Ann Inst Stat Math (2015) 67:39–73
DOI 10.1007/s10463-013-0438-5

The limited information maximum likelihood approach to dynamic panel structural equation models

Kentaro Akashi · Naoto Kunitomo

Received: 28 January 2013 / Revised: 15 October 2013 / Published online: 13 December 2013

Abstract We develop the panel-limited information maximum likelihood approach for estimating dynamic panel structural equation models. When there are dynamic effects and endogenous variables with individual effects at the same time, the LIML method for the filtered data does give not only a consistent estimator and asymptotic normality, but also attains the asymptotic bound when the number of orthogonal conditions is large. Our formulation includes Alvarez and Arellano (Econometrica 71:1121–1159, 2003), Blundell and Bond (Econ Rev 19-3:321–340, 2000) and other linear dynamic panel models as special cases.

Keywords Dynamic panel structural equation · LIML · Many orthogonal conditions · Forward and backward filters · Optimality

1 Introduction

There have been a number of panel data available and their analyses have been growing in many applied fields of economics in the past decades. Statistical methods on

---

Electronic supplementary material The online version of this article (doi:10.1007/s10463-013-0438-5) contains supplementary material, which is available to authorized users.

K. Akashi
Faculty of Economics, Gakushuin University,
Toshima-ku, Mejiro 1-5-1, Tokyo 113-0033, Japan
e-mail: kentaro.akashi@gakushuin.ac.jp

N. Kunitomo (✉)
Graduate School of Economics, University of Tokyo,
Bunkyo-ku, Hongo 7-3-1, Tokyo 113-0033, Japan
e-mail: kunitomo@e.u-tokyo.ac.jp

panel data have been developed, which are indispensable in econometrics (see Hsiao 2003, Arellano 2003 and Baltagi 2005 for instance). The dynamic panel models have been often used in empirical applications and the earlier investigations were from Anderson and Hsiao (1981, 1982). In a pioneering work Alvarez and Arellano (2003) investigated the asymptotic behavior of alternative estimation methods in a dynamic panel regression model when both N (the number of individuals) and T (the number of observation periods) are large. Our approach is related to these studies.

There are still non-trivial statistical problems on estimating dynamic panel econometric models to be investigated. In particular, when there are lagged endogenous variables with individual effects and the simultaneity effects in the structural equation of interest exist at the same time, the standard econometric methods including the GMM (generalized method of moments) in the econometric literature or the estimating equation (EE) method in the statistics literature do not necessarily work well due to the presence of *individual effects*, which cause the problem of *incidental parameters* when we have a long time-horizon.

In this paper, we propose *the panel-limited information maximum likelihood* (PLIML) approach to dynamic panel structural equation models. It is a simple extension of the limited information maximum likelihood (LIML) method, which was originally developed by Anderson and Rubin (1949, 1950). We intend to apply the LIML method to the dynamic panel structural models when there are dynamic effects and endogenous variables with individual effects at the same time. We need to modify the LIML method to handle the dynamic panel models with individual effects and possibly *many orthogonal conditions* because the individual effects in panel structural equations cause a source of endogeneity between the explanatory (or instrumental) variables and the explained variables, and we propose to use the filtering procedures. The PLIML method gives a consistent estimator and attains the asymptotic efficiency bound for general dynamic panel structural equation models when the relative ratio T/N is not small. In macro-panel data or long panel data, T (the number of observations over time) can be substantial and it is often important to estimate the dynamic effects in the structural equation of interest. When the panel dimensions (N, T) and the number of available instruments are not small, the approximations of the limiting distributions of estimators and test statistics based on the standard asymptotics are often poor and we need another asymptotic theory, which corresponds to the *large-K_2 asymptotics* developed by Kunitomo (1980) as an early study and it has been recently re-examined by Anderson et al. (2005, 2010, 2011).

In our framework of study, we shall consider alternative ways of filtering procedure for the original data before estimation systematically, namely, the forward-filtering and the backward-filtering. We shall show that the LIML estimation has an advantageous aspect when we use the forward-filtering and utilize many orthogonal conditions in particular. Also the usage of the backward-filtering for instruments can decrease the effects of a large number of possible instruments and the doubly-filtered LIML becomes asymptotically less biased. In a companion paper, Akashi and Kunitomo (2012) further have investigated the details of the finite sample properties of alternative estimation methods such as the WG (within groups), the GMM and the PLIML estimators in a simple setting when there are two equations. The formulation of this paper is much more general and we shall show that their findings and results on the finite sample and asymptotic properties of alternative estimators are relevant for more general dynamic panel structural equations. A related work to the LIML method in panel econometric analysis would be Alonso-Borrego and Arellano (1999).

In Sect. 2, we state the formulation of models and define the alternative estimation methods of unknown parameters in the dynamic panel structural equation model with possibly many instruments and the filtering procedures. Then in Sect. 3, we give the results on the asymptotic properties of the LIML and GMM estimation methods and the result on the asymptotic optimality. In Sect. 4, we shall report on the finite sample properties based on a set of Monte Carlo simulations. Then in Sect. 5, some concluding remarks will be given. The proofs of our theorems will be given in Sect. 6.

2 LIML approach to dynamic panel structural equation

2.1 Model

We consider the estimation problem of a dynamic panel structural equation with individual effects in the form

$$y_{it}^{(1)} = \boldsymbol{\beta}_2^{'} \mathbf{y}_{it}^{(2)} + \boldsymbol{\gamma}_1^{'} \mathbf{z}_{it-1}^{(1)} + \eta_i + u_{it}, \tag{1}$$

where $y_{it}^{(1)}$ and $\mathbf{y}_{it}^{(2)} = (y_{it}^{(j)})$, $(j = 2, \ldots, 1 + G_2)$ are $1 + G_2$ endogenous variables, $\mathbf{z}_{it-1}^{(1)}$ is the $K_1 \times 1$ vector of the included predetermined variables in (1), η_i $(i = 1, \ldots, N)$ are individual effects, u_{it} are mutually independent (over individuals and periods) disturbance terms with $\mathcal{E}[u_{it}] = 0$, $\mathcal{E}[u_{it}^2] = \sigma^2$, and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters. We allow that the explanatory variables include the lagged endogenous variables and the observations are for $i = 1, \ldots, N; t = 1, \ldots, T$ and the sample size is NT $(= n)$.

We assume that the reduced form is written as

$$\mathbf{y}_{it} = \boldsymbol{\Pi} \mathbf{z}_{it-1} + \boldsymbol{\pi}_i + \mathbf{v}_{it}, \tag{2}$$

where $\mathbf{y}_{it} = (y_{it}^{(j)})$, $(j = 1, \ldots, G)$, $\mathbf{z}_{i,t-1} = (z_{i,t-1}^{(j)}, j = 1, \ldots, K)$, and $\mathcal{E}[\mathbf{v}_{it}] = \mathbf{0}$ and $\mathcal{E}[\mathbf{v}_{it}\mathbf{v}_{it}^{'}] = \boldsymbol{\Omega} > 0$ (a positive definite matrix). We also assume that the instrumental variables $\mathbf{z}_{it-1}$ are $\mathcal{F}_{t-1}$ adapted, and $\mathcal{F}_{t-1}$ is the $\sigma-$field generated by $\{\mathbf{v}_{it-h}, \boldsymbol{\pi}_i\}_{h=1}^{\infty}$. (We use the notation $\mathcal{E}_t[\ .\ ] = \mathcal{E}[\ .\ |\mathcal{F}_{t-1}]$ for the conditional expectation operator.) The relation between the coefficients in (1) and (2) gives the condition $(1, -\boldsymbol{\beta}_2^{'})\boldsymbol{\Pi} = (\boldsymbol{\gamma}_1^{'}, \mathbf{0}^{'})$ and $\boldsymbol{\Pi}_{12}^{'} = \boldsymbol{\beta}_2^{'}\boldsymbol{\Pi}_{22}$, where $\boldsymbol{\Pi}_1^{'} = (\boldsymbol{\Pi}_{11}, \boldsymbol{\Pi}_{21}^{'})$ is a $K_1 \times G$ matrix, $\boldsymbol{\Pi}_2^{'} = (\boldsymbol{\Pi}_{12}, \boldsymbol{\Pi}_{22}^{'})$ is a $K_2 \times G$ matrix and the $G \times (K_1 + K_2)$ matrix of coefficients is partitioned as

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_{11}^{'} & \boldsymbol{\Pi}_{12}^{'} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix}. \tag{3}$$

Although we may call (2) as the *reduced form*, the predetermined variables in $\mathbf{z}_{it-1}$ are correlated with unobserved variables $(\boldsymbol{\pi}_i + \mathbf{v}_{it})$ since $\mathcal{E}[\mathbf{z}_{it-1}\boldsymbol{\pi}_i'] \neq \mathbf{O}$ in general and it makes the panel econometric model of (1) and (2) different from the classical simultaneous equation models. We give two examples in the econometric literature.

Example 1 Blundell and Bond (2000) have considered the simple model of a dynamic panel structural equation with two endogenous variables given by

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \eta_i + u_{it}, \tag{4}$$

$$y_{it}^{(2)} = \gamma_2 y_{it-1}^{(2)} + \delta\eta_i + v_{it}, \tag{5}$$

where the disturbance terms u_{it} and v_{it} are correlated. The equation (5) can be regarded as a reduced form equation and the estimation of γ_2 was considered by Alvarez and Arellano (2003). They used the forward-filtering to data and proposed to use all past values y_{is} $(s < t)$ at period t as instruments, i.e., the number of instruments is $T(T-1)/2$ $(= r_n)$. Hayakawa (2006, 2009), on the other hand, has suggested to apply the backward-filter to generate instruments.

Example 2 The Panel Vector Autoregressive (Panel VARs) model suggested by Holtz-Eakin et al. (1988) can be written as

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} x_{it} + \eta_i + u_{it}, \tag{6}$$

and the extended reduced form is defined by

$$\begin{pmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ y_{it-1}^{(2)} \\ x_{it+1} \\ x_{it} \end{pmatrix} = \begin{pmatrix} \pi_{11}^* & \pi_{12}^* & \pi_{13}^* & \pi_{14}^* & 0 \\ 0 & \pi_{21}^* & \pi_{22}^* & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{31}^* & \pi_{32}^* \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ y_{it-2}^{(2)} \\ x_{it} \\ x_{it-1} \end{pmatrix} + \begin{pmatrix} \pi_i^{*(1)} \\ \pi_i^{*(2)} \\ 0 \\ \pi_i^* \\ 0 \end{pmatrix} + \begin{pmatrix} v_{it}^{*(1)} \\ v_{it}^{*(2)} \\ 0 \\ \epsilon_{it+1}^* \\ 0 \end{pmatrix},$$

where the first two rows are the Panel VARs model and x_{it} is the included independent variable.

There are several important aspects of the problem of estimating equations with instrumental variables in the dynamic panel structural equations. First, the standard statistical estimation methods do not necessarily have desirable properties because of the presence of individual effects η_i $(i = 1, \ldots, N)$. To deal with this problem, there have been several statistical procedures developed for the estimating equations with individual effects. Second, some estimation procedures based on the standard asymptotics $(N \to \infty, T < \infty)$ have substantial bias when the panel models become dynamic in the sense that we have lagged endogenous variables as explanatory variables. This is because even if we used the appropriate filtering method to remove the individual effects, their influence cause the second-order bias through the past variables and it becomes serious when T becomes large as Akashi and Kunitomo (2012) illustrated. Although we can remove the source of correlations among the lagged endogenous variables and heterogeneity of individual using the filtering procedure, we cannot remove the simultaneity by the standard procedure.

We shall develop a new statistical procedure which may overcome these problems at the same time by applying the limited information maximum likelihood (PLIML) method. The asymptotic properties of the LIML method for estimating structural equations including its asymptotic optimality have been recently investigated by Anderson et al. (2010, 2011) when there are *many instruments*. We extend their analysis to the PLIML method when the number of instruments increases as T, which may be quite natural in the estimation problem of dynamic panel structural equations. Before we apply the LIML method, however, first we need to use the filtering procedure to the original data, which is a data transformation. There are alternative filtering procedures, which correspond to either the forward direction filtering or the backward direction filtering, to remove their individual effects.

2.2 Instrumental variables and filtering procedures

Let $\mathbf{y}_i^{(1)} = (y_{it}^{(1)})$, $\mathbf{Y}_i^{(2)} = (\mathbf{y}_{it}^{(2)'})$ and $\mathbf{Z}_{i(-1)}^{(1)} = (\mathbf{z}_{it-1}^{(1)'})$ be $T \times 1$, $T \times G_2$ and $T \times K_1$ matrices. We define the forward deviation operator $\mathbf{A}_f$, which is the $(T-1) \times T$ upper triangular matrix used by Arellano and Bover (1995) and Alvarez and Arellano (2003) such that $\mathbf{A}_f\mathbf{A}_f^{'} = \mathbf{I}_{T-1}$, $\boldsymbol{\iota} = (1, \ldots, 1)^{'}$ and $\mathbf{A}_f^{'}\mathbf{A}_f = \mathbf{Q}_T = \mathbf{I}_T - \boldsymbol{\iota}_T\boldsymbol{\iota}_T^{'}/T$. We apply the forward deviation operator $\mathbf{y}_i^{(1)}$, $\mathbf{Y}_i^{(2)}$, and $\mathbf{Z}_{i(-1)}^{(1)}$ and then denote the resulting variables as $\mathbf{y}_i^{(1,f)} = (y_{it}^{(1,f)})$, $\mathbf{Y}_i^{(2,f)} = (\mathbf{y}_{it}^{(2,f)'})$ and $\mathbf{Z}_i^{(1,f)} = (\mathbf{z}_{it-1}^{(1,f)'})$. For an example, we denote

$$\mathbf{y}_{it}^{(2,f)} = c_t\left[\mathbf{y}_{it}^{(2)} - \frac{1}{T-t}(\mathbf{y}_{it+1}^{(2)} + \cdots + \mathbf{y}_{iT}^{(2)})\right] \tag{7}$$

where $c_t^2 = (T-t)/(T-t+1)$ for $t = 1, \ldots, T-1,\ T \geq 2$. Using the forward-filtered variables, we write for $t = 1, \ldots, T-1$ as

$$y_{it}^{(1,f)} = \boldsymbol{\beta}_2^{'}\mathbf{y}_{it}^{(2,f)} + \boldsymbol{\gamma}_1^{'}\mathbf{z}_{it-1}^{(1,f)} + u_{it}^{(f)}, \tag{8}$$

where $\mathbf{u}_i^{(f)} = (u_{it}^{(f)})$ is the transformed $(T-1) \times 1$ vector by $\mathbf{u}_i^{(f)} = \mathbf{A}_f\mathbf{u}_i$ from the $T \times 1$ disturbance vector $\mathbf{u}_i = (u_{it})$. Here, we have the relation that $\mathcal{E}[z_{it}^{(1,f)}u_{it}^{(f)}] \neq \mathbf{0}$, consequently.

We also define the backward operator $\mathbf{A}_b$ with the filter direction to the past, which is the $(T-1) \times T$ lower triangular matrix as used by Hayakawa (2006). This procedure removes the individual effects from the instrumental variables. We denote the transformed instrumental variables $\mathbf{Z}_{i(-1)}^{(b)} = (\mathbf{z}_{it-1}^{(b)'})$ and

$$\mathbf{z}_{it-1}^{(b)} = b_t\Big[\mathbf{z}_{it-1} - \frac{1}{t}(\mathbf{z}_{it-2} + \cdots + \mathbf{z}_{i0} + \mathbf{z}_{i(-1)})\Big], \tag{9}$$

where $b_t^2 = t/(t+1)$ for $t = 1, \ldots, T-1$, and we include $\mathbf{z}_{i(-1)}$ to simplify the notation of the index range.

