\left(F_{\delta\delta} - F_{\delta\theta}F_{\theta\theta}^{-1}F_{\delta\theta}'\right)^{-1}$. *The weak convergence is independent of the way N or T or both go to infinity.*

See appendix for the proof.

Corollary 4.1 *Maximizing the log-likelihood function of N individuals of the $(T-1)$ system equations (4.7) ignoring the fact that $\tilde{\mathbf{y}}_{i1}$ is random (i.e. mistreating $\tilde{\mathbf{y}}_{i1}$ as fixed constants) yields an estimator that is consistent and asymptotically unbiased only if N is fixed and T tends to infinity. If $\frac{N}{T} \to c \neq 0$ and $c < \infty$ as $T \to \infty$, then the QMLE of the system (4.7) is asymptotically biased and the bias is of order $\sqrt{\frac{N}{T}}$.*

Remark 4.1 *The QMLE (4.12)-(4.14) remain consistent and asymptotically normally distributed whether $\boldsymbol{\eta}_i$ are fixed constants or random variables satisfying Assumption 5 below.*

Assumption 5 (A5): The individual-specific effects $\boldsymbol{\eta}_i$ are randomly distributed with mean zero and covariance matrix Ω_η and are independent of $\mathbf{x}_{it}$.

However, if $\boldsymbol{\eta}_i$ are indeed independent of $\mathbf{x}_{it}$, then for each individual i, in addition to the T equations of the form (4.1), there is also the distribution of the initial value $\mathbf{y}_{i0}$ in the form of (3.3). In other words, for each individual i, we have $(T+1)$ equations of the form

$$\mathbf{B}\mathbf{Y}_i^* + \Gamma\mathbf{Y}_{i,-1}^* + \mathbf{C}^*\mathbf{X}_i^* = \mathbf{V}_i^*, \tag{4.16}$$

where $\mathbf{Y}_i^* = (\mathbf{y}_{i0}, \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT})$, $\mathbf{Y}_{i,-1}^* = (\mathbf{0}, \mathbf{y}_{i0}, \ldots, \mathbf{y}_{i,T-1})$,

$$\mathbf{X}_i^* = \begin{pmatrix} \bar{\mathbf{x}}_i & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i1} & \ldots & \mathbf{x}_{iT} \end{pmatrix}, \mathbf{C}^* = \begin{pmatrix} \mathbf{A} \\ \mathbf{C} \end{pmatrix},$$

and $\mathbf{V}_i^* = \boldsymbol{\eta}_i \otimes 1_{T+1}' + (\mathbf{u}_{i0}^*, \mathbf{u}_{i1}, \ldots, \mathbf{u}_{iT})$ with $E\left(\mathbf{V}_i^*\right) = \mathbf{0}$ and

$$Cov\left(vec\left(\mathbf{V}_i^*\right)\right) = \Omega_V^* = \Omega_u \otimes \begin{pmatrix} \mathcal{W} & \mathbf{0} \\ \mathbf{0} & I_T \end{pmatrix} + \Omega_\eta \otimes \begin{pmatrix} \overline{\mathcal{W}} & \nabla \otimes 1_T' \\ \nabla' \otimes 1_T & 1_T 1_T' \end{pmatrix},$$

where $\mathcal{W} = \Omega_u^{-1}\Omega_{\mathbf{u}_{i0}^*}$, $\overline{\mathcal{W}} = \Omega_\eta^{-1}(I-\Pi)^{-1}\Omega_{\xi^*}(I-\Pi')^{-1}$, $\nabla = \Omega_\eta^{-1}Cov(\boldsymbol{\eta}_i, \mathbf{y}_{i0}'\mathbf{B}')$. Conditional on Ω_V^*, the MLE of $\hat{\boldsymbol{\delta}}_1$ and $\hat{\boldsymbol{\delta}}_2$ is of the form (4.14), where $\Omega_{\bar{V}}$ is replaced by Ω_V^* and $\mathbf{Z}_i$ is now a $(T+1)\times(G+k+k_1)$ matrix of the form

$$\mathbf{Z}_i = \left[\mathbf{Y}_{2i}^{*\prime}, \mathbf{Y}_{1i,-1}^{*\prime}, (\bar{\mathbf{x}}_i, \mathbf{x}_{1i})\right],$$

and $\mathbf{W}_i$ is a $(T+1)\times(G+k+k_1+k_2)$ matrix of the form,

$$\mathbf{W}_i = \left[\mathbf{Y}_{i,-1}^{*\prime}, \mathbf{X}_i^{*\prime}\right],$$

where $\mathbf{Y}_{2i}^* = (\mathbf{y}_{2,i0}, \mathbf{y}_{2,i1}, \ldots, \mathbf{y}_{2,iT})$. The random effects QMLE is consistent and asymptotically unbiased independent of the way N or T or both tend to infinity.

Remark 4.2 *If A5 holds, the random effects QMLE (REQMLE) has several advantages over the fixed effects QMLE (FEQMLE). First, the REQMLE combines $T+1$ time series observations for each i while the FEQMLE uses only T time series observations. Second, the REQMLE uses a weighted average of between group variation and within group variation while the FEQMLE only uses the within group variation*[4]*. Typically, the between group variation is much larger than the within group variation. Third, the FEQMLE can not estimate the impact of time-invariant variables, but the REQMLE can. On the other hand, the FEQMLE remains consistent and asymptotically unbiased even $\mathbf{x}_{it}$ and $\boldsymbol{\eta}_i$ are correlated as long as $E\left(\mathbf{x}_{it}\mathbf{u}_{js}'\right) = 0$.*

5 Other Estimators

Because $\mathbf{x}_{it}$ is strictly exogenous, there is no loss of generality to consider the asymptotic properties of different estimators by simply consider a dynamic system without exogenous variables as in Akashi and Kunitomo (2012).

5.1 IV estimation

As discussed in section 2, we can remove the individual specific effects by first differencing (4.1) (under the assumption that $\mathbf{c}_1 = 0$, $\Pi_{23} = 0$ and $\Pi_{24} = 0$) to yield

$$\Delta y_{1,it} = \boldsymbol{\beta}'\Delta\mathbf{y}_{2,it} + \gamma_1\Delta y_{1,it-1} + \Delta u_{1,it} \quad (5.1)$$

$$\Delta\mathbf{y}_{2,it} = \Pi_{21}\Delta y_{1,it-1} + \Pi_{22}\Delta\mathbf{y}_{2,it-1} + \Delta\mathbf{u}_{2,it}, \quad (5.2)$$

[4] Following the approach of Maddala (1971), we can decompose the inverse of the asymptotic covariance matrix of the REQMLE as the sum of the inverse of the asymptotic covariance matrix of the FEQMLE and a positive definite matrix.

where $t = 2, \ldots, T$ and $i = 1, \ldots, N$. Although both $\Delta y_{1,it}$ and $\Delta \mathbf{y}_{2,it}$ are correlated with $\Delta u_{1,it}$. However,

$$E\left(\mathbf{y}_{2i,t-2} \Delta u_{1,it}\right) = 0 \text{ and } E\left(\Delta \mathbf{y}_{i,t-2} \Delta u_{1,it}\right) = 0. \tag{5.3}$$

Just like Anderson and Hsiao (1981, 1982), we can use either $\mathbf{y}_{i,t-2}$ or $\Delta \mathbf{y}_{2i,t-2}$ as instruments for equation (5.1). Consequently, the panel simple IV (PIV) estimator for $\boldsymbol{\beta}$ and γ_1 are

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{IV} \\ \hat{\gamma}_{1,IV} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} \Delta \mathbf{y}_{2,it} \\ \Delta y_{1,it-1} \end{pmatrix} \mathbf{y}'_{i,t-2}\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \mathbf{y}_{i,t-2} \Delta y_{1,it}\right], \tag{5.4}$$

or

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_{IV} \\ \hat{\gamma}_{1,IV} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=3}^{T} \begin{pmatrix} \Delta \mathbf{y}_{2,it} \\ \Delta y_{1,it-1} \end{pmatrix} \Delta \mathbf{y}'_{i,t-2}\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta \mathbf{y}_{i,t-2} \Delta y_{1,it}\right]. \tag{5.5}$$

Theorem 5.1 *The PIV estimator (5.4) (or (5.5)) is consistent and asymptotically unbiased independent of the way N or T or both tend to infinity and*

$$\sqrt{NT} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta} \\ \hat{\gamma}_{1,IV} - \gamma \end{pmatrix} \xrightarrow{d} N\left(0, \Omega_{IV}\right),$$

where $\Omega_{IV} = \Xi_1^{-1} \Omega_1 \Xi_1^{-1}$ for (5.4) or $\Omega_{IV} = \Xi_2^{-1} \Omega_2 \Xi_2^{-1}$ for (5.5), with

$$\begin{aligned}
\Xi_1 &= p \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} \Delta \mathbf{y}_{2,it} \\ \Delta y_{1i,t-1} \end{pmatrix} \mathbf{y}'_{i,t-2}, \\
\Omega_1 &= p \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \mathbf{y}_{i,t-2} \mathbf{y}'_{i,t-2} \left(\Delta u_{1,it}\right)^2, \\
\Xi_2 &= p \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} \Delta \mathbf{y}_{2,it} \\ \Delta y_{1,it-1} \end{pmatrix} \Delta \mathbf{y}'_{i,t-2}, \\
\Omega_2 &= p \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{y}_{i,t-2} \Delta \mathbf{y}'_{i,t-2} \left(\Delta u_{1,it}\right)^2.
\end{aligned}$$

5.2 The (generalized) two-stage least squares estimator (PG2SLS)

Stacking the $(T-1)$ time series observations of (5.1) for the i-th individual yields

$$\begin{aligned}
\Delta \mathbf{y}_{1i} &= \boldsymbol{\beta}' \Delta \mathbf{Y}_{2i} + \gamma_1 \Delta \mathbf{y}_{1i,-1} + \Delta \mathbf{u}_{1i} \\
&= \Delta \mathbf{X}_i \boldsymbol{\theta} + \Delta \mathbf{u}_{1i},
\end{aligned} \tag{5.6}$$

where $\Delta\mathbf{y}_{1i} = (\Delta y_{1,i2}, \ldots, \Delta y_{1,iT})'$, $\Delta\mathbf{y}_{1i,-1} = (\Delta y_{1,i1}, \ldots, \Delta y_{1,iT-1})'$, $\Delta\mathbf{Y}_{2i} = (\Delta\mathbf{y}_{2,i2}, \ldots, \Delta\mathbf{y}_{2,iT})'$, $\Delta\mathbf{u}_{1i} = (\Delta u_{1,i2}, \ldots, \Delta u_{1,iT})'$, $\Delta\mathbf{X}_i = (\Delta\mathbf{Y}_{2i}, \Delta\mathbf{y}_{1i,-1})$ and $\boldsymbol{\theta} = \left(\boldsymbol{\beta}', \gamma_1\right)'$. We note that

$$E\left(\mathbf{y}_{i,t-2}\Delta u_{1,it}\right) = 0, \tag{5.7}$$

and

$$E\left(\Delta\mathbf{u}_{1i}\Delta\mathbf{u}_{1i}'\right) = \sigma_{u,11}A, \tag{5.8}$$

where

$$A = \begin{pmatrix} 2 & -1 & \ldots & 0 \\ -1 & 2 & \ldots & \vdots \\ \vdots & \ldots & \ddots & -1 \\ 0 & \ldots & -1 & 2 \end{pmatrix}.$$