The forward-filtering enables us to make use of the orthogonal conditions for the disturbance terms. The backward-filtering removes the individual effects from instrumental variables. In our analysis, we use two types of transformations on the instrumental variables, and the instrumental matrices at period t are defined by

$$\mathbf{Z}_t^{(a)} = \begin{pmatrix} \mathbf{z}_{1(t-1)}^{(a)} & \cdots & \mathbf{z}_{N(t-1)}^{(a)} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_{10}^{(a)} & \cdots & \mathbf{z}_{N0}^{(a)} \end{pmatrix}', \ \mathbf{Z}_t^{(b)} = \left(\mathbf{z}_{1(t-1)}^{(b)}, \ldots, \mathbf{z}_{N(t-1)}^{(b)}\right)', \tag{10}$$

where $\mathbf{z}_{it-1}^{(a)}$ is the $K_* \times 1$ vector such that $\mathbf{z}_{it-1}^{(a)} = \mathbf{J}_{K_*}' \mathbf{z}_{it-1}$, and the selection matrix $\mathbf{J}_{K_*}'$ chooses the nearest lagged variables to $t-1$ while $\mathbf{Z}_t^{(b)}$ is the $N \times K$ matrix. The dimension reduction from K to K_* is often needed for the full rank of $(\mathbf{Z}_t^{(a)'}\mathbf{Z}_t^{(a)})$, where $\mathbf{Z}_t^{(a)}$ is the $N \times (K_* t)$.

We shall consider two alternative ways of the instrumental variables.

(a) At period t, we use all available lagged variables after applying the forward-filtering to the structural equation as suggested by Arellano and Bover (1995) and Alvarez and Arellano (2003). Since the instruments $\mathbf{z}_{is}^{(a)}$ $(0 \le s < t)$ are generated by the past information at t and the individual effects, the orthogonal conditions at period t can be written as

$$\mathcal{E}\left[\mathbf{z}_{is}^{(a)} u_{it}^{(f)}\right] = \mathbf{0} \ (0 \le s < t \le T). \tag{11}$$

When T is large, the number of orthogonal conditions can be large if we use all orthogonal conditions imposed.

(b) Alternatively, at period t, we can use the only (a fixed number of) lagged variables included in the reduced form after applying the backward-filtering to all instruments. Since the instruments $\mathbf{z}_{is}^{(b)}$ $(0 \le s < t)$ are generated by the past information at t and the individual effects are removed, the orthogonal conditions at period t used can be written as

$$\mathcal{E}\left[\mathbf{z}_{it-1}^{(b)} u_{it}^{(f)}\right] = \mathbf{0},$$

which are the standard orthogonal conditions except the effects of the forward-filtering and backward-filtering for the original data.

Then, we consider two asymptotic sequences with respect to two dimensions N and T in alternative ways. We define the total number of orthogonal conditions used as r_n and consider the ratio r_n/n (the total sample NT $(= n)$) as

$$\text{(a)} \quad \frac{K_* T(T-1)}{2NT} \overset{N,T\to\infty}{\to} c_a = \left(\frac{K_*}{2}\right) \lim_{N,T\to\infty} \left(\frac{T}{N}\right). \tag{12}$$

$$\text{(b)} \quad \frac{K(T-1)}{N_0 T} \overset{T\to\infty}{\to} c_b = \frac{K}{N_0}, \tag{13}$$

where we use the notation $N\ (= N_0)$ as a fixed integer. Then we shall investigate the asymptotic behaviors of estimators when the sequence of ratio can be a reasonable approximation when r_n and N are large under panel structural equation model. When the number of instruments used is reduced to $O(T)$, the doubly-filtered LIML estimator does not need the double asymptotics $N, T \to \infty$ and the number of individuals is regarded as a fixed number.

2.3 The LIML and GMM Estimation

Let $\mathbf{y}_t^{(f)} = (y_{it}^{(1,f)}, \mathbf{y}_{it}^{(2,f)'})'$ be $(1 + G_2)$ vectors and

$$\mathbf{Y}_t^{(f)'} = \left(\mathbf{y}_{1t}^{(f)}, \ldots, \mathbf{y}_{Nt}^{(f)}\right),\ \mathbf{Z}_t^{(1,f)'} = \left(\mathbf{z}_{1t}^{(1,f)}, \ldots, \mathbf{z}_{Nt}^{(1,f)}\right),$$

be $(1+G_2) \times N$, and $K_1 \times N$ matrices of the forward-filtered variables, respectively. Using these notations, we define two $(1 + G_2 + K_1) \times (1 + G_2 + K_1)$ matrices as

$$\mathbf{G}^{(f)} = \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} \mathbf{M}_t \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{14}$$

and

$$\mathbf{H}^{(f)} = \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} [\mathbf{I}_N - \mathbf{M}_t] \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{15}$$

where the projection matrices $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$ are defined by $\mathbf{M}_t^{(a)} = \mathbf{Z}_t^{(a)}(\mathbf{Z}_t^{(a)'}\mathbf{Z}_t^{(a)})^{-1}\mathbf{Z}_t^{(a)'}$ and $\mathbf{M}_t^{(b)} = \mathbf{Z}_t^{(b)}(\mathbf{Z}_t^{(b)'}\mathbf{Z}_t^{(b)})^{-1}\mathbf{Z}_t^{(b)'}$. Then the LIML estimator $\hat{\boldsymbol{\theta}}_{\text{LI}}^{(.)} = (\hat{\boldsymbol{\beta}}_{2.\text{LI}}^{'}, \hat{\boldsymbol{\gamma}}_{1.\text{LI}}^{'})'$ of $(1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})' = (1, -\boldsymbol{\theta}')'$ is defined by

$$\left[\frac{1}{n}\mathbf{G}^{(f)} - \lambda_n \frac{1}{q_n}\mathbf{H}^{(f)}\right] \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\theta}}_{\text{LI}}^{(.)} \end{bmatrix} = \mathbf{0}, \tag{16}$$

where $n = NT$, $q_n = n - r_n$ and λ_n is the smallest root of

$$\left|\frac{1}{n}\mathbf{G}^{(f)} - l\frac{1}{q_n}\mathbf{H}^{(f)}\right| = 0. \tag{17}$$

In this formulation, we use the notation $\hat{\boldsymbol{\theta}}_{\mathrm{LI}}^{(.)} = \hat{\boldsymbol{\theta}}_{\mathrm{LI}}^{(a)}$ in the case of $\mathbf{M}_t^{(a)}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{LI}}^{(b)}$ in the case of $\mathbf{M}_t^{(b)}$, respectively. The solution to (16) gives the minimum of the variance ratio

$$\mathrm{VR}_n = \frac{\left[1, -\boldsymbol{\theta}^{'}\right]\mathbf{G}^{(f)}\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}}{\left[1, -\boldsymbol{\theta}^{'}\right]\mathbf{H}^{(f)}\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}}. \tag{18}$$

Similarly, we define the panel GMM (or two-stage least squares TSLS) estimator, $\hat{\boldsymbol{\theta}}_{GM}^{(.)} = (\hat{\boldsymbol{\beta}}_{2.GM}^{'}, \hat{\boldsymbol{\gamma}}_{1.GM}^{'})^{'}$ of $(1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})^{'} = (1, -\boldsymbol{\theta}')'$ by

$$\left[\mathbf{0}, \mathbf{I}_{G_2+K_1}\right]\sum_{t=1}^{T-1}\begin{bmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{bmatrix}\mathbf{M}_t\left[\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right]\begin{bmatrix}1\\ -\hat{\boldsymbol{\theta}}_{GM}^{(.)}\end{bmatrix} = \mathbf{0} \tag{19}$$

and we denote $\hat{\boldsymbol{\theta}}_{GM}^{(a)}$ and $\hat{\boldsymbol{\theta}}_{GM}^{(b)}$ accordingly. It minimizes the numerator of the variance ratio in (18). The LIML and TSLS estimation methods were originally developed by Anderson and Rubin (1949, 1950), and we modify them slightly to develop the panel LIML and the panel GMM (or TSLS) methods for the dynamic panel simultaneous equations model with individual effects.

3 Asymptotic properties of the LIML and GMM estimators

3.1 Asymptotic distributions

In this section, we shall derive the limiting distributions of the LIML and the GMM estimators when we have the representation $\mathbf{y}_{it}^* = \boldsymbol{\Pi}^*\mathbf{y}_{i,t-1}^* + \boldsymbol{\pi}_i^* + \mathbf{v}_{it}^*$ and $G_* \times 1$ vector $\mathbf{y}_{it}^*$ includes $\mathbf{y}_{it}$ $(1 + G_2 \le G \le G_*)$. Then we shall investigate the case when $1 + G_2 \le G \le G_*$ when $\mathbf{y}_{it}^*$ can be degenerated as above and both Examples 1 and 2 are some special cases. There can be some possible ways of extensions of our arguments, but then we would need quite lengthy derivations as Sect. 6 of this paper has suggested.

Let $\mathbf{w}_{it} = \mathbf{y}_{it}^* - \boldsymbol{\mu}_i$ and $(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)\boldsymbol{\mu}_i = \boldsymbol{\pi}_i^*$. We make a set of assumptions on the moments of disturbances and the dynamics of the underlying process $\{\mathbf{w}_{it}\}$ satisfying

$$\mathbf{w}_{it} = \boldsymbol{\Pi}^*\mathbf{w}_{it-1} + \mathbf{v}_{it}^*. \tag{20}$$

(A1) $\{\mathbf{v}_{it}^*\}$ $(i = 1, \ldots, N;\ t = 1, \ldots, T)$ are i.i.d. across time and individuals and independent of $\boldsymbol{\pi}_i^*$ and $\mathbf{z}_{i0}$ with $\mathcal{E}[\mathbf{v}_{it}^*] = \mathbf{0}$, $\mathcal{E}[\mathbf{v}_{it}^*\mathbf{v}_{it}^{*'}] = \boldsymbol{\Omega}^*$ and $\mathcal{E}[\|\mathbf{v}_{it}\|^8]$ exists.
(A2) The initial observation satisfies $\mathbf{y}_{i0}^* = (\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}\boldsymbol{\pi}_i^* + \mathbf{w}_{i0}$ $(i = 1, \ldots, N)$, where $\mathbf{w}_{i0}$ is independent of $\boldsymbol{\pi}_i^*$ and i.i.d. with the steady state distribution of the homogenous process such that $\mathbf{w}_{i0} = \sum_{j=0}^{\infty}\boldsymbol{\Pi}^j\mathbf{v}_{i,-j}$. All roots of $\left|\boldsymbol{\Pi}^* - \lambda\mathbf{I}_{G_*}\right| = 0$ satisfy the stationarity condition $|\lambda_k| < 1$ $(k = 1, \ldots, G_*)$.

(A3) There exists an $K_* \times 1$ $(K_* \leq K)$ vector of instrumental variables $\mathbf{z}_{i,t-1}$ in (2) such that $\mathbf{Z}_t^{(a)}$ and $\mathbf{Z}_t^{(b)}$ in (10) are non-degenerate.

The assumptions (A1) and (A2) are analogous to some conditions used by Alvarez and Arellano (2003). They imply that the underlying processes for $\{\mathbf{y}_{it}\}$ and $\{\mathbf{w}_{it}\}$ are stationary. We shall make use of the assumption on initial condition to prove Lemmas in Sect. 6, but they could be relaxed at the expense of the resulting lengthy derivations.

To state main theoretical results in concise ways, we prepare some notations such that $\mathcal{E}[\mathbf{v}_{it}\mathbf{v}_{it}'] = \boldsymbol{\Omega}$, $\sigma^2 = \mathcal{E}[u_{it}^2] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$, $\mathbf{u}_{it}^{\perp} = [\mathbf{0}, \mathbf{I}_{G_2}]\Big[\mathbf{v}_{it} - \mathrm{Cov}(\mathbf{v}_{it}, u_{it})u_{it}/\sigma^2\Big]$, $\boldsymbol{\Phi}^* = \mathbf{D}'\mathbf{J}_K'\mathcal{E}[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}']\mathbf{J}_K\mathbf{D}$, $\mathbf{D} = (\boldsymbol{\Pi}_2, \mathbf{J}_{K_1})$ and $\mathbf{J}_{K_1}' = (\mathbf{I}_{K_1}, \mathbf{O})$.

We first discuss Case (a) when we take the forward-filtering procedure and then apply the LIML and the GMM estimation. We denote $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and we have the next result whose proof will be in Sect. 6.

Theorem 1 *Suppose Assumptions* (A1)–(A3) *hold and* $\boldsymbol{\Phi}^*$ *is a positive definite matrix. Consider the double asymptotics* $N, T \to \infty$ *and assume that* $0 \leq K_* \lim_{N,T\to\infty}(T/N) < 1$.

(i) Assume $T/N \to c_a > 0$ *as* $N \to \infty$ *and* $T \to \infty$. *Then*

$$\hat{\boldsymbol{\theta}}_{GM}^{(a)} - \boldsymbol{\theta} \xrightarrow{p} \left[\boldsymbol{\Phi}^* + c_a \begin{pmatrix} \mathbf{J}_{*G_2}'\boldsymbol{\Omega}\mathbf{J}_{*G_2} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\right]^{-1} \begin{bmatrix} c_a\mathbf{J}_{*G_2}'\boldsymbol{\Omega}\boldsymbol{\beta} \\ \mathbf{O} \end{bmatrix},$$

where $\mathbf{J}_{*G_2}' = [\mathbf{0}, \mathbf{I}_{G_2}]$.

(ii) For $c_a = 0$, $0 \leq \lim_{N,T\to\infty}(T^3/N) = d_a < \infty$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_0^{(a)}, \sigma^2\boldsymbol{\Phi}^{*-1}\right), \quad (21)$$

where

$$\mathbf{b}_0^{(a)} = \left[\frac{d_a^{1/2}K_*}{2}\right]\boldsymbol{\Phi}^{*-1}\begin{pmatrix} \mathbf{J}_{*G_2}'\boldsymbol{\Omega}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}.$$

(iii) For $c_a = 0$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{*-1}). \quad (22)$$

(iv) For $0 < c_a < 1/2$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_c^{(a)}, \boldsymbol{\Psi}^{*(a)}\right), \quad (23)$$

where $c_{*a} = c_a/(1-c_a)$, $\mathbf{J}_{G_2}^{'} = (\mathbf{I}_{G_2}, \mathbf{O})$,

$$\boldsymbol{\Psi}^{*(a)} = \boldsymbol{\Phi}^{*-1}\left[\sigma^2\boldsymbol{\Phi}^* + \mathbf{J}_{G_2}(c_a[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega}]_{22} + \boldsymbol{\Xi}_4^{(a)})\mathbf{J}_{G_2}^{'}\right]\boldsymbol{\Phi}^{*-1},$$

$[\cdot]_{22}$ *is the* (2,2)*-th element* ($G_2 \times G_2$ *matrix*) *of the partitioned* $(1+G_2)\times(1+G_2)$ *matrix,*

$$\boldsymbol{\Xi}_4^{(a)} = \left(\frac{1}{1-c_a}\right)^2 \mathcal{E}\left[\left(u_{it}^2 - \sigma^2\right)\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}\right]\left[\operatorname{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{d}_t^{(a)'}\mathbf{d}_t^{(a)} - c_a^2\right],$$

$$\mathbf{b}_c^{(a)} = -\left(\frac{K_*}{2}\right)^{1/2}\frac{c_a^{1/2}}{(1-c_a)}\boldsymbol{\Phi}^{*-1}\mathbf{D}'\mathbf{J}_{G_*}^{'}(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}\mathcal{E}\left[\mathbf{v}_{it}^* u_{it}\right],$$

$\mathbf{d}_t^{(a)} = (d_{it}^{(a)}) = \operatorname{diag}(\mathbf{M}_t^{(a)})$ *and* $\mathbf{W}_{t-1} = (\mathbf{w}_{1(t-1)}, \ldots, \mathbf{w}_{N(t-1)})'$ *is the* $N \times K_*$ *matrix consisting of* $\{\mathbf{w}_{it}\}$.

The asymptotic covariances in Theorem 1 in some case look complicated due to the term $\boldsymbol{\Xi}_4^{(a)}$ which depends on the fourth moments of disturbances. In our numerical analysis, the effects of fourth moments are usually negligible. When $c_a = 0$, both the LIML and the GMM estimators are consistent and they have the asymptotic normality. But the GMM estimator has an extra asymptotic bias $\mathbf{b}_0^{(a)}$ due to the presence of endogenous variables. This result agrees with the one by Anderson et al. (2010) for linear structural equation models with many instruments. The asymptotic bias $\mathbf{b}_c^{(a)}$ due to the presence of forward-filtering is similar to the one by Alvarez and Arellano (2003) for a simple dynamic regression model. When $c_a > 0$, however, the LIML estimator is still consistent and it has the asymptotic normality while the GMM estimator is inconsistent.

Next, we consider Case (b). When we apply the backward-filtering procedure to the set of instrumental variables including the lagged endogenous variables. We take $\mathbf{M}_t = \mathbf{M}_t^{(b)}$. In the case (b) we have the next result whose proof will be in Section 6.

Theorem 2 *Suppose Assumptions* (A1)–(A3) *hold and* $\boldsymbol{\Phi}^*$ *is a positive definite matrix. Let* $T \to \infty$ *and set* $K/N = c_b$. *We take* $c_b = 0$ *when* $N \to \infty$ *while* K *is a fixed positive integer. We denote* $c_b > 0$ *when* N *and* K *are bounded and positive integers.*

(i) Consider the case when $c_b > 0$ *and* N *is bounded. Then as* $T \to \infty$,

$$\hat{\boldsymbol{\theta}}_{GM}^{(b)} - \boldsymbol{\theta} \xrightarrow{p} \left[\boldsymbol{\Phi}^* + c_b\begin{pmatrix}\mathbf{J}_{*G_2}^{'}\boldsymbol{\Omega}\mathbf{J}_{*G_2} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{pmatrix}\right]^{-1}\begin{bmatrix}c_b\mathbf{J}_{*G_2}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}\\ \mathbf{O}\end{bmatrix},$$

where $\mathbf{J}_{*G_2}^{'} = [\mathbf{0}, \mathbf{I}_{G_2}]$.

(ii) For $c_b = 0$ *or* $N \to \infty$, $0 \le \lim_{N,T\to\infty}(T/N) = d_b < \infty$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_0^{(b)}, \sigma^2\boldsymbol{\Phi}^{*-1}\right), \quad (24)$$

where

$$\mathbf{b}_0^{(b)} = [d_b^{1/2} K]\boldsymbol{\Phi}^{*-1} \begin{pmatrix} \mathbf{J}'_{*G_2}\boldsymbol{\Omega\beta} \\ \mathbf{0} \end{pmatrix}.$$

(iii) For $c_b = 0$ *or* $N \to \infty$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{*-1}). \tag{25}$$

(iv) For $0 < c_b < 1$ *or* $N = N_0$ *is fixed,*

$$\sqrt{N_0 T}\left(\hat{\boldsymbol{\theta}}_{LM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^{*(b)}), \tag{26}$$

where $c_{*b} = c_b/(1-c_b)$,

$$\boldsymbol{\Psi}^{*(b)} = \boldsymbol{\Phi}^{*-1}\left[\sigma^2\boldsymbol{\Phi}^* + \mathbf{J}_{G_2}(c_{*b}\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega\beta\beta}'\boldsymbol{\Omega}\right]_{22} + \boldsymbol{\Xi}_4^{(b)})\mathbf{J}'_{G_2} + \boldsymbol{\Xi}_3^{(b)} + \boldsymbol{\Xi}_3^{(b)'}\right]\boldsymbol{\Phi}^{*-1},$$

$$\boldsymbol{\Xi}_3^{(b)'} = \begin{pmatrix} \frac{1}{1-c_b}\mathcal{E}\left[u_{it}^2\mathbf{u}_{it}^{\perp}\right] \text{ p}\lim_{T\to\infty}\frac{1}{N_0 T}\sum_{t=1}^{T-1}\left[\mathbf{d}_t^{(b)'}\mathbf{W}_{t-1}\right]\mathbf{J}_K\mathbf{D} \\ \mathbf{O} \end{pmatrix},$$

$$\boldsymbol{\Xi}_4^{(b)} = \left(\frac{1}{1-c_b}\right)^2 \mathcal{E}\left[(u_{it}^2 - \sigma^2)\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}\right]\left[\text{p}\lim_{T\to\infty}\frac{1}{N_0 T}\sum_{t=1}^{T-1}\mathbf{d}_t^{(b)'}\mathbf{d}_t^{(b)} - c_b^2\right],$$

$\mathbf{d}_t^{(b)} = (d_{it}^{(b)}) = \text{diag}(\mathbf{M}_t^{(b)})$ *and* $\mathbf{W}_{t-1} = (\mathbf{w}_{1(t-1)}, \ldots, \mathbf{w}_{N(t-1)})'$ *is the* $N \times G_*$ *matrix consisting of* $\{\mathbf{w}_{it}\}$.

When $c_b = 0$, both the LIML and the GMM estimators are consistent and they have the asymptotic normality. But the GMM estimator has an extra asymptotic bias. When $c_b > 0$, however, the LIML estimator is also consistent and almost medium-unbiased (i.e., $\mathbf{b}_c^{(b)} = \mathbf{0}$) and it has the asymptotic normality while the GMM estimator is inconsistent.

We notice that $\boldsymbol{\Phi}^*$ is the same in both our theorems, so that the backward-filtered instruments can be considered as the optimal instruments in the double asymptotics. But when $c_b > 0$ and the fixed-N or the large-K_2 asymptotics holds, then the second term of the asymptotic covariance becomes large, so that the large-K_2 improves the approximation of limiting distributions by capturing the number K and possibly large fixed N_0. On the other hand, the GMM estimator has the asymptotic bias even when $N \to \infty$. If $N \to \infty$, the doubly-filtered LIML has no bias asymptotically and attains the asymptotic efficiency bound $\sigma^2\boldsymbol{\Phi}^{*-1}$, which is the standard bound when $\boldsymbol{\pi}_i^* = \mathbf{0}$ $(i = 1, \ldots, N)$ and T is a fixed integer. In our numerical analysis, the third- and fourth-order moments of disturbances are usually negligible.

3.2 An asymptotic bound and optimality

For the estimation problem of the vector of structural parameters $\boldsymbol{\theta}$, it may be natural to consider a set of statistics of two $(1+G_2+K_1)\times(1+G_2+K_1)$ random matrices $\mathbf{G}^{(f)}$ and $\mathbf{H}^{(f)}$. It includes the GMM and the LIML estimators for instance. As a related earlier study, Hahn and Kuersteiner (2002) have investigated the asymptotic bound in the panel estimation when there does not exist any endogeneity and the disturbances are normally distributed. Anderson et al. (2010) have developed the asymptotic bound when there are *many instruments* in the structural equation models.

We consider a class of estimators which are some smooth functions of $\mathbf{G}^{(f)}$ and $\mathbf{H}^{(f)}$. It may be natural to restrict this class because they are sufficient statistics when there are no individual effects under the normal disturbances. Then we have some results on the asymptotic optimality and the proof is quite similar to the corresponding ones in Anderson et al. (2010).

Theorem 3 *In the panel structural equation models of* (1) *and* (2), *define the class of consistent estimators for* $\boldsymbol{\theta} = (\boldsymbol{\beta}_2^{'}, \boldsymbol{\gamma}_1^{'})^{'}$ *by*

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\gamma}}_1 \end{pmatrix} = \phi(\mathbf{G}^{(f)},\ \mathbf{H}^{(f)}), \tag{27}$$

where ϕ *is continuously differentiable and its derivatives are bounded at the probability limits of random matrices* $(1/n)\mathbf{G}^{(f)}$ *and* $(1/q_n)\mathbf{H}^{(f)}$.

(i) Then either under the conditions of Theorem 1 *or Theorem* 2, *as* $T\to\infty$ *with* $c_a = 0$ *or* $c_b = 0$,

$$\sqrt{NT}\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{b}_0^{(\cdot)}, \boldsymbol{\Psi}), \tag{28}$$

where

$$\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^{*} \tag{29}$$

and $\boldsymbol{\Psi}^{*}$ *and* $\mathbf{b}_0^{(\cdot)}$ *are given in Theorems* 1 *and* 2. *The LIML estimator and the bias-adjusted GMM estimator attain the asymptotic bound.*

(ii) When $0 < c_a < 1/2$ *or* $0 < c_b < 1$ *in Theorem* 1 *or Theorem* 2

$$\sqrt{NT}\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{b}_c^{(\cdot)}, \boldsymbol{\Psi}), \tag{30}$$

where $\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^{*}$ *and the asymptotic bias* $\mathbf{b}_c^{(\cdot)}$ *caused by the forward-filter depends on* $\phi(\mathbf{G}^{(f)}, \mathbf{H}^{(f)})$. *The LIML estimator attains the asymptotic bound.*

This is a result on the asymptotic efficiency bound for dynamic panel structural equation models with individual effects and endogeneity at the same time. It can be regarded as an extension of Theorem 4 of Anderson et al. (2010) for the linear structural equation in the simultaneous equation systems. Because of individual effects in the panel structural equations and the filtering problem, there are additional features on the asymptotic optimality.

4 On finite sample properties

It is important to investigate the finite sample properties of estimators partly because they are not necessarily similar to their asymptotic properties. One simple example would be the fact that the exact moments of some estimators do not necessarily exist. In that case, it may be meaningless to compare the exact MSEs of alternative estimators and their Monte Carlo analogs.

In our experiments, we took Example 2 ($K = 4, K_* = 3, K_1 = 2, G_2 = 1$) in Sect. 2 as a typical example.[1] In Example 2 we set the unknown parameters such as $(\beta_2, \gamma_{11}) = (.5, .5)$, $\gamma_{12} = .3$, and $(\omega_{11}, \omega_{12}, \omega_{22}) = (1.0, .3, 1.0), (1.5, 1.0, 1.0)$ where we take $\boldsymbol{\gamma}_1 = (\gamma_{11}, \gamma_{12})'$ and $\boldsymbol{\Omega} = (\omega_{ij})$. Also we control the variance of each components of $\boldsymbol{\pi}_i$ as 1. Our experiments are similar to the ones reported in Akashi (2008), and Akashi and Kunitomo (2012). Then we generate a large number of random variables by simulations and calculate the empirical distribution functions of the GMM and LIML estimators in their normalized forms. We repeat 5,000 replications for each case and the smoothing technique to estimate the empirical distribution functions. The details of simulations are similar to those explained by Anderson et al. (2005, 2011). We report only the results for $(N, T) = (100, 25)$ and $(100, 50)$ as the typical cases among a large number of our simulations. All simulations reported in this section have been done under the situation that the disturbance terms are normally distributed. Although we have done a number of simulations when the disturbances are non-normal such as the t-distribution, we have omitted them to save space because they are quite similar to the results reported.

We examine the distribution functions of the LIML and GMM estimators in two normalizations. The first one is in terms of

$$\frac{\sqrt{NT}}{\sigma}\begin{bmatrix}(\phi^{11})^{-1/2} & 0\\ 0 & (\phi^{22})^{-1/2}\end{bmatrix}\begin{bmatrix}\hat{\beta}_2-\beta_2\\ \hat{\gamma}_1-\gamma_1\end{bmatrix}, \tag{31}$$

where ϕ^{11} and ϕ^{22} are the (1,1)-th element and (2.2)-th element of $\boldsymbol{\Phi}^{*-1}$, respectively. The second normalization is

$$\sqrt{NT}\begin{bmatrix}\psi_{11}^{-1/2} & 0\\ 0 & \psi_{22}^{-1/2}\end{bmatrix}\left[\begin{pmatrix}\hat{\beta}_2-\beta_2\\ \hat{\gamma}_1-\gamma_1\end{pmatrix}-\frac{1}{\sqrt{NT}}\mathbf{b}_c^{(\cdot)}\right], \tag{32}$$

where $\mathbf{b}_c^{(\cdot)}$ is the asymptotic bias term, ψ_{11} and ψ_{22} are the (1,1)-th element and the (2,2)-th element of $\boldsymbol{\Psi}^*$, respectively. We have chosen these standardizations because

[1] We have used Example 1 in Akashi and Kunitomo (2012) to investigate Case (a) in more details. Example 1 can be regarded as a special case of Example 2.

of the form of the limiting distribution of the LIML estimator in Theorems 1 and 2. We may call the large-r_n case when $c = 0$ ($c = c_a$ or c_b) and $c \neq 0$ as the general case.

Since Akashi and Kunitomo (2012) have given many figures on Case (a), we only give Figs. 5 and 6 in Appendix. We have shown the estimated distribution functions of the GMM and the LIML estimators of (β_2, γ_1) and we confirm the findings of Akashi and Kunitomo (2012) in a simple case. The GMM estimator is badly biased when N and T are large while the LIML estimator is almost median-unbiased after correcting the bias term in (32). However, the normalization by the limiting covariance matrix of the LIML estimator when $c = 0$ is not appropriate. This aspect can be observed because the circles in figures are the standard normal distribution function N(0,1).

For Case (b) with the backward-filtering procedure, we show the estimated distribution functions of the GMM and LIML estimators of β_2 and γ_1 as Figs. 1, 2, 3 and 4 among many results. From these figures, we first observe that the GMM estimator is often biased when N and T are large while the LIML estimator is almost median-unbiased. Second, the normalization by the limiting covariance matrix of the LIML estimator when $c = 0$ is often not appropriate. Since the normal approximations based on the general case $c \neq 0$, it is important to use the formulas in Sect. 3.

5 Conclusions

In this paper, we develop the panel-limited information maximum likelihood (PLIML) approach for estimating dynamic panel simultaneous equation models. When there are dynamic effects and endogenous variables with individual effects at the same time, the LIML method for the filtered data does give not only the consistency and the asymptotic normality, but also attains the asymptotic efficiency bound when the order of orthogonal conditions is large or many instruments.

The consistency of LIML method does not depend on specified panel asymptotics and the total number of instruments as long as it is less than the total number of observations. The approximation of its limiting distribution embodies the influence of the number of instrumental variables automatically and our method gives an unified approach for solving the statistical problem with panel data when N and T are large.

We have examined the effects of alternative filtering procedures. When we only apply the forward-filtering and use $O(T^2)$ instruments, the GMM estimator is badly biased while the LIML estimator can be almost-unbiased. If we use the backward-filtering to instruments, the GMM estimator is often biased, but its magnitude is significantly reduced. Since we often do not know the precise form of lag structures of the reduced form in the panel simultaneous equation models, we conclude that the LIML method has the asymptotic robustness in both cases of (a) and (b) while the GMM does not have such robustness.

In a companion paper, Akashi and Kunitomo (2012) have investigated the finite sample properties of alternative estimation methods such as the WG (Within Groups), the GMM and the PLIML estimators in a simpler setting. Although they have used a particular case of dynamic panel models and the forward filtering procedure, their results are relevant for more general panel structural equations. In this sense, the LIML method is quite useful and relevant in dynamic panel econometric modeling.

Finally, there is an important issue on the individual heteroscedasticity in panel structural equation models. An optimal modification of the LIML method to remove the possible bias due to individual heteroscedasticity can be developed. The details have been discussed in Kunitomo and Akashi (2010), which was along the arguments originally developed by Kunitomo (2012).

6 Mathematical details

In this section, we give the proofs of Theorems 1 and 2. The method of proofs are similar to those used in Akashi and Kunitomo (2012). When we use notations $(\mathbf{M}_t, \mathbf{N}_t, c, c_*)$, the relevant derivations are valid for each case of $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$ under the corresponding asymptotics. (We shall use c and c_* $(= c/(1-c))$ for c_a, c_b (, which are different from c_t), and c_{*a} $(= c_a/(1-c_a)), c_{*b}$ $(= c_b/(1-c_b))$ without any confusion.) Also we use notations $\mathbf{J}$ for $\mathbf{J}_K$ and $\mathbf{z}_{it} = \mathbf{J}\mathbf{w}_{i,t-1}$, which means that the relevant information is in the first K-variables for the sake of convenience. Because some derivations have been given in Akashi and Kunitomo (2012) when $G_2 = 1$ and it is often straight-forward to extend their analysis to the general case, we refer to their results in such cases.

Derivations of Theorems 1 and 2

Step 1: First, we shall investigate the effects of the forward-filtering. We drive the probability limits of sample quantities and obtain the representations for the LIML estimator. Substitution of (7) into (14) yields

$$\mathbf{G}^{(f)} = \mathbf{G}^{(f,1)} + \mathbf{G}^{(f,2)} + \mathbf{G}^{(f,2)'} + \mathbf{G}^{(f,3)}, \tag{33}$$

where $\mathbf{G}^{(f,1)} = \mathbf{D}^{*'} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{M}_t \mathbf{Z}_{t-1}^{(f)} \mathbf{D}^*$, $\mathbf{G}^{(f,2)} = \mathbf{D}^{*'} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{M}_t (\mathbf{V}_t^{(f)}, \mathbf{O})$, $\mathbf{G}^{(f,3)} = \sum_{t=1}^{T-1} (\mathbf{V}_t^{(f)}, \mathbf{O})' \mathbf{M}_t (\mathbf{V}_t^{(f)}, \mathbf{O})$, $\mathbf{V}_t^{(f)'} = (\mathbf{v}_{1t}^{(f)}, \ldots, \mathbf{v}_{Nt}^{(f)})$ and $\mathbf{v}_{it}^{(f)}$ $(i = 1, \ldots, N)$ are the corresponding forward-filtered disturbances of $\mathbf{v}_{it}$, and a $K \times (1 + G_1 + K_1)$ matrix $\mathbf{D}^* = \mathbf{D}[\boldsymbol{\theta}, \mathbf{I}_{G_2+K_1}]$.