Let $\mathbf{Y}_{i,-2} = (\mathbf{y}_{i0}, \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT-2})$, an analogus panel generalized 2SLS (PG2SLS) estimator can be defined as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{PG2SLS} &= \left\{\left[\sum_{i=1}^{N}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right]\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}_{i,-2}'\right]^{-1}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{X}_i\right\}^{-1} \\ &\quad \times\left\{\left[\sum_{i=1}^{N}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right]\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}_{i,-2}'\right]^{-1}\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{y}_{1i}\right]\right\}. \end{aligned} \tag{5.9}$$

Proposition 5.2 *The PG2SLS estimator of $\boldsymbol{\theta} = \left(\boldsymbol{\beta}', \gamma_1\right)'$ is consistent, asymptotically unbiased and*

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{PG2SLS} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, \Omega_{PG2SLS}\right),$$

where

$$\Omega_{PG2SLS} = \sigma_{u,11}\left\{\frac{1}{NT}\left[\sum_{i=1}^{N}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right]\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}_{i,-2}'\right]^{-1}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{X}_i\right\}^{-1},$$

independent of the way N or T or both tend to infinity.

5.3 Generalized Method of Moment Estimator (GMM)

Either the PIV or PG2SLS is a form of the method of moments estimator. However, in addition to $\mathbf{y}_{i,t-2}$, all past $\mathbf{y}_{i,t-j}$, $j \geq 2$ are legimate instruments. Following Arellano-Bond (1991), one can use the moment conditions

$$E\left(\Delta\mathbf{u}_{1i}\mathbf{Q}_i\right) = 0, \tag{5.10}$$

where $\Delta\mathbf{u}_{1i} = (\Delta u_{1i,2}, \ldots, \Delta u_{1i,T})'$ and $\mathbf{Q}_i$ is a $(T-1) \times K_*$ block diagonal matrix with the t-th block being a vector of the form[5]

$$\mathbf{q}_{it} = (y_{1,i0}, \ldots, y_{1,it-2}, y_{2,i0}, \ldots, y_{2,it-2})', \tag{5.11}$$

where $K_* = T(T-1)$ for model (5.1) and (5.2). Thus, an Arellano-Bond type GMM can be defined as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{GMM} &= \left\{ \left(\sum_{i=1}^{N} \Delta\mathbf{X}_i'\mathbf{Q}_i \right) \left(\sum_{i=1}^{N} \mathbf{Q}_i' A \mathbf{Q}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Q}_i' \Delta\mathbf{X}_i \right) \right\}^{-1} \\ &\quad \times \left\{ \left(\sum_{i=1}^{N} \Delta\mathbf{X}_i'\mathbf{Q}_i \right) \left(\sum_{i=1}^{N} \mathbf{Q}_i' A \mathbf{Q}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Q}_i' \Delta\mathbf{y}_{1i} \right) \right\}^{-1}. \end{aligned} \tag{5.12}$$

To ensure the existence of the inversion of the matrix, $\frac{T}{N} < \frac{1}{2}$, i.e., the time series dimension of the panel data must be less that one half of the cross-section dimension. Akashi and Kunitomo (2012, 2014) show that Arellano-Bond type GMM suffers from:

Proposition 5.3 *The Arellano-Bond type GMM (5.12) is inconsistent if $\frac{T}{N} \to c \neq 0 < \infty$ as $(N,T) \to \infty$. Even when $c = 0$ as $(N,T) \to \infty$, if $\frac{T^3}{N} \to d \neq 0 < \infty$, $\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}\right)$ is asymptotically biased of order $\sqrt{\frac{T^3}{N}}$.*

The Arellano-Bond type GMM uses $(\mathbf{y}_{i0}, \mathbf{y}_{i1}, \ldots, \mathbf{y}_{it-2})$ as instruments for each $\Delta u_{1,it}$ equation. As t increases, so is the number of available instruments. One way to control the impact of ever increasing number of instruments on the asymptotic distribution as T increases is to fix the number of instruments used for each $\Delta u_{1,it}$ equation. Akashi and Kunitomo (2014) show that

Proposition 5.4 *When a fixed number of instruments is used for each $\Delta u_{1,it}$ equation (say, only $\mathbf{y}_{it-2}$ is used), the modified GMM, $\hat{\boldsymbol{\theta}}_{MGMM}$, is consistent as $(N,T) \to \infty$. However, $E\left[\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}\right)\right] = O\left(\sqrt{\frac{T}{N}}\right)$.*

The difference between the Arellano-Bond type GMM and the PIV or PG2SLS is that the former uses the cross-sectional mean $\frac{1}{N}\sum_{i=1}^{N} \mathbf{q}_{it}\Delta u_{1it}$ to approximate the population moments

[5] As pointed out by a referee that "nowadays, it is more and more appreciated that combining this approach with estimating the model in levels with moments in differences greatly adds to the quality of the estimation process." The issue of finding optimal combination of moment is important and deserves further study. However, the purpose of this paper is to study the order of asymptotic bias order, and the order of asymptotic bias is mainly due to the error of approximating the population moments using cross-sectionally mean.

$E\left(\mathbf{q}_{it}\Delta u_{1,it}\right)$, while the latter uses $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{y}_{i,t-2}\Delta u_{1it}$ to approximate the population moments (5.7). Using the cross-sectional mean, the regressor, $\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_{it}\Delta\mathbf{x}_{it}$, and the error $\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_{it}\Delta u_{1it}$ are correlated of order $\frac{1}{N}$. When T is fixed and N is large, they are asymptotically uncorrelated. However, when T is also large, as long as $\frac{T}{N}\to c\neq 0$, the Arellano-Bond type GMM magnifies the impact of the correlation between $\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_{it}\Delta\mathbf{x}_{it}$ and $\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_{it}\Delta u_{1it}$ by the time dimension, T. On the other hand, the PIV or PG2SLS uses the simple or weighted average across i and t to approximate the moment condition $E\left(\mathbf{y}_{i,t-2}\Delta u_{1it}\right)=0$ and the correlation of the transformed regressor of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{y}_{i,t-2}\Delta\mathbf{x}_{it}$ and the transformed error $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{y}_{i,t-2}\Delta u_{1it}$ is of order $\frac{1}{NT}$, hence is asymptotically uncorrelated either N or T or both tend to infinity, and is independent of the value of $\frac{T}{N}$.

5.4 Least Variance Ratio Estimator

Akashi and Kunitomo (2012) propose a panel generalization of the Anderson-Rubin (1949) least variance ratio estimator[6] of $\left(\boldsymbol{\beta},\gamma_1\right)$, where $\hat{\theta}_{LI}=\left(\boldsymbol{\beta}',\gamma_1\right)'$ are obtained by

$$\left(\frac{1}{n}G^{(f)}-\lambda_n\frac{1}{q_n}H^{(f)}\right)\begin{pmatrix}1\\-\hat{\theta}_{LI}\end{pmatrix}=0, \tag{5.13}$$

where

$$\begin{aligned}G^{(f)}&=\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{y}_t^{(1,f)'}\\ \mathbf{Y}_{t\cdot t-1}^{(f)'}\end{pmatrix}M_t\left(\mathbf{y}_t^{(1,f)},\mathbf{Y}_{t\cdot t-1}^{(f)}\right)\\ H^{(f)}&=\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{y}_t^{(1,f)'}\\ \mathbf{Y}_{t\cdot t-1}^{(f)'}\end{pmatrix}\left[I_N-M_t\right]\left(\mathbf{y}_t^{(1,f)},\mathbf{Y}_{t\cdot t-1}^{(f)}\right),\end{aligned}$$

where $\mathbf{y}_t^{(1,f)}=\left(y_{1,1t}^{(f)},\ldots,y_{1,Nt}^{(f)}\right)'$, $\mathbf{y}_t^{(2,f)}=\left(y_{2,1t}^{(f)},\ldots,y_{2,Nt}^{(f)}\right)'$ with $\mathbf{y}_t^{(1,f)}$ and $\mathbf{y}_t^{(2,f)}$ being the forward defferencing transformation, and $\mathbf{Y}_{t\cdot t-1}^{(f)}=\left(\mathbf{y}_t^{(2,f)},\mathbf{y}_{t-1}^{(1,f)}\right)$, $M_t=\mathbf{Z}_t\left(\mathbf{Z}_t'\mathbf{Z}_t\right)^{-1}\mathbf{Z}_t'$. Two versions of $\mathbf{Z}_t$ are suggested. Version A defines $\mathbf{Z}_t=\left(\mathbf{y}_{10},\mathbf{y}_{20},\ldots,\mathbf{y}_{1,t-1},\mathbf{y}_{2,t-1}\right)$ is the $N\times 2t$ instrumental variables matrix. Version B only uses the first lagged variables as instruments, so $\mathbf{Z}_t=\left(\mathbf{y}_{1,t-1},\mathbf{y}_{2,t-1}\right)$ (Akashi and Kunitomo (2014)).

Proposition 5.5 *For the least variance ratio estimator (5.13), under version A, Akashi and Kunitomo (2012) show that they are consistent and asymptotically normally distributed, but asymptotically biased of order* $\sqrt{\frac{T}{N}}$ *if* $0\leq c\leq\frac{1}{2}$ *where* $c=\frac{T}{N}$ *as* $(N,T)\to\infty$. *On the other hand, if*

[6] The Anderson-Rubin (1949) limited information maximum likelihood approach and the least variance ration approach yields identical estimators in the one-dimensional case (i.e., either $N=1$ but $T\neq 1$ or $T=1$ but $N\neq 1$). In the multi-dimensional case, the two approaches yields different estimator, Akashi and Kunitomo (2012a,b) (or Alvarez and Arellano (2003)) call their estimators the panel generalization of Anderson-Rubin (1949) LIML. We feel that their estimators is probably more in the spirit of least variance ratio approach.

only a fixed number of instruments are used, Akashi and Kunitomo (2014) show that the version B of the least variance ratio estimator is asymptotically unbiased.