We shall show that for $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ or $\mathbf{M}_t^{(b)}$,

$$\frac{1}{n}\mathbf{G}^{(f)} \xrightarrow{p} \mathbf{G}_0 = \begin{bmatrix} \boldsymbol{\theta}' \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \boldsymbol{\Phi}^* \left[\boldsymbol{\theta}, \mathbf{I}_{G_2+K_1} \right] + c \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \tag{34}$$

and

$$\frac{1}{q_n}\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0 = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}, \tag{35}$$

where $\boldsymbol{\Phi}^* = \mathbf{D}'\mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J}\mathbf{D} = \mathbf{D}'\mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J}\mathbf{D}$ and $n = NT$. Using the representation

$$\begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} = \mathbf{D}^{*'}\mathbf{Z}_{t-1}^{(f)'} + \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O}' \end{pmatrix} \text{ (say,)}$$

we show that $(1/n)\mathbf{G}^{(f,2)} \xrightarrow{p} \mathbf{O}_{G_2+K_1}$ using the same argument for $(1/\sqrt{n})\sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{u}_t^{(f)}$ in Akashi and Kunitomo (2012). We write

$$\mathbf{Z}_{t-1}^{(f)'} = \mathbf{J}\left(c_t\left[\mathbf{I}_{G_*} - \frac{1}{T-t}\left(\sum_{j=1}^{T-t}\boldsymbol{\Pi}^{*j}\right)\right]\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}'\right) = \boldsymbol{\Psi}_t\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}',$$

where $\tilde{\mathbf{V}}_{tT}'$ is defined in Step 3 below. We further decompose $(1/n)\mathbf{G}^{(f,1)}$ as

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{Z}_{t-1}^{(f)} &= \frac{1}{n}\sum_{t=1}^{T-1}\boldsymbol{\Psi}_t\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1}\boldsymbol{\Psi}_t' - \frac{1}{n}\sum_{t=1}^{T-1}c_t\boldsymbol{\Psi}_t\mathbf{W}_{t-1}'\mathbf{M}_t\tilde{\mathbf{V}}_{tT} \\ &\quad -\frac{1}{n}\sum_{t=1}^{T-1}c_t\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t\mathbf{W}_{t-1}\boldsymbol{\Psi}_t' + \frac{1}{n}\sum_{t=1}^{T-1}c_t^2\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t\tilde{\mathbf{V}}_{tT}. \qquad (36)\end{aligned}$$

Moreover, using Lemmas 2 and 3 in Steps 4 and 5, and $c_t^2 = 1 - 1/(T-t+1)$ after some calculations, it is possible to show

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{T-1}\boldsymbol{\Psi}_t\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1}\boldsymbol{\Psi}_t' \\ &= \frac{1}{n}\sum_{t=1}^{T-1}c_t^2\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1} \\ &\quad -\frac{1}{n}\sum_{t=1}^{T-1}\frac{c_t^2}{T-t}\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1}\left(\sum_{j=1}^{T-t}\boldsymbol{\Pi}^{*j}\right)' - \frac{1}{n}\sum_{t=1}^{T-1}\frac{c_t^2}{T-t}\left(\sum_{j=1}^{T-t}\boldsymbol{\Pi}^{*j}\right)\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1} \\ &\quad +\frac{1}{n}\sum_{t=1}^{T-1}\left(\frac{c_t}{T-t}\right)^2\left(\sum_{j=1}^{T-t}\boldsymbol{\Pi}^{*j}\right)\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1}\left(\sum_{j=1}^{T-t}\boldsymbol{\Pi}^{*j}\right)' \qquad (37)\end{aligned}$$

converges to $\mathcal{E}[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}']$ in probability. The second and third terms of (36) have zero means and their variances to tend to zeros. It is because

$$\mathrm{Var}\left[\frac{1}{NT}\sum_{t=1}^{T-1} c_t \mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k\right]$$

$$\leq \frac{1}{N^2T^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sqrt{c_t^2 \mathcal{E}\left[(\mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k)^2\right]}\sqrt{c_s^2 \mathcal{E}\left[(\mathbf{e}_k^{'} \tilde{\mathbf{V}}_{sT} \mathbf{M}_s \mathbf{W}_{s-1} \boldsymbol{\Psi}_s^{'} \mathbf{e}_j)^2\right]},$$

where $\mathbf{e}_j$ $(j, k = 1, \ldots, K)$ are j-th unit vector. Also we have

$$c_t^2 \mathcal{E}\left[(\mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k)^2\right] = c_t^2 \left[\mathbf{e}_k^{'} \mathcal{E}\left[\tilde{\mathbf{v}}_{itT} \tilde{\mathbf{v}}_{itT}^{'}\right] \mathbf{e}_k\right] \mathcal{E}\left[\mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} \boldsymbol{\Psi}_t^{'} \mathbf{e}_j\right]$$

$$\leq c_t^2 \left[\frac{1}{(T-t)^2} \mathbf{e}_k^{'} \sum_{h=1}^{T-t} \boldsymbol{\Phi}_h \mathcal{E}\left[\mathbf{v}_{i0}^{*} \mathbf{v}_{i0}^{*'}\right] \boldsymbol{\Phi}_h^{'} \mathbf{e}_k\right]$$

$$\times \left[\mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathcal{E}[\mathbf{W}_{t-1}^{'} \mathbf{W}_{t-1}] \boldsymbol{\Psi}_t^{'} \mathbf{e}_j\right],$$

which is $O(N/(T-t))$ because $\sum_{h=1}^{T-t} \mathbf{e}_k^{'} \boldsymbol{\Phi}_h \mathcal{E}[\mathbf{v}_{i0}^{*} \mathbf{v}_{i0}^{*'}] \boldsymbol{\Phi}_h^{'} \mathbf{e}_k = O(T-t)$. Then

$$\mathrm{Var}\left[\frac{1}{N_0}\sum_{t=1}^{T-1} c_t \mathbf{e}_j^{'} \boldsymbol{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k\right] = O\left(\frac{(\sqrt{T})^2}{N_0 T^2}\right).$$

For the fourth term of (36), the expected value is given by

$$\mathcal{E}\left[\frac{1}{n}\sum_{t=1}^{T-1} c_t^2 \mathbf{e}_j^{'} \tilde{\mathbf{V}}_{tT}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k\right] = \frac{1}{n}\sum_{t=1}^{T-1} c_t^2 \mathrm{tr}(\mathbf{M}_t) \mathcal{E}\left[\mathbf{e}_j^{'} \tilde{\mathbf{v}}_{itT} \tilde{\mathbf{v}}_{itT}^{'} \mathbf{e}_k\right]$$

$$= O\left(\frac{1}{n}\sum_t \frac{\mathrm{tr}(\mathbf{M}_t)}{T-t+1}\right)$$

and it converges to zero in probability. Its variance tends to zero in the same way as for $\Upsilon_{21n}^{(k)}$ and $\Upsilon_{22n}^{(k)}$ in Step 3 below.

Next, we consider $(1/n)\mathbf{G}^{(f,3)}$. Using the fact that $\mathcal{E}_t[\mathbf{v}_{it}^{(f)} \mathbf{v}_{it}^{(f)'}] = \boldsymbol{\Omega}$,

$$\mathcal{E}\left[\frac{1}{n}\sum_{t=1}^{T-1} \mathbf{e}_g^{'} \mathbf{V}_t^{(f)'} \mathbf{M}_t \mathbf{V}_t^{(f)} \mathbf{e}_h\right] = \frac{\mathbf{e}_g^{'} \boldsymbol{\Omega} \mathbf{e}_h}{n}\sum_{t=1}^{T-1} \mathrm{tr}(\mathbf{M}_t),$$

which converges to $c(\mathbf{e}_g^{'} \boldsymbol{\Omega} \mathbf{e}_h)$ as $n \to \infty$. Moreover, using $\mathbf{V}_t^{(f)} = (\mathbf{V}_t - \bar{\mathbf{V}}_{tT})/c_t$, we decompose

$$\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{V}_t^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)} = \frac{1}{n}\sum_{t=1}^{T-1}c_t^{-2}\mathbf{V}_t'\mathbf{M}_t\mathbf{V}_t - \frac{1}{n}\sum_{t=1}^{T-1}c_t^{-2}\mathbf{V}_t'\mathbf{M}_t\bar{\mathbf{V}}_{tT}$$
$$-\frac{1}{n}\sum_{t=1}^{T-1}c_t^{-2}\bar{\mathbf{V}}_{tT}'\mathbf{M}_t\mathbf{V}_t + \frac{1}{n}\sum_{t=1}^{T-1}c_t^{-2}\bar{\mathbf{V}}_t'\mathbf{M}_t\bar{\mathbf{V}}_{tT}. \qquad (38)$$

Because of Lemma 1 of Step 3 below, $\mathrm{Var}[\mathbf{v}_t^{(g)'}\mathbf{M}_t\mathbf{v}_t^{(h)}] = O(t)$ and $\mathrm{Cov}[\mathbf{v}_t^{(g)'}\mathbf{M}_t\mathbf{v}_t^{(h)}, \mathbf{v}_s^{(g)'}\mathbf{M}_t\mathbf{v}_s^{(h)}] = 0$ for $t \neq s$. Hence the variance of the first term satisfies

$$\mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{V}_t^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)}\mathbf{e}_h\right] = \frac{1}{n^2}\sum_{t=1}^{T-1}\left(1+\frac{1}{T-t}\right)^2 \times O(t),$$

which converges to zero.

The second and third terms of the right-hand side of (38) can be evaluated analogously as $\Upsilon_{21n}^{(k)}$ and $\Upsilon_{22n}^{(k)}$, and their variances tend to zeros using the similar arguments.

We turn to show that $(1/q_n)\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0$ by evaluating

$$\frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{pmatrix}(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)})$$
$$= \mathbf{D}^{*'}\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{Z}_{t-1}^{(f)}\mathbf{D}^* + \mathbf{D}^{*'}\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}(\mathbf{V}_t^{(f)}, \mathbf{O})$$
$$+\frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)'}\\ \mathbf{O}'\end{pmatrix}\mathbf{Z}_{t-1}^{(f)}\mathbf{D}^* + \frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)'}\\ \mathbf{O}'\end{pmatrix}(\mathbf{V}_t^{(f)}, \mathbf{O}). \qquad (39)$$

The expected values of the second and third terms of $1/(N_0T)\sum_t \mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{V}_t^{(f)}] = (1/T)(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*\mathbf{v}_{it}^{*'}] + O(1/T)$ converge to zeros as $T \to \infty$. We can establish the mean squared convergence similarly. Moreover,

$$\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{e}_j'\mathbf{Z}_{t-1}^{(f)'}\mathbf{Z}_{t-1}^{(f)}\mathbf{e}_k = \frac{1}{N_0T}\sum_{i=1}^{N_0}\sum_{t=1}^{T-1}w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)} - \frac{1}{N_0T}\sum_{i=1}^{N_0}\frac{1}{T}\boldsymbol{\iota}_T'\mathbf{w}_{i(t-1)}^{(j)}\mathbf{w}_{i(t-1)}^{(k)'}\boldsymbol{\iota}_T$$

converges to $\mathcal{E}[w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)}]$ in probability since $(1/T)\sum_{t=1}^{T-1}w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)} \xrightarrow{p} \mathcal{E}[w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)}]$ and the second term converges to $(1/N_0)\sum_{i=1}^{N_0}(0+o_p(1))^2 = o_p(1)$ using that $(1/T)\boldsymbol{\iota}_T'\mathbf{w}_{i(t-1)}^{(j)} \xrightarrow{p} 0$. Again using the similar argument, we have that $(1/n)\sum_{t=1}^{T-1}\mathbf{V}_t^{(f)'}\mathbf{V}_t^{(f)} \xrightarrow{p} \boldsymbol{\Omega}$. Hence

$$\frac{1}{q_n}\mathbf{H}^{(f)} \xrightarrow{p} \frac{1}{1-c}\left[plim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)\prime}\\ \mathbf{Z}_{t-1}^{(1,f)\prime}\end{pmatrix}(\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)})-\mathbf{G}_0\right]=\mathbf{H}_0.$$

Step 2: Using the convergence results in Step 1, we have

$$\left|\boldsymbol{\Phi}_\theta+[c-(plim_{n\to\infty}\lambda_n)]\begin{bmatrix}\boldsymbol{\Omega} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\right|=0, \tag{40}$$

where

$$\boldsymbol{\Phi}_\theta=\begin{bmatrix}\boldsymbol{\theta}'\\ \mathbf{I}_{G_2+K_1}\end{bmatrix}\boldsymbol{\Phi}^*\left[\boldsymbol{\theta},\mathbf{I}_{G_2+K_1}\right]. \tag{41}$$

By the assumption that $\boldsymbol{\Phi}^*$ is a positive definite matrix, $\lambda_n \xrightarrow{p} c$ and we have that $\hat{\boldsymbol{\theta}}_{\text{LI}} \xrightarrow{p} \boldsymbol{\theta}$ because (16) gives $\boldsymbol{\Phi}^*(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})=o_p(1)$. Define $\mathbf{G}_1^{(f)}=\sqrt{n}[(1/n)\mathbf{G}^{(f)}-\mathbf{G}_0]$, $\mathbf{H}_1^{(f)}=\sqrt{q_n}[(1/q_n)\mathbf{H}^{(f)}-\mathbf{H}_0]$, $\lambda_{1n}^{(f)}=\sqrt{n}[\lambda_n-c]$ and $\mathbf{b}_1=\sqrt{n}[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}]$. By substituting these variables into (16), we find

$$[\mathbf{G}_0-c\mathbf{H}_0]\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}+\frac{1}{\sqrt{n}}\left[\mathbf{G}_1^{(f)}-\lambda_{1n}^{(f)}\mathbf{H}_0\right]\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}+\frac{1}{\sqrt{n}}[\mathbf{G}_0-c\mathbf{H}_0]\,\mathbf{b}_1$$
$$-\frac{1}{\sqrt{q_n}}[c\mathbf{H}_1^{(f)}]\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}=o_p\left(\frac{1}{\sqrt{n}}\right). \tag{42}$$

Then using the relation of $\boldsymbol{\Phi}_\theta(1,-\boldsymbol{\theta}')'=\mathbf{0}$, we have

$$\boldsymbol{\Phi}_\theta\begin{bmatrix}0\\ \mathbf{b}_1\end{bmatrix}=\left[\mathbf{G}_1^{(f)}-\lambda_{1n}^{(f)}\mathbf{H}_0-\sqrt{cc_*}\mathbf{H}_1^{(f)}\right]\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}+o_p(1).$$

Multiplication of (42) from the left by $(1,-\boldsymbol{\theta})$ yields

$$\lambda_{1n}^{(f)}=\frac{(1,-\boldsymbol{\theta}')\left[\mathbf{G}_1^{(f)}-\sqrt{cc_*}\mathbf{H}_1^{(f)}\right](1,-\boldsymbol{\theta}')'}{(1,-\boldsymbol{\theta}')\mathbf{H}_0(1,-\boldsymbol{\theta}')'}+o_p(1). \tag{43}$$

Also the multiplication of (43) from the left by $(0,\mathbf{I}_{G_2+K_1})$ and substitution for $\lambda_{1n}^{(f)}$ for (43) yields

$$\boldsymbol{\Phi}^*\sqrt{n}\begin{bmatrix}\hat{\boldsymbol{\beta}}_{2LI}-\boldsymbol{\beta}_2\\ \hat{\boldsymbol{\gamma}}_{1LI}-\boldsymbol{\gamma}_1\end{bmatrix}$$
$$=[\mathbf{0},\mathbf{I}_{G_2+K_1}]\left[\mathbf{I}_{1+G_2+K_1}-\frac{1}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\begin{pmatrix}\boldsymbol{\Omega}\boldsymbol{\beta}\\ \mathbf{0}\end{pmatrix}(1,-\boldsymbol{\theta}')\right]\left[\mathbf{G}_1^{(f)}-\sqrt{cc_*}\mathbf{H}_1^{(f)}\right]\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}$$
$$+o_p(1). \tag{44}$$

Using the relation of (33), we have

$$\left[\mathbf{G}_1^{(f)} - \sqrt{cc_*}\mathbf{H}_1^{(f)}\right]\begin{bmatrix}1\\-\boldsymbol{\theta}\end{bmatrix}$$
$$= \frac{1}{\sqrt{n}}\mathbf{D}^{*\prime}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}\mathbf{M}_t\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\left[\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)\prime}\\\mathbf{O}\end{pmatrix}\mathbf{M}_t\mathbf{u}_t^{(f)} - r_n\begin{pmatrix}\boldsymbol{\Omega\beta}\\\mathbf{0}\end{pmatrix}\right]$$
$$-\frac{\sqrt{cc_*}}{\sqrt{q_n}}\mathbf{D}^{*\prime}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}\left[\mathbf{I}_N - \mathbf{M}_t\right]\mathbf{u}_t^{(f)}$$
$$-\frac{\sqrt{cc_*}}{\sqrt{q_n}}\left[\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)\prime}\\\mathbf{O}\end{pmatrix}[\mathbf{I}_N - \mathbf{M}_t]\mathbf{u}_t^{(f)} - q_n\begin{pmatrix}\boldsymbol{\Omega\beta}\\\mathbf{0}\end{pmatrix}\right]. \tag{45}$$

Also we use the relations $\sqrt{cc_*}/\sqrt{q_n} - c_*/\sqrt{n} = o(1)$, $[\mathbf{I}_{1+G_2} - (1/\sigma^2)\boldsymbol{\Omega\beta\beta}']\boldsymbol{\Omega\beta} = \mathbf{0}$ and then

$$\boldsymbol{\Phi}^*\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{2LI} - \boldsymbol{\beta}_2\\\hat{\boldsymbol{\gamma}}_{1LI} - \boldsymbol{\gamma}_1\end{pmatrix} = \frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}\mathbf{N}_t\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{U}_t^{(\perp,f)\prime}\\\mathbf{O}\end{pmatrix}\mathbf{N}_t\mathbf{u}_t^{(f)} + o_p(1). \tag{46}$$

where $\mathbf{N}_t = \mathbf{M}_t - c_*(\mathbf{I}_N - \mathbf{M}_t) = \frac{1}{1-c}[\mathbf{M}_t - c\mathbf{I}_N]$ and

$$\mathbf{U}_t^{(\perp,f)\prime} = \left[\mathbf{0}, \mathbf{I}_{G_2}\right]\left[\mathbf{I}_{1+G_2} - \frac{\boldsymbol{\Omega\beta\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega\beta}}\right]\mathbf{V}_t^{(f)\prime} = \left(\mathbf{u}_{1t}^{(\perp,f)}, \ldots, \mathbf{u}_{Nt}^{(\perp,f)}\right).$$

Step 3: We evaluate the additional effects of the forward-filtering on the LIML estimation at this step by setting $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and the k-th unit vector as $\mathbf{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)'$. Using (A2) and $\mathbf{u}_t^{(f)} = (\mathbf{u}_t - \mathbf{u}_{tT})/c_t$, we decompose the first and second terms of (46) as, for $k = 1, \ldots, K (= K_1 + K_2)$ and $g = 1, \ldots, G_2$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{Z}_{t-1}^{(f)\prime}\mathbf{N}_t^{(a)}\mathbf{u}_t^{(f)}$$
$$= \frac{1}{1-c_a}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{u}_t - \Upsilon_{11n}^{(k,a)} - \Upsilon_{12n}^{(k,a)}\right) - \left(\Upsilon_{21n}^{(k,a)} - \Upsilon_{22n}^{(k,a)}\right)\right]$$
$$-c_{*a}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{u}_t - \Upsilon_{3n}^{(k)}\right), \tag{47}$$

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{U}_t^{(\perp,f)'}\mathbf{N}_t^{(a)}\mathbf{u}_t^{(f)}\\
&=\frac{1}{1-c_a}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\left(\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{M}_t^{(a)}\mathbf{u}_t+\left(\frac{1}{T-t}\right)\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{M}_t^{(a)}\mathbf{u}_t\right.\right.\\
&\quad\left.\left.-c_t^{-2}\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}-c_t^{-2}\mathbf{e}_g'\bar{\mathbf{U}}_{tT}^{\perp'}\mathbf{M}_t^{(a)}\mathbf{u}_t+c_t^{-2}\mathbf{e}_g'\bar{\mathbf{U}}_{tT}^{\perp'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\right)\right]\\
&\quad-c_{*a}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{u}_t-\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\mathbf{e}_g'\bar{\mathbf{u}}_i^{\perp}\bar{u}_i\right)\\
&=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{N}_t^{(a)}\mathbf{u}_t-\Upsilon_{4n}^{(g,a)},
\end{aligned}
\tag{48}
$$

where

$$
\begin{aligned}
\Upsilon_{11n}^{(k,a)}&=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT},\ \Upsilon_{12n}^{(k,a)}=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\frac{c_t}{T-t}\mathbf{e}_k'\mathbf{J}'\tilde{\mathbf{W}}_{t-1}'\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)},\\
\Upsilon_{21n}^{(k,a)}&=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{J}'\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^{(a)}\mathbf{u}_t,\ \Upsilon_{22n}^{(k,a)}=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_k'\mathbf{J}'\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT},\\
\Upsilon_{3n}^{(k)}&=\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\mathbf{e}_k'\mathbf{J}'\bar{\mathbf{w}}_{i(-1)}\bar{u}_i,\\
\Upsilon_{4n}^{(g,a)}&=\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{U}_t^{(\perp,f)'}\mathbf{N}_t^{(a)}\mathbf{u}_t^{(f)}-\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{U}_t^{\perp'}\mathbf{N}_t^{(a)}\mathbf{u}_t,
\end{aligned}
$$

and we use notations: $\bar{\mathbf{u}}_{tT}=(\mathbf{u}_t+\cdots+\mathbf{u}_T)/(T-t+1)$, $\mathbf{u}_t'=(u_{1t},\ldots,u_{Nt})$, $\tilde{\mathbf{W}}_{t-1}'=(\sum_{h=1}^{T-t}\boldsymbol{\Pi}^{*h})\mathbf{W}_{t-1}'$, $\tilde{\mathbf{V}}_{tT}'=\frac{1}{T-t}\sum_{h=1}^{T-t}\boldsymbol{\Phi}_h\mathbf{V}_{T-h}^{*'}$, $\mathbf{V}_h^{*'}=(\mathbf{v}_{1h}^*,\ldots,\mathbf{v}_{Nh}^*)=(\mathbf{v}_h^{*(1)},\ldots,\mathbf{v}_h^{*(K)})'$, $\boldsymbol{\Phi}_h=(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^{*h})$, $\bar{\mathbf{w}}_{i(-1)}=\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{w}_{i(t-1)}$, $\bar{u}_i=\frac{1}{T}\sum_{t=1}^{T-1}u_{it}$, $\bar{\mathbf{U}}_t^{\perp}=(\mathbf{U}_t^{\perp}+\cdots+\mathbf{U}_T^{\perp})/(T-t+1)$, $\bar{\mathbf{u}}_i^{\perp}=\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{u}_{it}^{\perp}$ and

$$
\mathbf{U}_t^{\perp'}=\left[\mathbf{0},\mathbf{I}_{G_2}\right]\left[\mathbf{I}_{1+G_2}-\frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\right]\mathbf{V}_t'=(\mathbf{u}_{1t}^{\perp},\ldots,\mathbf{u}_{Nt}^{\perp}).
$$

We shall show that some of variances go to zeros. The variances of each terms $\Upsilon_{4n}^{(g,a)}$ (and the corresponding $\Upsilon_{4n}^{(g,b)}$) can be evaluated using the same argument of Akashi and Kunitomo (2012) and we give the *Key Lemma* below (the proof is in Akashi and Kunitomo (2012)).

Lemma 1 *Let* $\mathbf{d}_t$ *and* $\mathbf{d}_s$ *be* $N \times 1$ *vectors containing the diagonal elements of* $\mathbf{M}_t$ *and* $\mathbf{M}_s$*, respectively, such that* $\mathrm{tr}(\mathbf{M}_t) = \mathbf{d}_t'\boldsymbol{\iota}_N$, $\mathrm{tr}(\mathbf{M}_s) = \mathbf{d}_s'\boldsymbol{\iota}_N$, $\mathbf{d}_t'\mathbf{d}_s \leq \max\{\mathrm{tr}(\mathbf{M}_t), \mathrm{tr}(\mathbf{M}_s)\}$ *and* $\mathrm{tr}(\mathbf{M}_t\mathbf{M}_s) \leq \max\{\mathrm{tr}(\mathbf{M}_t), \mathrm{tr}(\mathbf{M}_s)\}$*. Then, for* $l \geq r \geq t$, $p \geq q \geq s, t \geq s$,

$$\mathrm{Cov}\left[\boldsymbol{\epsilon}_l^{*'}\mathbf{M}_t\boldsymbol{\epsilon}_r^{**}, \boldsymbol{\epsilon}_p^{*'}\mathbf{M}_s\boldsymbol{\epsilon}_q^{**}\right]$$

$$= \begin{cases} (m^{(3)} + m^{(2)})\mathrm{tr}(\mathbf{M}_t\mathbf{M}_s) + m^{(0)}\mathcal{E}[\mathbf{d}_t'\mathbf{d}_s] & \text{if } l = r = p = q, \\ \mathcal{E}\left[\epsilon_{it}^{*2}\epsilon_{it}^{**}\right]\mathcal{E}[\mathbf{d}_t'\mathbf{M}_s\boldsymbol{\epsilon}_q^{**}] & \text{if } l = r = p \neq q < t, \\ m^{(3)}\mathrm{tr}(\mathbf{M}_t\mathbf{M}_s) & \text{if } l = p \neq r = q, \\ 0 & \text{otherwise,} \end{cases} \tag{49}$$

where $|\mathcal{E}\left[\mathbf{d}_t'\mathbf{M}_s\boldsymbol{\epsilon}_q^{**}\right]| \leq (\mathrm{tr}(\mathbf{M}_t)\mathrm{tr}(\mathbf{M}_s)\mathcal{E}\left[\epsilon_{it}^{**2}\right])^{1/2}$,

$$m^{(1)} = m^{(1)}(\boldsymbol{\epsilon}_t^*, \boldsymbol{\epsilon}_t^{**}) = \mathcal{E}\left[\boldsymbol{\epsilon}_{it}^{*2}\boldsymbol{\epsilon}_{it}^{**2}\right], m^{(2)} = m^{(2)}(\boldsymbol{\epsilon}_t^*, \boldsymbol{\epsilon}_t^{**}) = (\mathcal{E}\left[\boldsymbol{\epsilon}_{it}^*\boldsymbol{\epsilon}_{it}^{**}\right])^2,$$

$$m^{(3)} = m^{(3)}(\boldsymbol{\epsilon}_t^*, \boldsymbol{\epsilon}_t^{**}) = \mathcal{E}\left[\boldsymbol{\epsilon}_{it}^{*2}\right]\mathcal{E}\left[\boldsymbol{\epsilon}_{it}^{**2}\right], m^{(0)} = m^{(0)}(\boldsymbol{\epsilon}_t^*, \boldsymbol{\epsilon}_t^{**}) = m^{(1)} - 2m^{(2)} - m^{(3)}.$$

Now we go back to the original derivation. First, it it straightforward to show that $\mathrm{Var}[\Upsilon_{3n}^{(k)}] \to 0$ as $T \to \infty$ by the similar argument as used for Alvarez and Arellano (2003). Second, we have

$$\mathrm{Var}[\Upsilon_{11n}^{(k,a)}] = \frac{1}{n}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathcal{E}\left[\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(a)}\mathbf{W}_{t-1}\mathbf{J}\mathbf{e}_k\right].$$

For $t \geq s$,

$$\mathcal{E}\left[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}\right] = \frac{\sigma^2}{(T-s+1)}\mathcal{E}\left[\mathcal{E}_s\left[\mathbf{w}_{t-1}^{(k)'}\right]\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}\right]$$

$$= \frac{\sigma^2}{(T-s+1)}\mathcal{E}\left[\sum_{j=1}^{G_*}\left(\mathbf{e}_{kJ}'\boldsymbol{\Pi}^{*t-s}\mathbf{e}_j\right)\mathbf{w}_{s-1}^{(j)'}\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}\right],$$

where $\mathbf{w}_{t-1}^{(k)'} = \mathbf{e}_{kJ}'\mathbf{W}_{t-1}'$, $\mathbf{w}_{t-1}^{(j)'} = \mathbf{e}_j'\mathbf{W}_{t-1}'$ and $\mathbf{e}_{kJ}' = \mathbf{e}_k'\mathbf{J}'$, which is an unit k-th vector. The second equality is due to the fact that $\mathbf{M}_t^{(a)}\mathbf{M}_s^{(a)} = \mathbf{M}_s^{(a)}$. Using the relation that for any s, j, k, $|\mathcal{E}[\mathbf{w}_{s-1}^{(j)'}\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}]| \leq (\mathcal{E}[(\mathbf{w}_0^{(j)'}\mathbf{w}_0^{(j)})(\mathbf{w}_0^{(k)'}\mathbf{w}_0^{(k)})])^{1/2}$ and $(\mathcal{E}[(\mathbf{w}_0^{(j)'}\mathbf{w}_0^{(j)})(\mathbf{w}_0^{(k)'}\mathbf{w}_0^{(k)})])^{1/2} = O(N)$, we can evaluate

$$\mathrm{Var}[\Upsilon_{11n}^{(k,a)}] \leq O(1) \times \frac{1}{T}\left[\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\frac{1}{T-s+1}\sum_{j=1}^{G_*}|\mathbf{e}_{kJ}'\boldsymbol{\Pi}^{*|t-s|}\mathbf{e}_j|\right]$$

$$= \frac{O(1)}{T}\left[\left(\frac{1}{T}+\cdots+\frac{1}{2}\right)+2\sum_{j=1}^{G_*}S_T^{(k,j)}\right],$$

which is $O((\log T)/T)$ because

$$S_T^{(k,j)} = \frac{1}{T}\left(|\mathbf{e}_{kJ}'\boldsymbol{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\boldsymbol{\Pi}^{*T-2}\mathbf{e}_j|\right)$$

$$+\frac{1}{T-1}\left(|\mathbf{e}_{kJ}'\boldsymbol{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\boldsymbol{\Pi}^{*T-3}\mathbf{e}_j|\right) + \cdots + \frac{1}{3}|\mathbf{e}_{kJ}'\boldsymbol{\Pi}^*\mathbf{e}_j|$$

$$\leq \left(\frac{1}{3}+\cdots+\frac{1}{T}\right)\left(|\mathbf{e}_{kJ}'\boldsymbol{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\boldsymbol{\Pi}^{*T-2}\mathbf{e}_j|\right) = O(\log T).$$

Next,

$$\mathrm{Var}[\Upsilon_{12n}^{(k,a)}] = \frac{\sigma^2}{n}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}\left[\tilde{\mathbf{w}}_{t-1}^{(k)\prime}\mathbf{M}_t^{(a)}\tilde{\mathbf{w}}_{t-1}^{(k)}\right]$$

$$\leq \frac{\sigma^2}{n}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}\left[\tilde{\mathbf{w}}_{t-1}^{(k)\prime}\tilde{\mathbf{w}}_{t-1}^{(k)}\right]$$

$$\leq \frac{\sigma^2\lambda_{\max}\{\mathcal{E}\left[\mathbf{w}_{i0}\mathbf{w}_{i0}'\right]\}}{T}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}\mathbf{e}_k'\mathbf{J}'\left(\sum_{h=1}^{T-t}\boldsymbol{\Pi}^{*h}\right)\left(\sum_{h=1}^{T-t}\boldsymbol{\Pi}^{*h}\right)'\mathbf{J}\mathbf{e}_k,$$

which is $O(1/T)$, where $\tilde{\mathbf{w}}_{t-1}^{(k)\prime} = \mathbf{e}_k'\mathbf{J}'\tilde{\mathbf{W}}_{t-1}'$ and λ_{max} stands for the largest eigenvalue. The last inequality follows from the fact that $c_t^2 < 1$ and $\sum_{h=1}^{T-t}\boldsymbol{\Pi}^{*h}$ is bounded for any t, T $(T-t>0)$.