6 Simulation[7]

We conduct a small scale Monte Carlo simulations to examine the finite sample properties of various estimators and report the results in this section. Following Akashi and Kunitomo (2012), we consider a dynamic simultaneous equations model of the form

$$\begin{aligned} y_{1,it} &= \beta y_{2,it} + \gamma_{11} y_{1,it-1} + \eta_{1i} + u_{1,it}, \\ y_{2,it} &= \gamma_{21} y_{1,it-1} + \gamma_{22} y_{2,it-1} + \eta_{2i} + u_{2,it}, \end{aligned}$$

with $\beta = 0.5$, $\gamma_{11} = 0.5$, $\gamma_{21} = 0$, $\gamma_{22} = 0.3$.

In data generation process 1 (DGP1), we assume that

$$\begin{aligned} \begin{pmatrix} \eta_{1i} \\ \eta_{2i} \end{pmatrix} &\sim {}_{iid} N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \\ \begin{pmatrix} u_{1,it} \\ u_{2,it} \end{pmatrix} &\sim {}_{iid} N\left(0, \begin{bmatrix} \sigma_{u_1}^2 & 0.2\sigma_{u_1}\sigma_{u_2} \\ 0.2\sigma_{u_1}\sigma_{u_2} & \sigma_{u_2}^2 \end{bmatrix}\right), \end{aligned}$$

where $\sigma_{u_1,i}^2$ and $\sigma_{u_2,i}^2$ are set as independently random draws from $0.5(1 + 0.5\chi^2(2))$ for $i = 1, 2, \ldots, N$, and $(\eta_{1i}, \eta_{2i})'$ and $(u_{1,it}, u_{2,it})'$ are independent over i and t.

In DGP2, we assume that

$$\begin{pmatrix} \eta_{1i} \\ \eta_{2i} \end{pmatrix} \sim_{iid} N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right),$$

and $u_{1,it} \sim_{iid} \chi^2(1) - 1$ and $u_{2,it} \sim_{iid} \chi^2(1) - 1$, but we set $Cov(u_{1,it}, u_{2,it}) = 0.2$. As before, $(\eta_{1i}, \eta_{2i})'$ and $(u_{1,it}, u_{2,it})'$ are independent over i and t.

We generate $100 + T$ observations of $\mathbf{y}_{it}$, starting with zero. We let $\mathbf{y}_{i0}$ be the 100th observations of $\mathbf{y}_{it}$. We report the bias, root mean square error (RMSE), iqr (interquantile range of 75%-25%) and size for MLE, PIV, PG2SLS, two versions of Akashi-Kunitomo least variance ratio estimator (LV$^{(A)}$ and LV$^{(B)}$), two versions of Akashi-Kunitomo type GMM (GMM$^{(A)}$ and GMM$^{(B)}$) and the MLE assuming initial value as fixed, denoted by MLE*, of β and γ_{11} when $N = 100, 200$ and $T = 25, 50$ for DGP1 in Tables 1-2 and DGP2 in Tables 3-4. The number of replication is set at 2000. For illustration, we also draw the empirical densities for different estimators of β and γ_{11} when $(N, T) = (200, 25)$ for DGP1.

[7] Codes are available from the authors upon request.

The results show clearly that the actual size of PIV, PG2SLS and PLIML is close to the nominal size. However, the size distortion of Arellano-Bond type GMM and version A of LV are significant. The size distortion of using a fixed number of instruments for the Arellano-Bond type GMM (GMM$^{(B)}$) is less, however, it is still significant. The version B of Akashi-Kunitomo LV (LV$^{(B)}$) has negligible size distortion if both N and T are large. However, if N is much larger than T, then the size distortion remains significant. The actual size of the GMM estimator for the coefficient of the joint dependent variable, β, could be near 100%, and for the lag dependent variable, γ, could be near 75% for a 5% significance level test. On the other hand, the actual size of the LV$^{(A)}$ estimator is about 1% for a 5% significant level test due to the large variance of the estimator. Overall, the empirical distribution of Arellano-Bond type GMM estimators for β is more concentrated than other estimators. Unfortunately, it is not centered at the true values. Among the asymptotically unbiased estimators, the root mean square error of the PLIML is only about one-half of the root mean square error of the PIV, PG2SLS or LV$^{(B)}$. Overall, our finding suggests that PLIML proposed in this paper is preferred for the estimation and inference for panel dynamic simultaneous equations models, in terms of bias, RMSE, and size.

7 Conclusion

We consider the identification and estimation of panel dynamic simultaneous equations models in this paper. We have shown that although the time-invariant individual-specific effects creates the dependence between the current and all the past joint dependent variables, they do not change the identification conditions for the Cowles Commission dynamic simultaneous equations models (e.g. Hood and Koopmans (1953), Hsiao (1983)). However, the presence of time-invariant individual-specific effects does affect the asymptotic properties of the estimator when the cross-sectional dimension, N, and the time-series dimension, T, are of the same magnitude. We consider both the likelihood approach and the methods of moments approach of inference. We show that the treatment of initial values plays a pivotal role in the likelihood approach. The asymptotic distribution of the quasi-maximum likelihood estimator (QMLE) is centered at the true value independent of the way N or T or both go to infinity if the distribution of initial values is properly formulated. On the other hand, mistreating initial values as fixed constant could yield an estimator that is asymptotically biased of order $\sqrt{\frac{N}{T}}$. For the method of moments estimators, the treatment of initial values plays no role. However, the asymptotic distribution depends critically on the way that population moments are approximated by the sample moments. The suggested panel instrumental variable estimator

(PIV) and panel generalized two stage least squares estimator (PG2SLS) both approximate the population moments by taking the average of NT sample observations are consistent and asymptotically unbiased independent of the way N or T or both tend to infinity. On the other hand, the Arellano-Bond (1991) type GMM estimators approximate the population moments by taking the cross-sectional mean is asymptotically unbiased only if T is fixed and N tends to infinity. When $\frac{T}{N} \to c \neq 0 < \infty$, the Arellano-Bond type GMM estimator is asymptotically biased of order $\sqrt{\frac{T}{N}}$. Our Monte Carlo studies confirm the importance of using asymptotically unbiased estimators to obtain valid statistical inference and the desirability of using a properly formulated likelihood approach for inference.[8]

References

[1] Akashi, K., and N. Kunitomo, 2012, Some properties of the LIML estimator in a dynamic panel structural equation, *Journal of Econometrics* 166, 167-183;

[2] Akashi, K., and N. Kunitomo, 2014, The limited information maximum likelihood approach to dynamic panel structural equations, forthcoming;

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

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

[5] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47–82;

[6] Anderson, T.W., and H. Rubin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63;

[7] Anderson, T.W., and H. Rubin, 1950, The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 21, 570–582;

[8] Arellano, M., 2003, *Panel Data Econometrics*, Cambridge University Press;

[8] Our focus in this paper is to derive the asymptotic distribution of an estimator. As a referee points out that an asymptotically normally distributed estimator may not have exact moments, which will be studied in future.

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

[10] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29–51;

[11] Avery, R., 1977, Error components and seemingly unrelated regressions, *Econometrica* 45, 199-209;

[12] Binder, M., C. Hsiao, and M.H. Pesaran, 2005, Estimation and inference in short panel vector autoregressions with unit roots and cointegration, *Econometric Theory* 21, 795-837;

[13] Box, G., and G. Jenkins, 1970, *Time Series Analysis: forcaseing and control*, Holden-Day Inc.;

[14] Dempster, A.P., N.M. Laird and D.B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B,* 39, 1-38;

[15] Grassetti, L., 2011, A note on transformed likelihood approach in linear dynamic panel models, *Statistical Methods & Applications* 20, 221-240;

[16] Hahn, J., J. Hausman, and G. Kuersteiner, 2007, Long difference instrumental variables estimation for dynamic panel models with fixed effects, *Journal of Econometrics* 140, 574–617;

[17] Hahn, J., and G. Kuersteiner, 2002, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* 70, 1639-1657;

[18] Hood, W., and T. Koopmans, 1953, *Studies in Eeconometric Method*, John Wiley & Sons;

[19] Hsiao, C., 1983, Identification, in *Handbook of Econometrics* vol 1, edited by Z. Griliches and M. Intriligator, pp. 223-283. Amsterdam;

[20] Hsiao, C., 2003, *Analysis of Panel Data*, Cambridge University Press;

[21] Hsiao, C., Pesaran, M.H., and K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods, *Jounal of Econometrics* 109, 107-150;

[22] Hsiao, C., and J. Zhang, 2013, IV, GMM or Likelihood approach to estimate dynamic panel models when either N or T or both are large, working paper;

[23] Maddala, G.S., 1971, The use of variance components models in pooling cross section and time series data, *Econometrica* 39, 341-358;

[24] Magnus, J., and H. Neucecker, 2007, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd Edition, John Wiley & Sons.

[25] Nerlove, M., 2000, An essay on the history of panel data econometrics, working paper;

[26] Shah, A., N. Laird and D. Schoenfeld, 1997, A random-effects model for multiple characteristics with possibly missing data, *Journal of the American Statistical Association* 92, 775-779;

[27] Wang, W-L., and T-H. Fan, 2010, ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors, *Computational Statistics & Data Analysis* 54, 1328-1341;

Table 1: Simulation results of β for DGP1

N	T		β_{MLE}	$\beta_{MLE,*}$	β_{IV}	β_{PG2SLS}	$\beta_{GMM}^{(A)}$	$\beta_{GMM}^{(B)}$	$\beta_{LV}^{(A)}$	$\beta_{LV}^{(B)}$
	25	estimate	0.4979	0.4920	0.4973	0.5064	0.6903	0.5852	0.4717	0.4976
		bias	-0.0021	-0.0080	-0.0027	0.0064	0.1903	0.0852	-0.0283	-0.0024
		rmse	0.0855	0.0875	0.1396	0.1268	0.1947	0.1320	1.6560	0.3709
		iqr	0.1176	0.1156	0.1738	0.1715	0.0494	0.1254	0.3310	0.2609
100		size	5.1%	5.25%	5.2%	5.05%	99%	13%	0.5%	2.9%
	50	estimate	0.5003	0.4990	0.4983	0.5020	0.6854	0.5864	0.4327	0.4929
		bias	0.0003	-0.0010	-0.0017	0.0020	0.1854	0.0864	-0.0673	-0.0071
		rmse	0.0590	0.0599	0.0904	0.0792	0.1864	0.1098	0.7791	0.1447
		iqr	0.0799	0.0814	0.1168	0.1063	0.0249	0.0892	0.3210	0.1820
		size	5.1%	4.75%	5.2%	5.15%	100%	24%	1%	5.1%
	25	estimate	0.5005	0.4951	0.4991	0.5033	0.6893	0.5510	0.4595	0.4826
		bias	0.0005	-0.0049	-0.0009	0.0033	0.1893	0.0510	-0.0405	-0.0174
		rmse	0.0608	0.0624	0.0951	0.0873	0.1936	0.0977	0.1619	0.1386
		iqr	0.0807	0.0826	0.1268	0.1110	0.0470	0.1127	0.1925	0.1741
200		size	4.9%	5.25%	4.95%	5%	99%	9.5%	5.95%	4.9%
	50	estimate	0.5005	0.4987	0.4987	0.5023	0.6868	0.5555	0.4480	0.4935
		bias	0.0005	-0.0013	-0.0013	0.0023	0.1868	0.0555	-0.0520	-0.0065
		rmse	0.0398	0.0404	0.0643	0.0570	0.1879	0.0795	0.3580	0.0863
		iqr	0.0503	0.0539	0.0861	0.0741	0.0240	0.0752	0.1721	0.1114
		size	4.9%	4.75%	5.4%	5.05%	100%	15%	0.1%	5.5%