Turning to evaluate the variance of $\Upsilon_{21n}^{(k,a)}$, in view of Lemma 1 the only non-zero terms to be considered are given by the quantities $a_{0n}^{(k,j,a)}$ and $a_{1n}^{(k,j,a)}$ $(j = 1, \ldots, K^*)$ which, are represented as

$$\mathrm{Var}[\Upsilon_{21n}^{(k,a)}] = \frac{1}{n}\mathrm{Var}\left[\sum_{t=1}^{T-1}\frac{1}{T-t}\sum_{h=1}^{T-t}\sum_{j=1}^{G_*}(\mathbf{e}_{kJ}'\boldsymbol{\Phi}_h\mathbf{e}_j)\mathbf{e}_j'\mathbf{V}_{T-h}^{*\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t\right]$$

$$= \frac{1}{n}\left[\sum_{j=1}^{G_*}\mathrm{Var}\left[\sum_{t=1}^{T-1}\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t\right]\right.$$

$$+\sum_{i,j}^{G_*} \mathrm{Cov}\left[\sum_{t=1}^{T-1} \tilde{\mathbf{v}}_{tT}^{*(k,i)'} \mathbf{M}_t^{(a)} \mathbf{u}_t, \sum_{t=1}^{T-1} \tilde{\mathbf{v}}_{tT}^{*(k,j)'} \mathbf{M}_t^{(a)} \mathbf{u}_t\right]\Bigg]$$

$$= \sum_{j=1}^{G_*} (a_{0n}^{(k,j,a)} + a_{1n}^{(k,j,a)}) + \frac{1}{n} \sum_{i,j}^{G_*} \mathrm{Cov}[.,.], \tag{50}$$

where

$$\tilde{\mathbf{v}}_{tT}^{*(k,j)'} = \frac{1}{T-t} \sum_{h=1}^{T-t} \left(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_h \mathbf{e}_j\right) \mathbf{e}_j' \mathbf{V}_{T-h}^{*'},$$

$$a_{0n}^{(k,j,a)} = \frac{1}{n} \sum_{t=1}^{T-1} \frac{1}{(T-t)^2} \Big[\left(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_{T-t} \mathbf{e}_j\right)^2 \mathrm{Var}\left[\mathbf{u}_t' \mathbf{M}_t^{(a)} \mathbf{v}_t^{*(j)}\right]$$

$$+ \cdots + \left(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_1 \mathbf{e}_j\right)^2 \mathrm{Var}\left[\mathbf{u}_t' \mathbf{M}_t^{(a)} \mathbf{v}_{T-1}^{*(j)}\right]\Big],$$

$$a_{1n}^{(k,j,a)} = \frac{2}{n} \sum_{t=1}^{T-2} \Bigg[\frac{\left(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_{T-t-1} \mathbf{e}_j\right)^2 \mathrm{Cov}\left[\mathbf{u}_t' \mathbf{M}_t^{(a)} \mathbf{v}_{t+1}^{*(j)}, \mathbf{u}_{t+1}' \mathbf{M}_{t+1}^{(a)} \mathbf{v}_{t+1}^{*(j)}\right]}{(T-t)(T-t-1)}$$

$$+ \cdots + \frac{(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_1 \mathbf{e}_j)^2 \mathrm{Cov}\left[\mathbf{u}_t' \mathbf{M}_t^{(a)} \mathbf{v}_{T-1}^{*(j)}, \mathbf{u}_{T-1}' \mathbf{M}_{T-1}^{(a)} \mathbf{v}_{T-1}^{*(j)}\right]}{(T-t)} \Bigg].$$

Using Lemma 1 and $(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_h \mathbf{e}_j)^2$ is bounded, we can evaluate as $a_{0n}^{(k,j,a)} = O(\frac{1}{NT} \sum_t \frac{t}{T-t}) = O(\frac{\log T}{N})$.

Moreover, using the fact that $|\mathcal{E}[\mathbf{d}_{t+j}' \mathbf{M}_t^{(a)} \mathbf{u}_t]| \leq O(tr(\mathbf{M}_{t+j}^{(a)}))$, we find a positive constant C_1 such that

$$|a_{1n}^{(k,j,a)}| \leq C_1 \frac{1}{n} \sum_{t=1}^{T-2} \frac{1}{(T-t)} \left(\frac{t+1}{T-t-1} + \cdots + \frac{T-1}{1}\right),$$

which is $O((\log T)^2/N)$. Finally, we shall evaluate the variance of $\Upsilon_{22n}^{(k,a)}$ as

$$\mathrm{Var}[\Upsilon_{22n}^{(k,a)}] = \frac{1}{n} \mathrm{Var}\left[\sum_{t=1}^{T-1} \frac{1}{T-t} \sum_{h=1}^{T-t} \sum_{j=1}^{G_*} \left(\mathbf{e}_{kJ}' \boldsymbol{\Phi}_h \mathbf{e}_j\right) \mathbf{e}_j' \mathbf{V}_{T-h}^{*'} \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT}\right]$$

$$= \frac{1}{n} \left[\sum_{j=1}^{G_*} \mathrm{Var}\left[\sum_{t=1}^{T-1} \tilde{\mathbf{v}}_{tT}^{*(k,j)'} \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT}\right] + \sum_{i,j}^{G_*} \mathrm{Cov}[.,.]\right].$$

Using the same argument as for Lemma 1, we find

$$\mathrm{Var}\left[\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\right] \le \mathrm{tr}\left(\mathbf{M}_t^{(a)}\right)\left[m^{(1)}\left(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}\right)+m^{(3)}\left(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}\right)\right]$$

and then

$$m^{(3)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) = \mathrm{Var}\left[\frac{1}{T-t}\left(\mathbf{e}_{kJ}'\boldsymbol{\Phi}_{T-t}\mathbf{e}_j v_{it}^{*(j)}+\cdots+\mathbf{e}_{kJ}'\boldsymbol{\Phi}_1\mathbf{e}_j v_{iT-1}^{*(j)}\right)\right] \times\mathrm{Var}\left[\frac{1}{T-t+1}(u_{it}+\cdots+u_{iT})\right],$$

which is $O(1/(T-t)^2)$ because $(v_{it}^{*(j)},u_{it})$ is independent of $(v_{is}^{*(j)},u_{is})$ for $s\neq t$. Similarly,

$$m^{(1)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) = \frac{1}{(T-t)^2(T-t+1)^2} \times\mathcal{E}\left[\left(\mathbf{e}_{kJ}'\boldsymbol{\Phi}_{T-t}\mathbf{e}_j v_{it}^{*(j)}+\cdots+\mathbf{e}_{kJ}'\boldsymbol{\Phi}_1\mathbf{e}_j v_{iT-1}^{*(j)}\right)^2(u_{it}+\cdots+u_{iT})^2\right],$$

which is $O(1/(T-t)^2)$. Therefore, for any j, we have $\mathrm{Var}[\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}] = O(t/(T-t)^2)$. From this result and the arguments as Alvarez and Arellano (2003), we conclude that $\mathrm{Var}[\Upsilon_{22n}^{(k,a)}]=O((\log T)^2/N)$.

Step 4: Now we evaluate the limiting distribution of the LIML estimator with the backward-filtered instruments. We replace $\mathbf{M}_t^{(b)}$ for $\mathbf{M}_t^{(a)}$ and define $\Upsilon_{11n}^{(k,b)}$, $\Upsilon_{12n}^{(k,b)}$, $\Upsilon_{21n}^{(k,b)}$ and $\Upsilon_{22n}^{(k,b)}$, accordingly. We first notice that the order of $\mathrm{Var}[\Upsilon_{12n}^{(k,.)}]$ is free with $\mathbf{M}_t$, and those of $\Upsilon_{21n}^{(k,b)}$ and $\Upsilon_{22n}^{(k,b)}$ are reduced by the fact that $\mathrm{tr}(\mathbf{M}_t^{(b)})=O(1)$. For instance, $\mathrm{Var}[\Upsilon_{12n}^{(k,b)}]=O(1/T)$, $\mathrm{Var}[\Upsilon_{21n}^{(k,b)}]=O((\log T)^2/(N_0T))$ and $\mathrm{Var}[\Upsilon_{22n}^{(k,b)}]=O((\log T)^2/(N_0T))$. To evaluate $\mathrm{Var}[\Upsilon_{11n}^{(k,b)}]$, we prepare the next lemma, which is a generalization of the corresponding one by Hayakawa (2006). The proof of Lemma 2 will be provided in the online supplementary appendix.

Lemma 2 *Define the $N\times 1$ error vectors of the linear projection of $\mathbf{W}_{t-1}\mathbf{J}$ on $\mathbf{Z}_t^{*(b)}\mathbf{J}$,*

$$\mathbf{E}_t^{(b)}=\left[\boldsymbol{\epsilon}_t^{(1,b)},\ldots,\boldsymbol{\epsilon}_t^{(K,b)}\right]=\mathbf{W}_{t-1}\mathbf{J}-\mathbf{Z}_t^{*(b)}\mathbf{J}\left[\boldsymbol{\gamma}_t^{*(1,b)},\ldots,\boldsymbol{\gamma}_t^{*(K,b)}\right], \tag{51}$$

where $\mathbf{Z}_t^{*(b)}=[\mathbf{z}_{1(t-1)}^{*(b)},\ldots,\mathbf{z}_{N(t-1)}^{*(b)}]'$, $\mathbf{Z}_t^{*(b)}\mathbf{J}=\mathbf{Z}_t^{(b)}$ and $\boldsymbol{\gamma}_t^{*(k,b)}$ is defined by $[\boldsymbol{\gamma}_t^{*(1,b)},\ldots,\boldsymbol{\gamma}_t^{*(K,b)}]=(b_t^2\lim_{t\to\infty}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{z}_{it-1}^{*(b)\prime}]\mathbf{J})^{-1}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{w}_{it-1}']\mathbf{J};$. Then, for $k=1,\ldots,K$,

$$\mathcal{E}[\epsilon_{it}^{(k,b)2}]=O\left(\frac{1}{t}\right) \tag{52}$$

and

$$\frac{1}{N_0T}\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(b)}\mathbf{W}_{t-1}\mathbf{J}\xrightarrow{p}\mathbf{J}'\mathcal{E}\left[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}'\right]\mathbf{J}=\mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J}. \tag{53}$$

We turn to evaluate the order of $\mathrm{Var}[\Upsilon_{11n}^{(k,b)}]$

$$\mathrm{Var}[\Upsilon_{11n}^{(k,b)}] = \frac{1}{N_0 T}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathcal{E}\left[\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(b)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(b)}\mathbf{W}_{t-1}\mathbf{J}\mathbf{e}_k\right]. \tag{54}$$

For $t \geq s$ and $k = 1, \ldots, K$,

$$\begin{aligned}
&\mathcal{E}\left[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(b)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(b)}\mathbf{w}_{s-1}^{(k)}\right]\\
&\quad= \frac{\sigma^2}{(T-s+1)}\left[\mathcal{E}\left[\mathbf{w}_{t-1}^{(k)'}\left(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)}\right)\boldsymbol{\epsilon}_{s-1}^{(k,b)}\right]-\mathcal{E}\left[\mathbf{w}_{t-1}^{(k)'}\mathbf{w}_{s-1}^{(k)}\right]\right.\\
&\qquad\left.-\mathcal{E}\left[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}\left(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)}\right)\left(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)}\right)\boldsymbol{\epsilon}_{s-1}^{(k,b)}\right]+\mathcal{E}\left[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}\left(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)}\right)\mathbf{w}_{s-1}^{(k)}\right]\right],
\end{aligned}$$

where we use the decomposition $\mathbf{w}_{h-1}^{(k)'}\mathbf{M}_h^{(b)} = \mathbf{w}_{h-1}^{(k)'} - \boldsymbol{\epsilon}_h^{(k,b)'}[\mathbf{I}_{N_0}-\mathbf{M}_h^{(b)}]$ for $h = t, s$. For the second term of the last equality, we write $\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{w}_{s-1}^{(k)}] = \mathcal{E}[\mathcal{E}_s(\mathbf{w}_{t-1}^{(k)'})\mathbf{w}_{s-1}^{(k)}]$ and then we find $\mathrm{Var}[\Upsilon_{11n}^{(k,a)}] = O(\log T/T)$. Hence for the first term, we have the same result as Step 3. As for the third term $|\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)})(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)})\boldsymbol{\epsilon}_{s-1}^{(k,b)}]|^2$ is less than

$$\mathcal{E}\left[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}\boldsymbol{\epsilon}_{t-1}^{(k,b)}\boldsymbol{\epsilon}_{s-1}^{(k,b)'}\boldsymbol{\epsilon}_{s-1}^{(k,b)}\right] = \sum_{i=1}^{N_0}\mathcal{E}\left[\epsilon_{i(t-1)}^{(k,b)2}\epsilon_{i(s-1)}^{(k,b)2}\right] + \sum_{i,j,i\neq j}^{N_0}\mathcal{E}\left[\epsilon_{i(t-1)}^{(k,b)2}\right]\left[\epsilon_{j(s-1)}^{(k,b)2}\right],$$

which $O(N/(ts))$ and the first equality is due to independence of random variables $\epsilon_{i(t-1)}^{(k,b)2}$. For the second inequality, we have applied Lemma 2 and the Cauchy-Schwarz inequality as

$$\left|\mathcal{E}\left[\epsilon_{i(t-1)}^{(k,b)2}\epsilon_{i(s-1)}^{(k,b)2}\right]\right|^2 \leq \left(\mathcal{E}\left[\epsilon_{i(t-1)}^{(k,b)4}\right]\right)\left(\mathcal{E}\left[\epsilon_{i(t-1)}^{(k,b)4}\right]\right) = O\left(\frac{1}{t^2}\right)\times O\left(\frac{1}{s^2}\right).$$

Thus we can take a positive constant C_2 such that

$$\begin{aligned}
&\frac{1}{N_0 T}\sum_{s=1}^{T-1}2\sum_{t\geq s}^{T-1}\frac{\sigma^2}{T-s+1}\left|\mathcal{E}\left[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}\left(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)}\right)\left(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)}\right)\boldsymbol{\epsilon}_{s-1}^{(k,b)}\right]\right|\\
&\quad\leq C_2\frac{N_0}{N_0 T}\sum_{s=1}^{T-1}2\sum_{t\geq s}^{T-1}\frac{1}{T-s+1}\frac{1}{\sqrt{t}}\frac{1}{\sqrt{s}} = O\left(\frac{(\log T)\sqrt{T}}{T}\right).
\end{aligned}$$

For the fourth term of (54), we have the same order by the similar arguments. Hence, we find that $\mathrm{Var}[\Upsilon_{11n}^{(k,b)}] = O((\log T)/\sqrt{T})$.

Step 5: At this step, we drive the asymptotic covariance and bias of the limiting distribution of the LIML estimator. First, we prepare the next lemma, which is useful for deriving the covariance formula for Case (a). The proof of Lemma 3 will be provided in the online supplementary appendix.