Note: 1. The true value of β in this case is $\beta = 0.5$;

2. For estimators, $_{MLE}$ refers to MLE, $_{MLE,*}$ refers to MLE ignoring $\mathbf{y}_{i1} - \mathbf{y}_{i0}$, ^{IV} refers to IV estimation, ^{PG2SLS} refers to PG2SLS estimation, ^{GMM,A} refers to Akashi-Kunitomo type GMM estimator of version A, ^{GMM,B} refers to Akashi-Kunitomo type GMM estimator of version B, ^{LV,A} refers to least variance estimation of version A, ^{LV,B} refers to least variance estimation of version B;

3. iqr is the 75th-25th interquartile range;

4. The number of replication is set at $R = 2000$, and the 95% confidence interval for size 5% is [4%, 6%];

Table 2: Simulation results of γ_{11} for DGP1

N	T		γ_{11}^{MLE}	$\gamma_{11}^{MLE,*}$	γ_{11}^{IV}	γ_{11}^{PG2SLS}	$\gamma_{11}^{GMM,A}$	$\gamma_{11}^{GMM,B}$	$\gamma_{11}^{LV,A}$	$\gamma_{11}^{LV,B}$
	25	estimate	0.5000	0.5035	0.5004	0.4986	0.4322	0.4698	0.4717	0.4816
		bias	0.0000	0.0035	0.0004	-0.0014	-0.0678	-0.0302	-0.0283	-0.0184
		rmse	0.0214	0.0227	0.0418	0.0392	0.0735	0.0450	0.1641	0.0696
		iqr	0.0284	0.0304	0.0536	0.0518	0.0261	0.0442	0.0551	0.0571
100		size	4.7%	5.15%	5.95%	5.05%	66%	15%	0.5%	2.9%
	50	estimate	0.4998	0.5005	0.5002	0.4993	0.4502	0.4801	0.4895	0.4936
		bias	-0.0002	0.0005	0.0002	-0.0007	-0.0498	-0.0199	-0.0105	-0.0064
		rmse	0.0143	0.0146	0.0282	0.0240	0.0519	0.0298	0.0829	0.0302
		iqr	0.0196	0.0196	0.0381	0.0328	0.0165	0.0290	0.0475	0.0390
		size	5.05%	4.85%	4.55%	4.95%	93%	14%	1%	5.9%
	25	estimate	0.4998	0.5027	0.5008	0.5000	0.4478	0.4794	0.4861	0.4894
		bias	-0.0002	0.0027	0.0008	0.0000	-0.0522	-0.0206	-0.0139	-0.0106
		rmse	0.0152	0.0162	0.0294	0.0266	0.0570	0.0330	0.0292	0.0326
		iqr	0.0205	0.0210	0.0392	0.0368	0.0202	0.0361	0.0336	0.0404
200		size	5.05%	5.65%	5.35%	4.45%	63%	13%	7.4%	5.8%
	50	estimate	0.5002	0.5009	0.5008	0.4999	0.4592	0.4851	0.4960	0.4940
		bias	0.0002	0.0009	0.0008	-0.0001	-0.0408	-0.0149	-0.0040	-0.0060
		rmse	0.0096	0.0099	0.0195	0.0165	0.0426	0.0229	0.0426	0.0209
		iqr	0.0131	0.0131	0.0261	0.0220	0.0128	0.0208	0.0262	0.0268
		size	4.75%	5.15%	4.75%	5.55%	91%	15%	1%	5.5%

Note: 1. The true value of γ_{11} in this case is γ_{11}= 0.5;

2. For estimators, ^{MLE} refers to MLE, ^{MLE,*} refers to MLE ignoring $\mathbf{y}_{i1}-\mathbf{y}_{i0}$, ^{IV} refers to IV estimation, ^{PG2SLS} refers to G2SLS estimation, ^{GMM,A} refers to Akashi-Kunitomo type GMM estimator of version A, ^{GMM,B} refers to Akashi-Kunitomo type GMM estimator of version B, ^{LV,A} refers to least variance estimation of version A, ^{LV,B} refers to least variance estimation of version B;

3. iqr is the 75th-25th interquartile range;

4. The number of replication is set at $R = 2000$, and the 95% confidence interval for size 5% is [4%, 6%];

Table 3: Simulation results of β for DGP2

N	T		β_{MLE}	$\beta_{MLE,*}$	β_{IV}	β_{PG2SLS}	$\beta_{GMM}^{(A)}$	$\beta_{GMM}^{(B)}$	$\beta_{LV}^{(A)}$	$\beta_{LV}^{(B)}$
	25	estimate	0.4978	0.4918	0.4946	0.5033	0.7003	0.5740	0.3634	0.4924
		bias	-0.0022	-0.0082	-0.0054	0.0033	0.2003	0.0740	-0.1366	-0.0076
		rmse	0.0817	0.0829	0.1341	0.1157	0.2054	0.1184	3.0518	0.1793
		iqr	0.1073	0.1085	0.1756	0.1491	0.0543	0.1226	0.3310	0.2099
100		size	5.35%	5.1%	5.35%	5.2%	99%	13%	1%	5.9%
	50	estimate	0.4975	0.4962	0.4928	0.4964	0.6949	0.5761	0.4006	0.5005
		bias	-0.0025	-0.0038	-0.0074	-0.0036	0.1949	0.0761	-0.0994	0.0005
		rmse	0.0571	0.0575	0.0933	0.0782	0.1965	0.0994	15.550	0.1071
		iqr	0.0767	0.0779	0.1221	0.1037	0.0310	0.0899	0.3192	0.1402
		size	4.7%	4.85%	4.6%	5.3%	100%	23%	0.2%	5.7%
	25	estimate	0.4980	0.4922	0.4965	0.5015	0.6949	0.5427	0.4656	0.4900
		bias	-0.0020	-0.0078	-0.0035	0.0015	0.1949	0.0427	-0.0344	-0.0100
		rmse	0.0584	0.0595	0.0926	0.0809	0.1999	0.0834	0.1856	0.0999
		iqr	0.0772	0.0797	0.1224	0.1107	0.0472	0.1022	0.1675	0.1383
200		size	5.3%	5.15%	5.15%	4.55%	99%	9.6%	1.8%	5.2%
	50	estimate	0.4977	0.4962	0.4959	0.4966	0.6930	0.5432	0.4720	0.4938
		bias	-0.0023	-0.0038	-0.0041	-0.0034	0.1930	0.0432	-0.0280	-0.0062
		rmse	0.0408	0.0412	0.0640	0.0552	0.1944	0.0654	0.1247	0.0654
		iqr	0.0552	0.0557	0.0836	0.0758	0.0270	0.0604	0.1441	0.0809
		size	5.3%	5.3%	5.3%	5.3%	100%	14%	5.3%	5.6%

Note: 1. The true value of β in this case is $\beta = 0.5$;

2. For estimators, $_{MLE}$ refers to MLE, $_{MLE,*}$ refers to MLE ignoring $\mathbf{y}_{i1} - \mathbf{y}_{i0}$, ^{IV} refers to IV estimation, ^{PG2SLS} refers to G2SLS estimation, ^{GMM,A} refers to Akashi-Kunitomo type GMM estimator of version A, ^{GMM,B} refers to Akashi-Kunitomo type GMM estimator of version B, ^{LV,A} refers to least variance estimation of version A, ^{LV,B} refers to least variance estimation of version B;

3. iqr is the 75th-25th interquartile range;

4. The number of replication is set at $R = 2000$, and the 95% confidence interval for size 5% is [4%, 6%];

Table 4: Simulation results of γ_{11} for DGP2

N	T		γ_{11}^{MLE}	$\gamma_{11}^{MLE,*}$	γ_{11}^{IV}	γ_{11}^{PG2SLS}	$\gamma_{11}^{GMM,A}$	$\gamma_{11}^{GMM,B}$	$\gamma_{11}^{LV,A}$	$\gamma_{11}^{LV,B}$
	25	estimate	0.4995	0.5023	0.5007	0.4983	0.4333	0.4752	0.4896	0.4871
		bias	-0.0005	0.0023	0.0007	-0.0017	-0.0667	-0.0248	-0.0104	-0.0129
		rmse	0.0191	0.0197	0.0368	0.0321	0.0717	0.0368	0.3198	0.0380
		iqr	0.0256	0.0252	0.0494	0.0418	0.0234	0.0359	0.0541	0.0444
100		size	5.35%	5.6%	5.25%	5.75%	72%	15%	1%	6.2%
	50	estimate	0.4998	0.5004	0.5010	0.5003	0.4489	0.4816	0.4885	0.4922
		bias	-0.0002	0.0004	0.0010	0.0003	-0.0511	-0.0184	-0.0115	-0.0078
		rmse	0.0128	0.0129	0.0253	0.0214	0.0529	0.0259	2.1272	0.0235
		iqr	0.0165	0.0169	0.0326	0.0290	0.0150	0.0253	0.0477	0.0295
		size	5.6%	5.55%	5%	4.85%	96%	19%	0.2%	6.2%
	25	estimate	0.5003	0.5031	0.5009	0.4996	0.4500	0.4813	0.4868	0.4890
		bias	0.0003	0.0031	0.0009	-0.0004	-0.0500	-0.0187	-0.0132	-0.0110
		rmse	0.0136	0.0143	0.0258	0.0227	0.0543	0.0273	0.0288	0.0249
		iqr	0.0188	0.0188	0.0353	0.0303	0.0182	0.0247	0.0296	0.0288
200		size	5.1%	5.55%	4.45%	5.1%	66%	15%	5.2%	8.5%
	50	estimate	0.5001	0.5007	0.5005	0.5006	0.4584	0.4868	0.4933	0.4938
		bias	0.0001	0.0007	0.0005	0.0006	-0.0416	-0.0132	-0.0067	-0.0062
		rmse	0.0089	0.0090	0.0178	0.0151	0.0431	0.0187	0.0193	0.0160
		iqr	0.0119	0.0118	0.0236	0.0202	0.0112	0.0182	0.0235	0.0199
		size	5.35%	5.6%	5.9%	5.3%	96%	17%	5.95%	7.5%