Lemma 3 *Let* $(\mu_i^{(1)}, \ldots, \mu_i^{(k)}, \ldots, \mu_i^{(G_*)})' = \boldsymbol{\mu}_i = \left[\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*\right]^{-1} \boldsymbol{\pi}_i^*$ *and* $\mathbf{M}_\mu = \left[\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N\right]'$. *Define the* $N \times 1$ *error vectors of the linear projection of* $\mathbf{M}_\mu \mathbf{J}$ *on* $\mathbf{Z}_t^{(a)}$,

$$\mathbf{E}_t^{(a)} = \left[\boldsymbol{\epsilon}_t^{(1,a)}, \ldots, \boldsymbol{\epsilon}_t^{(K,a)}\right] = \mathbf{M}_\mu \mathbf{J} - \mathbf{Z}_t^{(a)} \left[\boldsymbol{\gamma}_t^{*(1,a)}, \ldots, \boldsymbol{\gamma}_t^{*(K,a)}\right], \tag{55}$$

where for $k = 1, \ldots, K,\ \ h = 1, \ldots, t$ *we take each* $K_* t \times 1$ *coefficient vector* $\boldsymbol{\gamma}_t^{*(k,a)} = (\boldsymbol{\gamma}_{t1}^{*(k,a)'}, \ldots, \boldsymbol{\gamma}_{th}^{*(k,a)'}, \ldots, \boldsymbol{\gamma}_{tt}^{*(k,a)'})'$ *as* $\gamma_{thl}^{*(k,a)} = 1/t$ *(if* $l = k$*) and* $\gamma_{thl}^{*(k,a)} = 0$, *(if* $l \neq k$*), and* $\boldsymbol{\gamma}_{th}^{*(k,a)} = (\gamma_{th1}^{*(k,a)}, \ldots, \gamma_{thl}^{*(k,a)}, \ldots, \boldsymbol{\gamma}_{thK_*}^{*(k,a)})'$. *Then, for* $k = 1, \ldots, K$,

$$\mathcal{E}[\epsilon_{it}^{(k,a)2}] = O\left(\frac{1}{t}\right) \tag{56}$$

and

$$\frac{1}{NT} \sum_{t=1}^{T-1} \mathbf{J}' \mathbf{W}_{t-1}' \mathbf{M}_t^{(a)} \mathbf{W}_{t-1} \mathbf{J} \xrightarrow{p} \mathbf{J}' \mathcal{E}\left[\mathbf{w}_{i(t-1)} \mathbf{w}_{i(t-1)}'\right] \mathbf{J} = \mathbf{J}' \boldsymbol{\Gamma}_0 \mathbf{J}. \tag{57}$$

For the LIML estimator, we can re-write (46) as

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \mathbf{D}' \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{N}_t \mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{U}_t^{(\perp, f)'} \\ \mathbf{O} \end{pmatrix} \mathbf{N}_t \mathbf{u}_t^{(f)} \\
&= \frac{1}{\sqrt{n}} \mathbf{D}' \sum_{t=1}^{T-1} \mathbf{J}' \mathbf{W}_{t-1}' \mathbf{u}_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{U}_t^{\perp'} \\ \mathbf{O} \end{pmatrix} \mathbf{N}_t \mathbf{u}_t + O(1) + o_p(1) \\
&= \mathbf{a}_{1n} + \mathbf{a}_{2n} + O(1) + o_p(1), \text{ (say,)}
\end{aligned} \tag{58}$$

where $O(1)$ is the associated terms of the asymptotic bias to be discussed below. The first equality above is due to the result of Step 2 and the second equality follows from $\mathbf{N}_t = \mathbf{I}_N - (1 + c_*)(\mathbf{I}_N - \mathbf{M}_t)$ and

$$\begin{aligned}
\text{Var}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \mathbf{W}_{t-1}' (\mathbf{I}_N - \mathbf{M}_t) \mathbf{u}_t\right] &= \frac{\mathcal{E}[u_{it}^2]}{n} \sum_{t=1}^{T-1} \mathcal{E}\left[\boldsymbol{\epsilon}^{(k,.)'} (\mathbf{I}_N - \mathbf{M}_t) \boldsymbol{\epsilon}^{(k,.)}\right] \\
&\leq \frac{\sigma^2}{T} \sum_{t=1}^{T-1} \mathcal{E}\left[\epsilon_{it}^{(k,.)2}\right] = \frac{O(\log T)}{T},
\end{aligned}$$

where $\epsilon_{it}^{(k,.)} = \epsilon_{it}^{(k,a)}$ (or $\epsilon_{it}^{(k,a)}$), and we have used Lemmas 2 and 3. We notice

$$\mathcal{E}\left[\mathbf{a}_{1n}\mathbf{a}_{1n}^{'}\right] = \frac{\mathcal{E}_t[u_{it}^2]}{n}\mathbf{D}^{'}\mathcal{E}\left[\sum_{t=1}^{T-1}\mathbf{J}^{'}\mathbf{W}_{t-1}^{'}\mathbf{W}_{t-1}\mathbf{J}\right]\mathbf{D} \longrightarrow \sigma^2\boldsymbol{\Phi}^{*}. \tag{59}$$

Using the i-th unit vector $\mathbf{e}_i$ $(i = 1, \ldots, N)$, we find

$$\begin{aligned}
\mathcal{E}\left[\mathbf{a}_{1n}\mathbf{a}_{2n}^{'}\right] &= \left(\frac{1}{n}\mathbf{D}^{'}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{J}^{'}\mathbf{W}_{t-1}^{'}\mathcal{E}_t\left[\mathbf{u}_t\mathbf{u}_t^{'}\mathbf{N}_t\mathbf{U}_t^{\perp}\right]\right], \mathbf{O}\right) \\
&= \left(\frac{1}{n}\mathbf{D}^{'}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{J}^{'}\mathbf{W}_{t-1}^{'}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{e}_i\mathbf{e}_i^{'}\mathcal{E}_t\left[u_{it}^2\mathbf{N}_t\mathbf{e}_j\mathbf{u}_{jt}^{\perp'}\right]\right], \mathbf{O}\right) \\
&= \left(\frac{1}{n}\mathbf{D}^{'}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{J}^{'}\mathbf{W}_{t-1}^{'}\mathbf{d}_t\right]\mathcal{E}\left[u_{it}^2\mathbf{u}_{it}^{\perp'}\right]\left(\frac{1}{1-c}\right), \mathbf{O}\right),
\end{aligned}$$

since for any i, j, $\mathcal{E}_t[\mathbf{u}_{jt}^{\perp}u_{it}] = 0$ and $\mathcal{E}[\mathbf{W}_{t-1}^{'}c_*\mathbf{I}_N] = \mathbf{0}$. Furthermore, we use the decomposition

$$\mathcal{E}\left[\mathbf{J}_{G_2}^{'}\mathbf{a}_{2n}\mathbf{a}_{2n}^{'}\mathbf{J}_{G_2}\right] = \frac{1}{n}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{U}_t^{\perp'}\mathbf{N}_t\left[\sigma^2\mathbf{I}_N + (\mathbf{u}_t\mathbf{u}_t^{'} - \sigma^2\mathbf{I}_N)\right]\mathbf{N}_t\mathbf{U}_t^{\perp}\right]$$

and the first term converges $(1/n)\sum_{t=1}^{T-1}\mathrm{tr}(\mathbf{N}_t^2)\sigma^2\mathcal{E}[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}] \longrightarrow c_*\sigma^2\mathcal{E}[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}]$ because we have $\mathbf{N}_t^2 = \mathbf{M_t} + c_*^2(\mathbf{I}_N - \mathbf{M}_t)$ and

$$\frac{1}{n}\sum_{t=1}^{T-1}\mathrm{tr}(\mathbf{M}_t) + c_*^2\frac{1}{n}\sum_{t=2}^{T-1}\mathrm{tr}(\mathbf{I}_N - \mathbf{M}_t) = \frac{r_n}{n} + \frac{q_n}{n}c_*^2 \longrightarrow c_*.$$

Then for any constant vector $\mathbf{h}$, we write the second term as

$$\begin{aligned}
&\mathbf{h}^{'}\frac{1}{n}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{U}_t^{\perp'}\mathbf{N}_t(\mathbf{u}_t\mathbf{u}_t^{'} - \sigma^2\mathbf{I}_N)\mathbf{N}_t\mathbf{U}_t^{\perp}\right]\mathbf{h} \\
&= \frac{1}{n}\sum_{t=2}^{T-1}\sum_{j=1}^{N}\mathcal{E}\left[(\mathbf{e}_j^{'}\mathbf{N}_t\mathbf{e}_j)^2\mathcal{E}_t\left[(u_{it}^2 - \sigma^2)(\mathbf{u}_{it}^{\perp'}\mathbf{h})^2\right]\right].
\end{aligned}$$

Using the similar calculations as $\mathcal{E}[\mathbf{a}_{1n}\mathbf{a}_{2n}^{'}]$. Then under the assumptions we made, we need to evaluate the convergence of $[1/(NT)]\sum_{t=1}^{T-1}\mathbf{d}_t^{(\cdot)'}\mathbf{W}_{t-1}^{'}$ and $[1/(NT)]\sum_{t=1}^{T-1}\mathbf{d}_t^{(\cdot)'}\mathbf{d}_t^{(\cdot)}$. In each case, $d_{ti}^{(\cdot)}$ are bounded and $(1/T)\sum_{t=2}^{T-1}\mathbf{w}_{ti}$ converges to zero in probability. Further in the case of (a) $\max_i d_{it}$ converges to zero in probability as $N \to \infty$. Thus, the effects of third-order terms are negligible in the case of (a), while the effects of third-order terms may not be negligible in the case of (b) since N can be fixed. In both cases, we have the effects of fourth-order terms in the general case (see Condition (IV) of Anderson et al. 2010).

Next, we evaluate the asymptotic bias of LIML estimator. We notice $\mathcal{E}[\Upsilon_{4n}^{(g,a)}] = \mathcal{E}[\Upsilon_{4n}^{(g,b)}] = 0$ in (47) using the fact that for any i, j, s, t, $\mathcal{E}_t[\mathbf{u}_{it}^{\perp} u_{js}] = 0$. Then in the case of $\mathbf{M}_t = \mathbf{M}_t^{(a)}$, we evaluate the asymptotic bias as

$$\mathbf{b}_c^{(a)} = \boldsymbol{\Phi}^{*-1}\mathbf{D}' \lim_{N,T\to\infty} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} \mathcal{E}\left[\mathbf{Z}_{t-1}^{(f)'}\left(\frac{1}{1-c_a}\right)\left(\mathbf{M}_t^{(a)} - c_a\mathbf{I}_N\right)\mathbf{u}_t^{(f)}\right]. \quad (60)$$

For the term $\sum_{t=1}^{T-1} \mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{u}_t^{(f)}] = -(N/T)\mathbf{J}'\mathcal{E}[\mathbf{W}_{i(-1)}'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{u}_i]$, we have

$$\begin{aligned}
\mathcal{E}\left[\mathbf{W}_{i(-1)}'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{u}_i\right] &= \sum_{h=1}^{T-1}\sum_{j=0}^{T-1-h} \boldsymbol{\Pi}^{*j}\mathcal{E}[\mathbf{v}_{it}^* u_{it}] \\
&= T(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^* u_{it}] \\
&\quad -(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}[\mathbf{I}_{G_*} + \boldsymbol{\Pi}^*(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^{*T-1})]\mathcal{E}[\mathbf{v}_{it}^* u_{it}].
\end{aligned}$$

For the term $\sum_{t=1}^{T-1} \mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}] = -\sum_{t=1}^{T-1}\mathbf{J}'\mathcal{E}[c_t\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}]$, we can evaluate as

$$\begin{aligned}
&\mathcal{E}\left[c_t\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}\right] \\
&= \frac{K_* t}{T-t+1}\left[\boldsymbol{\Phi}_{T-t}\mathcal{E}[\mathbf{v}_{it}^* u_{it}] - \frac{1}{T-t}(\boldsymbol{\Phi}_{T-t-1} + \cdots + \boldsymbol{\Phi}_1)\mathcal{E}[\mathbf{v}_{it}^* u_{it}]\right] \\
&= \frac{K_* t}{T-t+1}(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}\left[(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^{*T-t}) - \mathbf{I}_{G_*}\right. \\
&\quad \left. + \left(\frac{1}{T-t}\right)[\mathbf{I}_{G_*} + \boldsymbol{\Pi}^*(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^{*T-t-1})]\right]\mathcal{E}[\mathbf{v}_{it}^* u_{it}] \quad (61)
\end{aligned}$$

and then

$$\begin{aligned}
\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}\right] &= -K_*(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*)^{-2}\left[(T-1)(\mathbf{I}_{G_*} - \boldsymbol{\Pi}^*) + O(\log T)\right] \\
&\quad \times\mathcal{E}[\mathbf{v}_{it}^* u_{it}].
\end{aligned}$$

Therefore, we obtain

$$\mathbf{b}_c^{(a)} = -\lim_{N,T\to\infty}\left(\frac{K_*}{1-c_a}\frac{T}{\sqrt{NT}} - \frac{c_a}{1-c_a}\frac{N}{\sqrt{NT}}\right)\boldsymbol{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^* u_{it}]$$

$$= -\frac{K_*\sqrt{\lim_{n\to\infty}(T/N)}}{2-K_*\lim_{n\to\infty}(T/N)}\boldsymbol{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^* u_{it}]. \quad (62)$$

Similarly, we consider the case of $\mathbf{M}_t = \mathbf{M}_t^{(b)}$ and

$$\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^{(b)}\mathbf{u}_t^{(f)}] = -K\mathbf{J}'(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)^{-2}\left[(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)-(1/T)(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^{*T})\right]$$
$$\times\mathcal{E}[\mathbf{v}_{it}^* u_{it}].$$

Then, regardless of whether $N_0\to\infty$ or fixed, the asymptotic bias becomes

$$\mathbf{b}_c^{(b)} = -\lim_{T\to\infty}\left(\frac{K}{1-c_b}\frac{1}{\sqrt{N_0T}} - \frac{c_b}{1-c_b}\frac{N_0}{\sqrt{N_0T}}\right)\boldsymbol{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{G_*}-\boldsymbol{\Pi}^*)^{-1}$$
$$\times\mathcal{E}[\mathbf{v}_{it}^* u_{it}] = \mathbf{0}.$$

Step 6: We now turn to consider the asymptotic covariance matrix and the bias of the GMM estimator. The necessary arguments are basically the same as those in Step 3 for the LIML estimator. If $c = 0$, the normalized GMM estimator are asymptotically equivalent to

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{GM}-\boldsymbol{\theta}) = \boldsymbol{\Phi}^{*-1}\left[\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{u}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{J}_{*G_2}'\mathbf{V}_t'\\ \mathbf{O}\end{pmatrix}\mathbf{M}_t\mathbf{u}_t\right] + o_p(1), \quad (63)$$

and $\mathbf{J}_{*G_2}' = [\mathbf{0}, \mathbf{I}_{G_2}]$. For any constant vector $\mathbf{h}$ we write $\mathbf{h}_G = \mathbf{J}_G\mathbf{h}$, and by Lemma 1

$$\mathrm{Var}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{h}_G'\mathbf{V}_t'\mathbf{M}_t\mathbf{u}_t\right] = \frac{1}{n}\sum_{t=1}^{T-1}\mathrm{Var}[\mathbf{h}_G'\mathbf{V}_t'\mathbf{M}_t\mathbf{u}_t] = O(c). \quad (64)$$

Thus in each case, $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ or $\mathbf{M}_t^{(b)}$, the asymptotic variance–covariance matrix becomes $\boldsymbol{\Phi}^{*-1}(\sigma^2\boldsymbol{\Phi}^*)\boldsymbol{\Phi}^{*-1} = \sigma^2\boldsymbol{\Phi}^{*-1}$. Also under the condition $\sum_{t=1}^{T-1}\mathrm{tr}(\mathbf{M}_t)/(\sqrt{n})$ is bounded, the asymptotic bias becomes

$$\mathbf{b}_0^{(.)} = \lim_{N,T\to\infty}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathrm{tr}(\mathbf{M}_t)\right]\boldsymbol{\Phi}^{*-1}\begin{pmatrix}\mathbf{J}_{*G_2}'\mathcal{E}[\mathbf{v}_{it}u_{it}]\\ \mathbf{0}\end{pmatrix}. \quad (65)$$

Step 7: Finally, we consider the asymptotic normality of the LIML estimator (the asymptotic normality of the GMM estimator can be proven as the special case of $\boldsymbol{\alpha}_{2t} = \mathbf{0}$ below). Define the $(G_2 + K_1) \times 1$ martingale difference sequence by

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{1t} + \boldsymbol{\alpha}_{2t} = \frac{1}{\sqrt{N}} \left[\mathbf{D}'\mathbf{J}' \sum_{i=1}^{N} \mathbf{w}_{i(t-1)} \mathbf{u}_t + \begin{pmatrix} \mathbf{U}_t^{\perp'} \\ \mathbf{O} \end{pmatrix} \mathbf{N}_t \mathbf{u}_t \right] \text{(say,)} \tag{66}$$

and then $\mathbf{a}_{1n} + \mathbf{a}_{2n} = (1/\sqrt{T}) \sum_{t=1}^{T-1} (\boldsymbol{\alpha}_{1t} + \boldsymbol{\alpha}_{2t})$.
In the present situation, we have the conditions (i) $(1/n) \sum_{t=1}^{T-1} \mathbf{W}'_{t-1}(\mathbf{W}_{t-1}, \boldsymbol{\iota}_N) \xrightarrow{p} (\boldsymbol{\Gamma}_0, \mathbf{0})$, and (i) the same evaluations as used for the asymptotic covariance evaluation, for any constant vector $\mathbf{h}$ and any N (because u_{it} is uncorrelated with $\mathcal{F}_{t-1}$ and $\mathcal{F}_{t-1}$ is the σ−field generated the random variables given at $t - 1$), we have

$$\frac{1}{T} \sum_{t=1}^{T-1} \mathcal{E} \left[\mathbf{h}'\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t' \mathbf{h} | \mathcal{F}_{t-1} \right] \longrightarrow \text{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} \mathcal{E} \left[\mathbf{h}'\boldsymbol{\alpha}_t \boldsymbol{\alpha}_t' \mathbf{h} \right] \tag{67}$$

and for some constant Δ' and any t, N, $\mathcal{E}[[\mathbf{h}'(\boldsymbol{\alpha}_{1t} + \boldsymbol{\alpha}_{2t})]^4] < \Delta'$. It is because $\mathcal{E}[[\mathbf{h}'\boldsymbol{\alpha}_{1t}]^4] < \infty$ and $\mathcal{E}[[(c_*/\sqrt{N})\mathbf{t}'(\mathbf{U}^{\perp'}(\mathbf{I}_N - \mathbf{M}_t)\mathbf{u}_t)]^4] < \infty$ using the similar arguments in the next Lemma 4 below. Then the Lyapounov conditions for the central limit theorem hold for both cases when $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$.