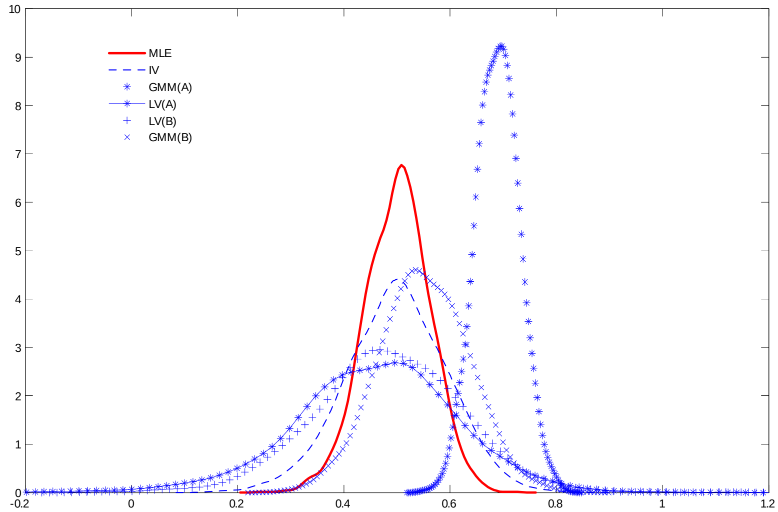
Note: 1. The true value of γ_{11} in this case is γ_{11}= 0.5;

2. For estimators, ^{MLE} refers to MLE, ^{MLE,*} refers to MLE ignoring $\mathbf{y}_{i1}-\mathbf{y}_{i0}$, ^{IV} refers to IV estimation, ^{PG2SLS} refers to G2SLS estimation, ^{GMM,A} refers to Akashi-Kunitomo type GMM estimator of version A, ^{GMM,B} refers to Akashi-Kunitomo type GMM estimator of version B, ^{LV,A} refers to least variance estimation of version A, ^{LV,B} refers to least variance estimation of version B;

3. iqr is the 75th-25th interquartile range;

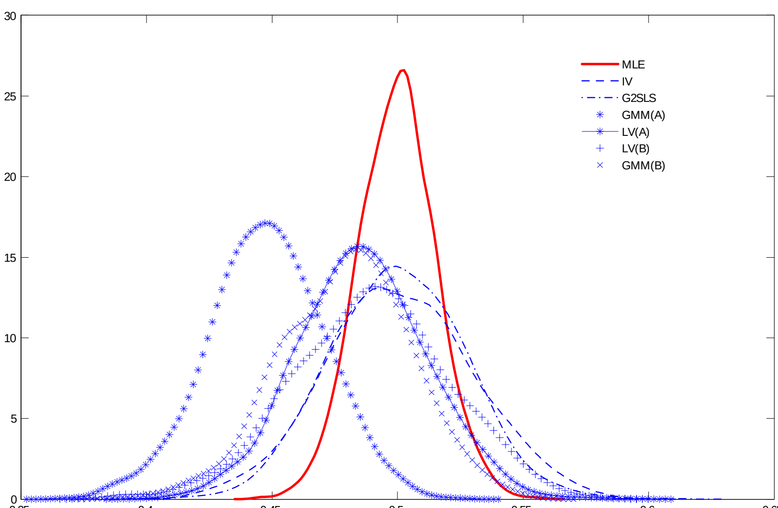
4. The number of replication is set at $R = 2000$, and the 95% confidence interval for size 5% is [4%, 6%];

Figure 1: Empirical densities for MLE, IV, 2SLS, GMM, MGMM and LV estimators of β for DGP1 when $(N, T) = (200, 25)$



These empirical densities are drawn based on 2000 replications of DGP1, the true value of β is 0.5.

Figure 2: Empirical densities for MLE, IV, 2SLS, GMM, MGMM and LV estimators of γ_{11} for DGP1 when $(N, T) = (200, 25)$



These empirical densities are drawn based on 2000 replications of DGP1, the true value of γ_{11} is 0.5.

Appendix

We sketch the derivation of the asymptotic distribution of the MLE, IV, G2SLS. Details are available on request.

Under the assumption that $\mathbf{x}_{it}$ are independent of $\mathbf{u}_{it}$, the presence of $\mathbf{x}_{it}$ does not affect the limiting distribution, nor the order of asymptotic bias, for ease of exposition, we consider an exactly identified model of the form

$$\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1,it} \\ y_{2,it} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1,it-1} \\ y_{2,it-1} \end{pmatrix} + \begin{pmatrix} \eta_{1i} \\ \eta_{2i} \end{pmatrix} = \begin{pmatrix} u_{1,it} \\ u_{2,it} \end{pmatrix} \tag{A.1}$$

where $\gamma_{22} \neq 0$ with the reduced form

$$\begin{pmatrix} y_{1,it} \\ y_{2,it} \end{pmatrix} = \Pi \begin{pmatrix} y_{1,it-1} \\ y_{2,it-1} \end{pmatrix} + \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} + \begin{pmatrix} v_{1,it} \\ v_{2,it} \end{pmatrix}, \tag{A.2}$$

where

$$\Pi = -\mathbf{B}^{-1}\Gamma = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}.$$

Let $\tilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \mathbf{y}_{i0}$, following the discussion in the paper, we have

$$\tilde{\mathbf{Y}}_i = I_2 \otimes \left(\tilde{Y}_{1i,-1}, \tilde{Y}_{2i,-1}\right) vec\left(\Pi'\right) + \tilde{\mathbf{V}}_i, \quad i = 1, 2, \ldots, N, \tag{A.3}$$

where $\tilde{\mathbf{Y}}_i = \left(\tilde{Y}'_{1i}, \tilde{Y}'_{2i}\right)'$ with $\tilde{Y}_{1i} = (\tilde{y}_{1,i2}, \ldots, \tilde{y}_{1,iT})'$ and $\tilde{Y}_{1i,-1} = (\tilde{y}_{1,i0}, \ldots, \tilde{y}_{1,iT-1})'$, and

$$\tilde{\mathbf{V}}_i = \mathbf{V}_i + \boldsymbol{\xi}_i \otimes 1_T,$$

and

$$\begin{aligned} \Omega_{\tilde{V}} &= E\left(\tilde{\mathbf{V}}_i \tilde{\mathbf{V}}'_i\right) = E\left(\mathbf{V}_i \mathbf{V}'_i\right) + E\left(\boldsymbol{\xi}_i \boldsymbol{\xi}'_i \otimes 1_T 1'_T\right) \\ &= \Omega_v \otimes I_T + \Omega_\xi \otimes 1_T 1'_T \\ &= \Omega_v \otimes Q + \Omega \otimes J. \end{aligned}$$

where $\Omega = \Omega_u + T\Omega_\xi$ and

$$\Omega_v = E\left(\mathbf{v}_{it}\mathbf{v}'_{it}\right) = \begin{pmatrix} \sigma_{v,11} & \sigma_{v,12} \\ \sigma_{v,21} & \sigma_{v,22} \end{pmatrix}, \Omega_\xi = E\left(\boldsymbol{\xi}_i \boldsymbol{\xi}'_i\right) = \begin{pmatrix} \sigma_{\xi,11} & \sigma_{\xi,12} \\ \sigma_{\xi,21} & \sigma_{\xi,22} \end{pmatrix}.$$

A. Limiting distribution of MLE

Conditional on Ω_v and Ω, the MLE of $vec\left(\Pi'\right) = \boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})'$ is given by

$$\hat{\boldsymbol{\pi}} = \left\{ \left[I_2 \otimes \begin{pmatrix} \tilde{Y}_{1,-1}' \\ \tilde{Y}_{2,-1}' \end{pmatrix} \right] \Omega_{\tilde{V}}^{-1} \left[I_2 \otimes \left(\tilde{Y}_{1,-1}, \tilde{Y}_{2,-1} \right) \right] \right\}^{-1} \times \left\{ \left[I_2 \otimes \begin{pmatrix} \tilde{Y}_{1,-1}' \\ \tilde{Y}_{2,-1}' \end{pmatrix} \right] \Omega_{\tilde{V}}^{-1} \tilde{Y} \right\}. \qquad \text{(A.4)}$$

where $\tilde{Y}_1 = \left(\tilde{Y}_{1,1}', \ldots, \tilde{Y}_{1,N}' \right)'$ with $\tilde{Y}_{1,i} = \left(\tilde{y}_{1,i1}, \ldots, \tilde{y}_{1,iT}\right)'$, $\tilde{Y}_{1,-1} = \left(\tilde{Y}_{11,-1}', \ldots, \tilde{Y}_{1N,-1}' \right)'$ with $\tilde{Y}_{1i,-1} = (0, \tilde{y}_{1,i2}, \ldots, \tilde{y}_{1,iT-1})'$, and $\Omega_{\tilde{V}}^{-1} = \Omega_v^{-1} \otimes Q + \Omega^{-1} \otimes J$. Thus,

$$\begin{aligned} \sqrt{NT}\left(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\right) &= \left\{ \frac{1}{NT} \left[I_2 \otimes \begin{pmatrix} \tilde{Y}_{1,-1}' \\ \tilde{Y}_{2,-1}' \end{pmatrix} \right] \Omega_{\tilde{V}}^{-1} \left[I_2 \otimes \left(\tilde{Y}_{1,-1}, \tilde{Y}_{2,-1} \right) \right] \right\}^{-1} \\ &\quad \times \left\{ \frac{1}{\sqrt{NT}} \left[I_2 \otimes \begin{pmatrix} \tilde{Y}_{1,-1}' \\ \tilde{Y}_{2,-1}' \end{pmatrix} \right] \Omega_{\tilde{V}}^{-1} \tilde{\mathbf{V}} \right\}, \qquad \text{(A.5)} \end{aligned}$$

It's easy to see the first term on the righthand size of (A.5) converges to a nonsingular constant matrix as $(N, T) \to \infty$. Let

$$\Omega_v^{-1} = \begin{pmatrix} \sigma_v^{11} & \sigma_v^{12} \\ \sigma_v^{12} & \sigma_v^{22} \end{pmatrix}, \Omega^{-1} = \begin{pmatrix} w^{11} & w^{12} \\ w^{12} & w^{22} \end{pmatrix}.$$