Lemma 4 *For any $G_2 \times 1$ constant vector $\mathbf{h}$ and any t, N, there is a positive constant Δ such that*

$$\mathcal{E} \left[\left[\left(\frac{1}{\sqrt{N}} \right) \mathbf{h}'\mathbf{U}_t^{\perp'} \mathbf{M}_t \mathbf{u}_t \right]^4 \right] < \Delta. \tag{68}$$

The proof of Lemma 4 will be provided in the online supplementary appendix.

7 Appendix: Some figures

In figures, the distribution functions of estimators are shown with the standard normalization (the case of $c = 0$) and the large-K_2 normalization (the case of $c > 0$). The limiting distributions for LIML in the large-K asymptotics are $N_2(\mathbf{0}, \mathbf{I}_2)$ and its marginal distributions are $N(0, 1)$, which are denoted as "o". For the sake of comparisons, the distribution of GMM are normalized in the same way and our settings are similar to those in Anderson et al. (2005, 2011) and Akashi and Kunitomo (2012).

See Figs. 1, 2, 3, 4, 5 and 6.

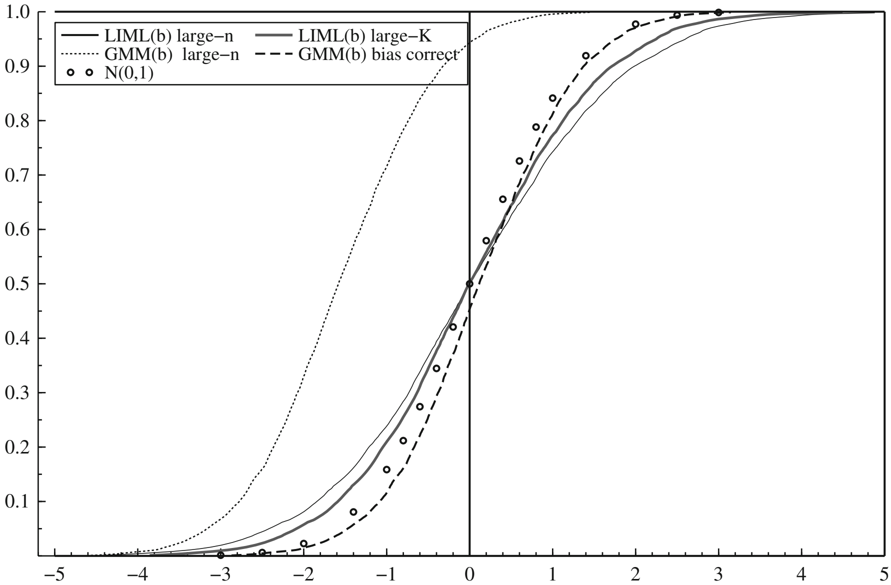

Fig. 1 $\beta_2 : \ N = 100, \ T = 50, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

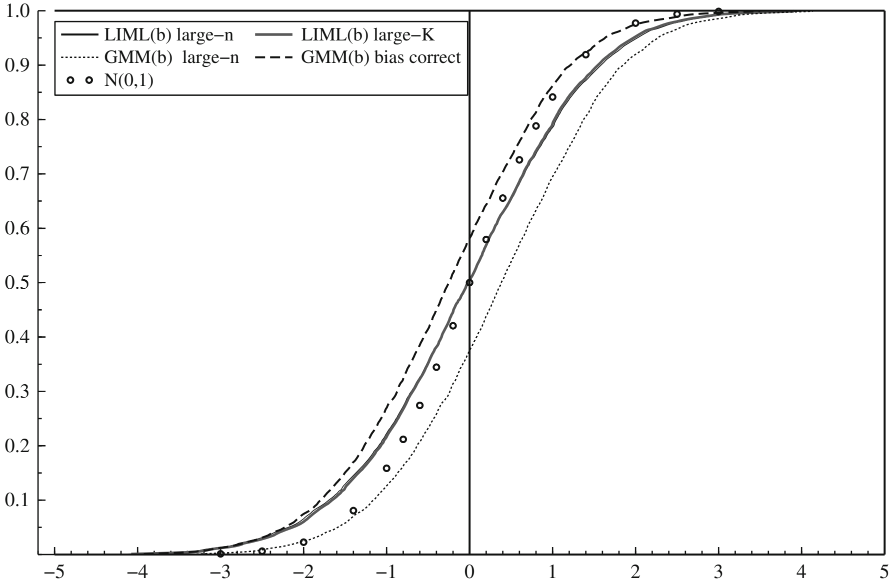

Fig. 2 $\gamma_{11} : \ N = 100, \ T = 50, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

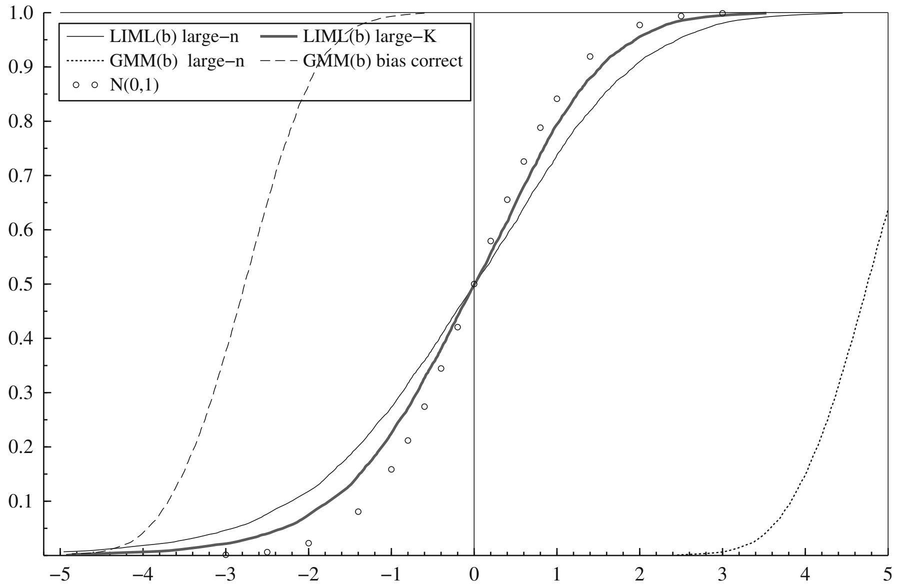

Fig. 3 $\beta_2: \; N = 100, \; T = 50, \; c_b = \frac{4}{100}, \; (\omega_{11}, \omega_{12}) = (1.5, 1.0)$

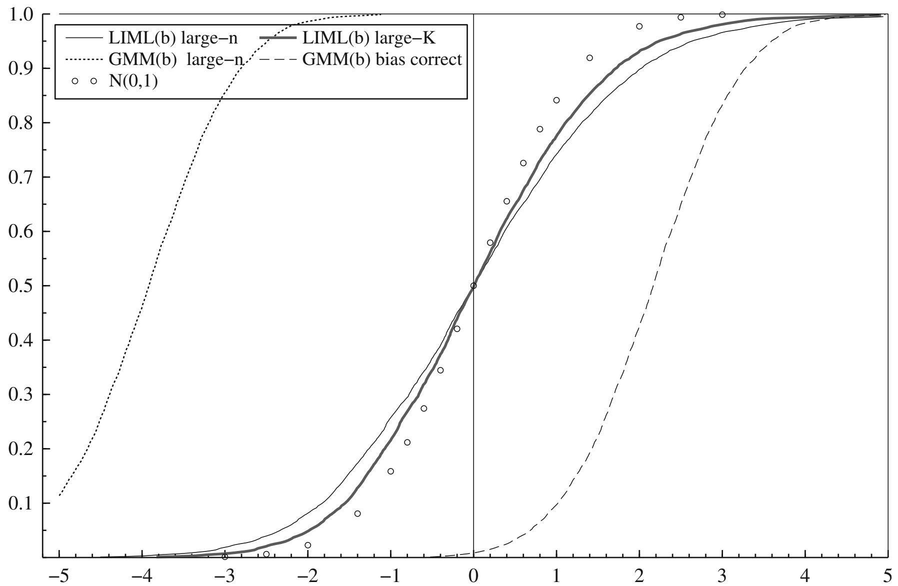

Fig. 4 $\gamma_{11}: \; N = 100, \; T = 50, \; c_b = \frac{4}{100}, \; (\omega_{11}, \omega_{12}) = (1.5, 1.0)$

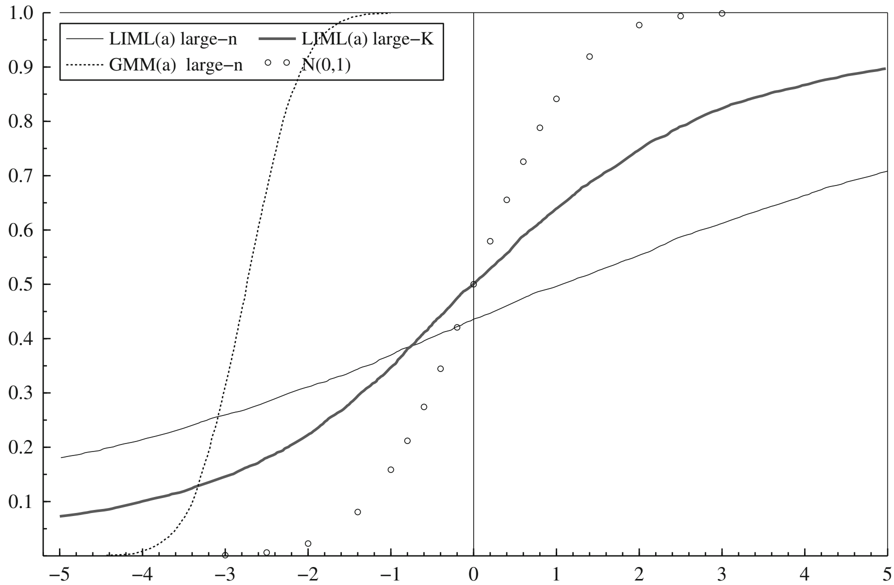

Fig. 5 β_2 : $N = 100,\ T = 25,\ c_a = \frac{3}{2}\frac{25}{100},\ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

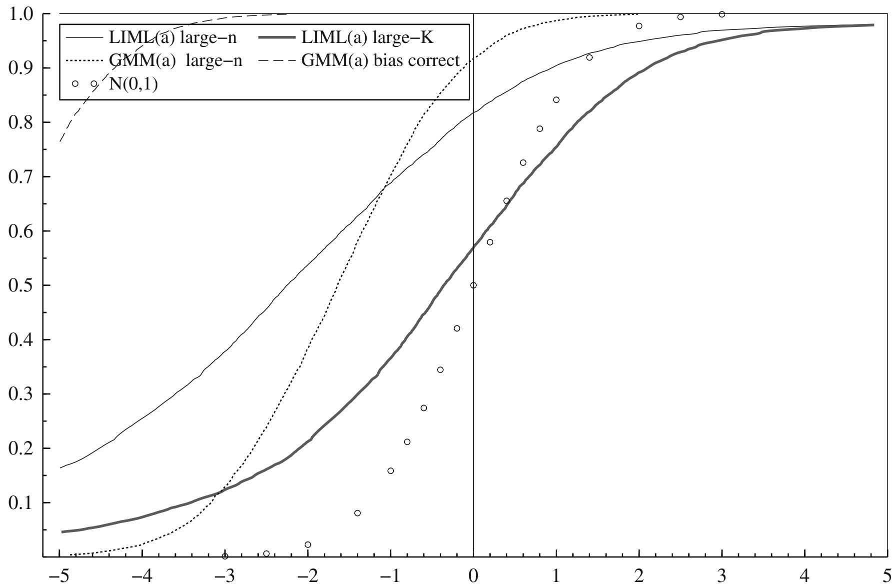

Fig. 6 γ_{11} : $N = 100,\ T = 25,\ c_a = \frac{3}{2}\frac{25}{100},\ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

Acknowledgments We thank the editor and two referees of this journal for helpful comments to the previous version. The earlier versions of this paper were presented at the 2006 Japan Statistical Society (JSS) meeting, the 62nd European Meeting of the Econometric Society, and the 2010 Econometric Society World Congress (2010, Shanghai) under the title "The Conditional Limited Information Maximum Likelihood Approach to Panel Dynamic Structural Equations". We also thank T.W. Anderson, Y. Matsushita, K. Hayakawa, R. Okui, K. Hitomi and Y. Nishiyama for comments to the earlier versions.

References

Akashi, K. (2008). t-test in dynamic panel structural equations (unpublished manuscript).

Akashi, K., Kunitomo, N. (2012). Some properties of the LIML estimator in a dynamic panel structural equation. *Journal of Econometrics*, *166*, 167–183.

Alonso-Borrego, C., Arellano, M. (1999). Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business and Economic Statistics*, *17*, 36–49.

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

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

Anderson, T. W., Hsiao, C. (1982). Formulation and estimation of dynamic models with panel data. *Journal of Econometrics*, *18*, 47–82.

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

Anderson, T. W., Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equation. *Annals of Mathematical Statistics*, *21*, 570–582.

Anderson T.W., et al. (2005). A new light from old wisdoms: alternative estimation methods of simultaneous equations with possibly many instruments, Discussion Paper CIRJE-F-321. Tokyo: Graduate School of Economics, University of Tokyo.

Anderson, T.W., et al. (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments. *Journal of Econometrics*, *157*, 191–204.

Anderson, T.W., et al. (2011). On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments. *Journal of Econometrics*, *165*, 58–69.

Arellano, M. (2003). *Panel Data Econometrics*. New York: Oxford University Press.

Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*, 29–51.

Baltagi, B.H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). New York: Wiley, Hoboken.

Blundell, R., Bond, S. (2000). GMM estimation with persistent panel data : an application to production function. *Econometric Review*, *19–3*, 321–340.

Hahn, J., Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, *70*, 1639–1657.

Hayakawa, K. (2006). Efficient GMM estimation of dynamic panel data models where large heterogeneity may be present. Tokyo: Hi-Stat Discussion Paper No.130, Hitotsubashi University.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator in panel AR(p) models. *Econometric Theory*, *25*, 873–890.

Holtz-Eakin, D., et al. (1988). Estimating vector autoregressions with panel data. *Econometrica*, *56*, 1371–1395.

Hsiao, C. (2003). *Analysis of Panel Data* (2nd ed.). Cambridge: Cambridge University Press.

Kunitomo, N. (1980). Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association*, *75*, 693–700.

Kunitomo, N. (2012). An optimal modification of the LIML estimation for many instruments and persistent heteroscedasiticity. *Annals of Institute of Statistical Mathematics*, *64*, 881–910.

Kunitomo, N., Akashi, K. (2010). An asymptotically optimal modification of the panel LIML estimation for individual heteroscedasticity. Tokyo: Discussion paper CIRJE-F-780, Graduate School of Economics, University of Tokyo. http://www.cirje.e.u-tokyo.ac.jp/research/dp