Then, the numerator of (A.5) can be rewritten as $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \vartheta_i$ where $\vartheta_i = (\vartheta_{i,1}, \vartheta_{i,2}, \vartheta_{i,3}, \vartheta_{i,4})'$ with

$$\begin{aligned} \vartheta_{i,1} &= \sigma_v^{11} \tilde{Y}_{1i,-1}' Q \mathbf{V}_{1i} + \sigma_v^{12} \tilde{Y}_{1i,-1}' Q \mathbf{V}_{2i} + w^{11} \tilde{Y}_{1i,-1}' J \left(\mathbf{1}_T \xi_{1i} + \mathbf{V}_{1i} \right) + w^{12} \tilde{Y}_{1i,-1}' J \left(\mathbf{1}_T \xi_{2i} + \mathbf{V}_{2i} \right) \\ \vartheta_{i,2} &= \sigma_v^{11} \tilde{Y}_{2i,-1}' Q \mathbf{V}_{1i} + \sigma_v^{12} \tilde{Y}_{2i,-1}' Q \mathbf{V}_{2i} + w^{11} \tilde{Y}_{2i,-1}' J \left(\mathbf{1}_T \xi_{1i} + \mathbf{V}_{1i} \right) + w^{12} \tilde{Y}_{2i,-1}' J \left(\mathbf{1}_T \xi_{2i} + \mathbf{V}_{2i} \right) \\ \vartheta_{i,3} &= \sigma_v^{21} \tilde{Y}_{1i,-1}' Q \mathbf{V}_{1i} + \sigma_v^{22} \tilde{Y}_{1i,-1}' Q \mathbf{V}_{2i} + w^{21} \tilde{Y}_{1i,-1}' J \left(\mathbf{1}_T \xi_{1i} + \mathbf{V}_{1i} \right) + w^{22} \tilde{Y}_{1i,-1}' J \left(\mathbf{1}_T \xi_{2i} + \mathbf{V}_{2i} \right) \\ \vartheta_{i,4} &= \sigma_v^{21} \tilde{Y}_{2i,-1}' Q \mathbf{V}_{1i} + \sigma_v^{22} \tilde{Y}_{2i,-1}' Q \mathbf{V}_{2i} + w^{21} \tilde{Y}_{2i,-1}' J \left(\mathbf{1}_T \xi_{1i} + \mathbf{V}_{1i} \right) + w^{22} \tilde{Y}_{2i,-1}' J \left(\mathbf{1}_T \xi_{2i} + \mathbf{V}_{2i} \right), \end{aligned} \qquad \text{(A.6)}$$

We note that

$$\begin{aligned} \begin{pmatrix} \tilde{y}_{1,it} \\ \tilde{y}_{2,it} \end{pmatrix} &= \Pi \begin{pmatrix} \tilde{y}_{1i,t-1} \\ \tilde{y}_{2i,t-1} \end{pmatrix} + \boldsymbol{\xi}_i + \mathbf{v}_{it} \qquad \text{(A.7)} \\ &= \left(I_2 - \Pi\right)^{-1} \left(I_2 - \Pi^t\right) \boldsymbol{\xi}_i + \sum_{j=0}^{t} \Pi^j \mathbf{v}_{i,t-j}, \end{aligned}$$

then

$$\begin{aligned} E\left[\begin{pmatrix} \tilde{y}_{1,it-1} \\ \tilde{y}_{2,it-1} \end{pmatrix} \boldsymbol{\xi}_i' \right] &= \left(I_2 - \Pi\right)^{-1} \left(I_2 - \Pi^{t-1}\right) \begin{pmatrix} \sigma_{\xi,11} & \sigma_{\xi,12} \\ \sigma_{\xi,12} & \sigma_{\xi,22} \end{pmatrix}, \\ E\left[\begin{pmatrix} \tilde{y}_{1,it-1} \\ \tilde{y}_{2,it-1} \end{pmatrix} \mathbf{v}_{i,t-j}' \right] &= \Pi^j \begin{pmatrix} \sigma_{v,11} & \sigma_{v,12} \\ \sigma_{v,12} & \sigma_{v,22} \end{pmatrix} \quad \text{for } 1 \le j \le t-1, \end{aligned}$$

and

$$E\left\{\begin{pmatrix}\tilde{Y}'_{1i,-1}\\ \tilde{Y}'_{2i,-1}\end{pmatrix}J\mathbf{1}_T\left(\xi_{1i},\xi_{2i}\right)\right\}=E\left\{\sum_{j=0}^{T-1}\begin{pmatrix}\tilde{y}_{1,iT-1-j}\\ \tilde{y}_{2,iT-1-j}\end{pmatrix}\left(\xi_{1i},\xi_{2i}\right)\right\}$$

$$= \left(I_2-\Pi\right)^{-2}\left\{\left(T-1\right)I_2-T\Pi+\Pi^T\right\}\begin{pmatrix}\sigma_{\xi,11} & \sigma_{\xi,12}\\ \sigma_{\xi,12} & \sigma_{\xi,22}\end{pmatrix}.$$

Thus

$$\begin{aligned}E\left\{\begin{pmatrix}\tilde{Y}'_{1i,-1}\\ \tilde{Y}'_{2i,-1}\end{pmatrix}Q\left(V_{1i},V_{2i}\right)\right\} &= -\frac{1}{T}E\left\{\sum_{j=0}^{T-1}\begin{pmatrix}\tilde{y}_{1,iT-1-j}\\ \tilde{y}_{2,iT-1-j}\end{pmatrix}\sum_{j=0}^{T-1}\left(v_{1,iT-j},v_{2,iT-j}\right)\right\}\\ &= -\frac{1}{T}\left(I_2-\Pi\right)^{-2}\left\{\left(T-1\right)I_2-T\Pi+\Pi^T\right\}\begin{pmatrix}\sigma_{v,11} & \sigma_{v,12}\\ \sigma_{v,12} & \sigma_{v,22}\end{pmatrix},\end{aligned}$$

and

$$\begin{aligned}E\left\{\begin{pmatrix}\tilde{Y}'_{1i,-1}\\ \tilde{Y}'_{2i,-1}\end{pmatrix}J\left(V_{1i},V_{2i}\right)\right\} &= \frac{1}{T}E\left\{\sum_{j=0}^{T-1}\begin{pmatrix}\tilde{y}_{1,iT-1-j}\\ \tilde{y}_{2,iT-1-j}\end{pmatrix}\sum_{j=0}^{T-1}\left(v_{1,iT-j},v_{2,iT-j}\right)\right\}\\ &= \frac{1}{T}\left(I_2-\Pi\right)^{-2}\left\{\left(T-1\right)I_2-T\Pi+\Pi^T\right\}\begin{pmatrix}\sigma_{v,11} & \sigma_{v,12}\\ \sigma_{v,12} & \sigma_{v,22}\end{pmatrix}.\end{aligned}$$

Let

$$\left(I_2-\Pi\right)^{-2}\left\{\left(T-1\right)I_2-T\Pi+\Pi^T\right\}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}.$$

Then

$$\begin{aligned}&E\left(\sigma_v^{11}\tilde{Y}'_{1i,-1}Q\mathbf{V}_{1i}+\sigma_v^{12}\tilde{Y}'_{1i,-1}Q\mathbf{V}_{2i}\right)\\ =\ & -\frac{1}{T}\left\{a_{11}\left[\sigma_v^{11}\sigma_{v,11}+\sigma_v^{12}\sigma_{v,21}\right]+a_{12}\left[\sigma_v^{11}\sigma_{v,12}+\sigma_v^{12}\sigma_{v,22}\right]\right\}\\ =\ & -\frac{a_{11}}{T},\end{aligned}$$

since $\sigma_v^{11}\sigma_{v,11}+\sigma_v^{12}\sigma_{v,21}=1$ and $\sigma_v^{11}\sigma_{v,12}+\sigma_v^{12}\sigma_{v,22}=0$ from the fact that $\Omega_v\Omega_v^{-1}=I_2$. Also,

$$\begin{aligned}&E\left(w^{11}\tilde{Y}'_{1i,-1}J\mathbf{1}_T\xi_{1i}+w^{12}\tilde{Y}'_{1i,-1}J\mathbf{1}_T\xi_{2i}\right)\\ =\ & a_{11}\left[w^{11}\sigma_{\xi,11}+w^{12}\sigma_{\xi,21}\right]+Ta_{12}\left[w^{11}\sigma_{\xi,12}+w^{12}\sigma_{\xi,22}\right],\end{aligned}$$

and

$$\begin{aligned}&E\left(w^{11}\tilde{Y}'_{1i,-1}J\mathbf{V}_{1i}+w^{12}\tilde{Y}'_{1i,-1}J\mathbf{V}_{2i}\right)\\ =\ & \frac{1}{T}\left\{a_{11}\left[w_{11}\sigma_{v,11}+w_{12}\sigma_{v,12}\right]+a_{12}\left[w_{11}\sigma_{v,21}+w_{12}\sigma_{v,22}\right]\right\}.\end{aligned}$$

Combining these two equations we have

$$\begin{aligned}
& E\left(w^{11}\tilde{Y}_{1i,-1}'J\left(\mathbf{1}_T\xi_{1i}+\mathbf{V}_{1i}\right)+w^{12}\tilde{Y}_{1i,-1}'J\left(\mathbf{1}_T\xi_{2i}+\mathbf{V}_{2i}\right)\right) \\
= & \frac{a_{11}}{T}\left[w^{11}\left(\sigma_{v,11}+T\sigma_{\xi,11}\right)+w^{12}\left(\sigma_{v,12}+T\sigma_{\xi,21}\right)\right] \\
& +\frac{a_{12}}{T}\left[w^{11}\left(\sigma_{v,12}+T\sigma_{\xi,12}\right)+w^{12}\left(\sigma_{v,22}+T\sigma_{\xi,22}\right)\right] \\
= & \frac{a_{11}}{T},
\end{aligned}$$

Thus

$$E\left[\sigma_v^{11}\tilde{Y}_{1i,-1}'Q\mathbf{V}_{1i}+\sigma_v^{12}\tilde{Y}_{1i,-1}'Q\mathbf{V}_{2i}+w^{11}\tilde{Y}_{1i,-1}'J\left(\mathbf{1}_T\xi_{1i}+\mathbf{V}_{1i}\right)+w^{12}\tilde{Y}_{1i,-1}'J\left(\mathbf{1}_T\xi_{2i}+\mathbf{V}_{2i}\right)\right]=0,$$

or

$$E\left(\vartheta_{i,1}\right)=0.$$

Similarly, we can show that $\vartheta_{i,2},\vartheta_{i,3},\vartheta_{i,4}$ have zero mean.

Following Magnus and Neudecker (2007, Ch16), we can establish that

$$\sqrt{NT}\left(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}\right)\xrightarrow{d}N\left(0,\Omega_{\boldsymbol{\pi}}\right),$$

where

$$\Omega_{\boldsymbol{\pi}}=-E\left(\frac{1}{NT}\frac{\partial^2\log L}{\partial\boldsymbol{\pi}\partial\boldsymbol{\pi}'}\right),$$

and

$$\begin{aligned}
\log L = & -\frac{NT}{2}\log\left|\Omega_{\tilde{V}}\right| \\
& -\frac{1}{2}\sum_{i=1}^{N}\left\{\left[\tilde{\mathbf{Y}}_i-I_2\otimes\left(\tilde{Y}_{1i,-1},\tilde{Y}_{2i,-1}\right)\boldsymbol{\pi}\right]\Omega_{\tilde{V}}^{-1}\left[\tilde{\mathbf{Y}}_i-I_2\otimes\left(\tilde{Y}_{1i,-1},\tilde{Y}_{2i,-1}\right)\boldsymbol{\pi}\right]'\right\}.
\end{aligned}$$

The structural form parameter β can be derived from the relation $\beta=\frac{\pi_{12}}{\pi_{22}}$. Thus, the MLE of β is simply $\frac{\hat{\pi}_{12}}{\hat{\pi}_{22}}$. From

$$\begin{aligned}
\hat{\beta}-\beta & = \frac{\hat{\pi}_{12}}{\hat{\pi}_{22}}-\frac{\pi_{12}}{\pi_{22}} \\
& = \frac{\pi_{22}\left(\hat{\pi}_{12}-\pi_{12}\right)-\pi_{12}\left(\hat{\pi}_{22}-\pi_{22}\right)}{\hat{\pi}_{22}\pi_{22}},
\end{aligned}\tag{A.8}$$

using the delta method, once can show that

$$\sqrt{NT}\left(\hat{\beta}-\beta\right)=\frac{\pi_{22}\sqrt{NT}\left(\hat{\pi}_{12}-\pi_{12}\right)-\pi_{12}\sqrt{NT}\left(\hat{\pi}_{22}-\pi_{22}\right)}{\pi_{22}^2}+o_p\left(1\right).\tag{A.9}$$

Thus, $\sqrt{NT}\left(\hat{\beta}-\beta\right)$ is asymptotically normally distributed with mean 0.

B. QMLE when $\mathbf{y}_{i1}-\mathbf{y}_{i0}$ is treated as fixed constant

Consider the system (A.1), let $\tilde{\mathbf{y}}_{i,-1}=(\tilde{\mathbf{y}}_{1i,-1},\tilde{\mathbf{y}}_{2i,-1})$, $\boldsymbol{\gamma}_2=(\gamma_{21},\gamma_{22})'$ and

$$\tilde{\mathbf{y}}_{1i}=\begin{pmatrix} y_{1,i2}-y_{1,i1} \\ \vdots \\ y_{1,iT}-y_{1,i1}\end{pmatrix}, \tilde{\mathbf{y}}_{1i,-1}=\begin{pmatrix} y_{1,i1}-y_{1,i0} \\ \vdots \\ y_{1,i,T-1}-y_{1,i0}\end{pmatrix}.$$

Let $\mathbf{Z}_i=(\tilde{\mathbf{y}}_{2i},\tilde{\mathbf{y}}_{1i,-1})$ and $\mathbf{W}_i=\left(\tilde{\mathbf{Y}}_{i,-1}\right)=(\tilde{\mathbf{y}}_{1i,-1},\tilde{\mathbf{y}}_{2i,-1})$, then

$$\begin{pmatrix}\tilde{\mathbf{y}}_{1i}\\ \tilde{\mathbf{y}}_{2i}\end{pmatrix}=\begin{pmatrix}\mathbf{Z}_i & 0\\ 0 & \mathbf{W}_i\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\gamma}_2\end{pmatrix}+\mathbf{V}_i,$$

where $\boldsymbol{\delta}_1=(\beta,\gamma_{11})'$ and $\mathbf{V}_i=(\mathbf{v}_{1i}',\mathbf{v}_{2i}')'$ with

$$\begin{aligned}\Omega_V &= E\left(\mathbf{V}_i\mathbf{V}_i'\right)=\Omega_u\otimes I_{T-1}+\Omega_u\otimes 1_{T-1}1_{T-1}' \\ &= \Omega_u\otimes\left(I_{T-1}+1_{T-1}1_{T-1}'\right),\end{aligned}$$

It follows that (e.g. Hsiao (2003)) $\Omega_V^{-1}=\Omega_u^{-1}\otimes\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)$, and

$$\begin{pmatrix}\hat{\boldsymbol{\delta}}_1\\ \hat{\boldsymbol{\gamma}}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\gamma}_2\end{pmatrix}+\left\{\sum_{i=1}^N\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\Omega_V^{-1}\begin{pmatrix}\mathbf{Z}_i & 0\\ 0 & \mathbf{W}_i\end{pmatrix}\right\}^{-1}\left\{\sum_{i=1}^N\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\Omega_V^{-1}\begin{pmatrix}\mathbf{v}_{1i}\\ \mathbf{v}_{2i}\end{pmatrix}\right\}, \tag{A.10}$$

It is easy to show that

$$\frac{1}{NT}\sum_{i=1}^N\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\Omega_V^{-1}\begin{pmatrix}\mathbf{Z}_i & 0\\ 0 & \mathbf{W}_i\end{pmatrix}\rightarrow_p\Sigma_1$$

where Σ_1 is a positive definite matrix. The numerator of (A.10) is

$$\begin{aligned}&\sum_{i=1}^N\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\Omega_V^{-1}\begin{pmatrix}\mathbf{v}_{1i}\\ \mathbf{v}_{2i}\end{pmatrix}\\ =&\sum_{i=1}^N\begin{pmatrix}\tilde{\mathbf{y}}_{2i}'\left(\sigma_u^{11}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{1i}+\sigma_u^{12}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{2i}\right)\\ \tilde{\mathbf{y}}_{1i,-1}'\left(\sigma_u^{11}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{1i}+\sigma_u^{12}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{2i}\right)\\ \mathbf{y}_{1i,-1}\left(\sigma_u^{21}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{1i}+\sigma_u^{22}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{2i}\right)\\ \mathbf{y}_{2i,-1}\left(\sigma_u^{21}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{1i}+\sigma_u^{22}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{2i}\right)\end{pmatrix},\end{aligned} \tag{A.11}$$

Thus, for the expectation of the first element of (A.11), we have

$$\begin{aligned}&E\left(\tilde{\mathbf{y}}_{2i}'\left(\sigma_u^{11}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{1i}+\sigma_u^{12}\left(I_{T-1}-\frac{1}{T}1_{T-1}1_{T-1}'\right)\mathbf{v}_{2i}\right)\right)\\ =&\ \sigma_u^{11}\left[E\left(\tilde{\mathbf{y}}_{2i}'\mathbf{v}_{1i}\right)-\frac{1}{T}E\left(\tilde{\mathbf{y}}_{2i}'1_{T-1}1_{T-1}'\mathbf{v}_{1i}\right)\right]+\sigma_u^{12}\left[E\left(\tilde{\mathbf{y}}_{2i}'\mathbf{v}_{2i}\right)-\frac{1}{T}E\left(\tilde{\mathbf{y}}_{2i}'1_{T-1}1_{T-1}'\mathbf{v}_{2i}\right)\right],\end{aligned}$$

To show this, we note that

$$\underset{(T-1)\times 1}{\tilde{\mathbf{y}}_{2i}} = \tilde{\mathbf{y}}_{1i,-1}\gamma_{21} + \tilde{\mathbf{y}}_{2i,-1}\gamma_{22} + \mathbf{v}_{2i},$$

then

$$E\left(\tilde{\mathbf{y}}_{2i}'\mathbf{v}_{1i}\right) = E\left(\gamma_{21}\mathbf{y}_{1i,-1}'\mathbf{v}_{1i} + \gamma_{22}\mathbf{y}_{2i,-1}'\mathbf{v}_{1i} + \mathbf{v}_{2i}'\mathbf{v}_{1i}\right),$$

because $y_{1,i0}$ is treated as a fixed constant. Also, since

$$\mathbf{y}_{it} = \left(I_2 - \Pi\right)^{-1}\left(I_2 - \Pi^t\right)\boldsymbol{\eta}_i + \sum_{j=0}^{t-1}\Pi^j\mathbf{B}^{-1}\mathbf{u}_{i,t-j} + \Pi^t\mathbf{y}_{i,0},$$

and $(y_{1,i0}, y_{2,i0})$ are fixed constants, then $E\left(\mathbf{y}_{it}'\mathbf{u}_{i,1}\right) = \Pi^{t-1}\mathbf{B}^{-1}\Omega_u$, and

$$\sum_{t=1}^{T-1}E\left(\mathbf{y}_{it}'\mathbf{u}_{i,1}\right) = \sum_{t=1}^{T-1}\Pi^{t-1}\mathbf{B}^{-1}\Omega_u = \left(I_2 - \Pi\right)^{-1}\left(I_2 - \Pi^T\right)\mathbf{B}^{-1}\Omega_u = O_p\left(1\right),$$

as $T \to \infty$. This suggests that

$$E\left(\mathbf{y}_{1i,-1}'\mathbf{v}_{1i}\right) = O_p\left(1\right).$$

Similarly, it can be shown that

$$E\left(\mathbf{y}_{1i,-1}'\mathbf{v}_{2i}\right) = O_p\left(1\right), E\left(\mathbf{y}_{2i,-1}'\mathbf{v}_{1i}\right) = O_p\left(1\right), E\left(\mathbf{y}_{2i,-1}'\mathbf{v}_{2i}\right) = O_p\left(1\right).$$

Also,

$$\begin{aligned}\frac{1}{T}E\left(\tilde{\mathbf{y}}_{2i}'1_{T-1}1_{T-1}'\mathbf{v}_{2i}\right) &= \frac{1}{T}\sum_{s,t}E\left(y_{2,is}v_{2,it}\right) - \frac{T-1}{T}E\left(y_{2,i1}1_{T-1}'\mathbf{v}_{2i}\right)\\ &= O_p\left(1\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{T}\sum_{t>s}E\left(\mathbf{y}_{it}\mathbf{v}_{is}'\right) &= \frac{1}{T}\sum_{t>s}\left(\Pi^{t-s}\mathbf{B}^{-1} - \Pi^{t-1}\mathbf{B}^{-1}\right)\mathbf{B}^{-1}\Omega_u\\ &= O_p\left(1\right),\end{aligned}$$

as $T \to \infty$. Similarly, we have $\frac{1}{T}E\left(\tilde{\mathbf{y}}_{2i}'1_{T-1}1_{T-1}'\mathbf{v}_{1i}\right) = O_p\left(1\right)$, $\frac{1}{T}E\left(\tilde{\mathbf{y}}_{1i}'1_{T-1}1_{T-1}'\mathbf{v}_{1i}\right) = O_p\left(1\right)$, $\frac{1}{T}E\left(\tilde{\mathbf{y}}_{1i}'1_{T-1}1_{T-1}'\mathbf{v}_{2i}\right) = O_p\left(1\right)$.

Combining these results, we have

$$\begin{pmatrix}\mathbf{Z}_i' & 0\\ 0 & \mathbf{W}_i'\end{pmatrix}\Omega_V^{-1}\begin{pmatrix}\mathbf{v}_{1i}\\ \mathbf{v}_{2i}\end{pmatrix} = O_p\left(1\right),$$

as $T \to \infty$. Consequently, we have

$$E\left[\sqrt{NT}\left(\begin{pmatrix}\hat{\boldsymbol{\delta}}_1 \\ \hat{\gamma}_2\end{pmatrix} - \begin{pmatrix}\boldsymbol{\delta}_1 \\ \gamma_2\end{pmatrix}\right)\right] = \frac{1}{\sqrt{NT}} O_p\left(N\right) = O_p\left(\sqrt{\frac{N}{T}}\right),$$

as required. This means that when the initial values are treated as fixed constant, the MLE are asymptotically biased of order $\sqrt{\frac{N}{T}}$.

C. Limiting distribution of IV estimator

For the PIV estimator (5.4) (or (5.5)), by using the orthogonal condition (5.3) and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\begin{pmatrix}\Delta\mathbf{y}_{2,it} \\ \Delta y_{1,it-1}\end{pmatrix}\mathbf{y}'_{i,t-2}$ (or $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=3}^{T}\begin{pmatrix}\Delta\mathbf{y}_{2,it} \\ \Delta y_{1,it-1}\end{pmatrix}\Delta\mathbf{y}'_{i,t-2}$) converges to a constant matrix as either N or T or both tend to infinity, it is obvious that (5.4) (or (5.5)) is consistent and asymptotically unbiased independent of the way N or T or both tend to infinity. For the limiting distribution, by following the standard textbook such as Hsiao (2003) or Hahn and Kuersteiner (2002), it can be easily verified that

$$\sqrt{NT}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta} \\ \hat{\gamma}_{1,IV} - \gamma\end{pmatrix} \xrightarrow{d} N\left(0, \Omega_{IV}\right),$$

where $\Omega_{IV} = \Xi_1^{-1}\Omega_1\Xi_1^{-1}$ for (5.4) or $\Omega_{IV} = \Xi_2^{-1}\Omega_2\Xi_2^{-1}$ for (5.5), where Ω_1, Ξ_1, Ω_2 and Ξ_2 are given in the paper.

D. Limiting distribution of PG2SLS estimator

For the PG2SLS estimator (5.9), for ease of exposition, we shall assume there is only one endogenous variables in (5.6), i.e, $\boldsymbol{\beta}$ is a scalar. Extension to more than one endogenous variables is straightforward. We first notice that

$$\begin{aligned}
\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{PG2SLS} - \boldsymbol{\theta}\right) &= \left\{\frac{1}{NT}\left[\sum_{i=1}^{N}\Delta\mathbf{X}'_i\mathbf{Y}'_{i,-2}\right]\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}'_{i,-2}\right]^{-1}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{X}_i\right\}^{-1} \\
&\times\left\{\left[\sum_{i=1}^{N}\Delta\mathbf{X}'_i\mathbf{Y}'_{i,-2}\right]\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}'_{i,-2}\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\right]\right\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}'_{i,-2} &= \frac{1}{NT}\sum_{i=1}^{N}\left[\left(2\mathbf{y}_{i0} - \mathbf{y}_{i1}\right)\mathbf{y}'_{i0} + \sum_{t=1}^{T-3}\left(2\mathbf{y}_{it} - \mathbf{y}_{it-1} - \mathbf{y}_{it+1}\right)\mathbf{y}'_{it} + \left(2\mathbf{y}_{iT-2} - \mathbf{y}_{iT-3}\right)\mathbf{y}'_{iT-2}\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-3}\left(2\mathbf{y}_{it} - \mathbf{y}_{it-1} - \mathbf{y}_{it+1}\right)\mathbf{y}'_{it} + o_p\left(1\right),
\end{aligned}$$

where it can be easily verified that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-3}\left(2\mathbf{y}_{it}-\mathbf{y}_{it-1}-\mathbf{y}_{it+1}\right)\mathbf{y}_{it}'$ converges to a positive definite constant matrix as N or T or both tend to infinity, and we shall denote the limit as A_{yy}. Also,

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}' &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\begin{pmatrix}\Delta y_{2,it}\\ \Delta y_{1,it-1}\end{pmatrix}\mathbf{y}_{it}',\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\mathbf{y}_{it-2}\Delta u_{1,it},\end{aligned}$$

thus, for the asymptotic unbiasness of the first element of $\boldsymbol{\theta}$ (which is β), by denoting $\boldsymbol{\tau}=(1,0)$, the numerator becomes

$$\begin{aligned}&E\left\{\sum_{i=1}^{N}\boldsymbol{\tau}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\left[\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}_{i,-2}'\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\right]\right\}\\ =\;& tr\left\{E\left(\left[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Y}_{i,-2}A\mathbf{Y}_{i,-2}'\right]^{-1}\frac{1}{N^{3/2}T^{3/2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\boldsymbol{\tau}\Delta\mathbf{X}_j'\mathbf{Y}_{j,-2}'\right)\right\}\\ =\;& tr\left\{A_{yy}^{-1}\frac{1}{N^{3/2}T^{3/2}}\sum_{i=1}^{N}E\left(\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\boldsymbol{\tau}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right)\right\}+o_p(1),\end{aligned}$$

where

$$\begin{aligned}E\left(\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\boldsymbol{\tau}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right) &= \sum_{s,t=2}^{T}E\left[\begin{pmatrix}y_{1,it-2}\Delta u_{1,it}\\ y_{2,it-2}\Delta u_{1,it}\end{pmatrix}\left(y_{1,is-2}\Delta y_{2,is},y_{2,is-2}\Delta y_{2,is}\right)\right]\\ &= \sum_{s,t=2}^{T}E\begin{pmatrix}y_{1,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is} & y_{1,it-2}\Delta u_{1,it}y_{2,is-2}\Delta y_{2,is}\\ y_{2,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is} & y_{2,it-2}\Delta u_{1,it}y_{2,is-2}\Delta y_{2,is}\end{pmatrix}\\ &= \sum_{s\geq t-1}E\begin{pmatrix}y_{1,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is} & y_{1,it-2}\Delta u_{1,it}y_{2,is-2}\Delta y_{2,is}\\ y_{2,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is} & y_{2,it-2}\Delta u_{1,it}y_{2,is-2}\Delta y_{2,is}\end{pmatrix},\end{aligned}$$

and for $s\geq t-1$,, we have

$$\begin{aligned}&\sum_{s\geq t-1}E\left(y_{1,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is}\right)\\ =\;&\sum_{s\geq t-1}E\left(y_{1,it-2}u_{1,it}y_{1,is-2}y_{2,is}\right)-\sum_{s\geq t-1}E\left(y_{1,it-2}u_{1,it-1}y_{1,is-2}y_{2,is}\right)\\ &-\sum_{s\geq t-1}E\left(y_{1,it-2}u_{1,it}y_{1,is-2}y_{2,is-1}\right)+\sum_{s\geq t-1}E\left(y_{1,it-2}u_{1,it-1}y_{1,is-2}y_{2,is-1}\right)\\ =\;&O_p(T),\end{aligned}$$

because from (A.7), we have

$$\begin{aligned}
&\sum_{s\geq t+2} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}y_{2,is}\right)\\
=&\sum_{t=3}^{T}\sum_{s=t+2}^{T} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}y_{2,is}\right)\\
=&\sum_{t=3}^{T}\sum_{s=t+2}^{T}\left[E\left(\left(\Pi^2\right)^{(2,1)} y_{1,it-2}u_{1,it}y_{1,is-2}^2\right)+E\left(\left(\Pi^2\right)^{(2,2)} y_{1,it-2}u_{1,it}y_{1,is-2}y_{2,is-2}\right)\right],
\end{aligned}$$

and

$$\sum_{t=3}^{T}\sum_{s=t+2}^{T} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}^2\right)=\left(\Pi^2\right)^{(2,1)}\sum_{t=3}^{T}\sum_{s=t+2}^{T} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}^2\right)$$

where $A^{(i,j)}$ denotes the (i,j)-th element of matrix A. Moreover, we have $y_{1,is-2}=\phi_1\left(s-t\right)y_{1,it-2}+\phi_2\left(s-t\right)y_{2,it-2}+\sum_{s'=t-1}^{s}\left(\phi_1\left(s-s'\right)u_{1,is'}+\phi_2\left(s-s'\right)u_{2,is'}\right)$, where $\phi_i\left(s\right)=\left(\Pi^s\right)^{(i,i)}$ for $i=1,2$, then

$$\begin{aligned}
&\sum_{s=t+2}^{T} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}^2\right)\\
=&\ 2\sum_{s=t+2}^{T}\phi_1\left(s-t\right)E\left(y_{1,it-2}^2u_{1,it}^2\right)+2\sum_{s=t+2}^{T}\phi_2\left(s-t\right)E\left(y_{1,it-2}y_{2,it-2}u_{1,it}^2\right)+o_p\left(1\right)\\
=&\ O_p\left(1\right),
\end{aligned}$$

since $\sum_{s=t+2}^{T}\phi_1\left(s-t\right)$ and $\sum_{s=t+2}^{T}\phi_2\left(s-t\right)$ are finite by assumption A3. Consequently, we have

$$\sum_{t=3}^{T}\sum_{s=t+2}^{T} E\left(y_{1,it-2}u_{1,it}y_{1,is-2}^2\right)=O_p\left(T\right),$$

and

$$\sum_{s\geq t-1} E\left(y_{1,it-2}\Delta u_{1,it}y_{1,is-2}\Delta y_{2,is}\right)=O_p\left(T\right),$$

and

$$\frac{1}{N^{3/2}T^{3/2}}\sum_{i=1}^{N} E\left(\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\boldsymbol{\tau}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right)=O_p\left(\frac{1}{\sqrt{NT}}\right)=o_p\left(1\right),$$

by substituting back, we have

$$tr\left\{A_{yy}^{-1}\frac{1}{N^{3/2}T^{3/2}}\sum_{i=1}^{N} E\left(\mathbf{Y}_{i,-2}\Delta\mathbf{u}_{1i}\boldsymbol{\tau}\Delta\mathbf{X}_i'\mathbf{Y}_{i,-2}'\right)\right\}=O_p\left(\frac{1}{\sqrt{NT}}\right)=o_p\left(1\right),$$

and consequently,

$$E\left[\sqrt{NT}\left(\hat{\beta}_{PG2SLS}-\beta\right)\right]=o_p\left(1\right),$$

i.e., β is asymptotically unbiased. By using similar argument, we can show that γ_1 is also asymptotically unbiased.

For the limiting distribution of $\hat{\boldsymbol{\theta}}_{G2SLS}$, by following Arellano (2003), we can establish that

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{PG2SLS}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(0,\Omega_{PG2SLS}\right),$$

where Ω_{PG2SLS} is given in the paper.