CIRJE-F-708

The Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equations

Kentaro Akashi
Institute of Statistical Mathematics

Naoto Kunitomo
University of Tokyo

January 2010

The Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equations *

Kentaro Akashi [2]

and

Naoto Kunitomo [3]

January 25, 2010

abstract

We develop the panel limited information maximum likelihood (PLIML) approach for estimating dynamic panel structural equation models. When there are dynamic effects and endogenous variables with individual effects at the same time, the PLIML estimation method for the filtered data does give not only a consistent estimator, but also it has the asymptotic normality and often attains the asymptotic bound when the number of orthogonal conditions is large. Our formulation includes Alvarez and Arellano (2003), Blundell and Bond (2000) and other linear dynamic panel models as special cases. We also compare the PLIML and dynamic GMM (generalized method of moments) estimation methods and suggest an asymptotically optimal modification of LIML under heteroscedastic disturbances among individuals.

Keywords : Dynamic Panel Structural Equations, PLIML, Dynamic GMM, Long Panel, Many Orthogonal Conditions, Forward Filtering, Backward Filtering, Asymptotic Optimality, Individual Heteroscedasticity.

1 Introduction

Recently there has been a growing interest on dynamic panel econometric models in micro-econometrics. The main reason may be due to the fact that there have been a number of panel data available and their analyses have been growing in many applied fields of economics. Then the econometric methods of panel data have been indispensable tools in econometrics (See Hsiao (2003), Arellano (2003)

*KA10-1-25-3. The earlier versions of this paper were presented at the Japan Statistical Society (JSS) meeting, 62^{ed} European Meeting of the Econometric Society, and the Japanese Economic Association (JEA) meeting under the title "The Conditional Limited Information Maximum Likelihood Approach to Panel Dynamic Structural Equations". We thank Kousuke Ōya, Yukitoshi Matsushita and Kazuhiko Hayakawa for comments.

[2]Institute of Statistical Mathematics

[3]Graduate School of Economics, University of Tokyo

and Baltagi (2005) for instance.). However, there are still non-trivial statistical problems of 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, it has been known that the standard statistical 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 causes some kind of *the incidental parameters* when we have observations over a long time-horizon. Earlier investigations on some aspect of the dynamic panel models were Anderson and Hsiao (1981, 1982).

In this paper we propose a new econometric method called *the panel limited information maximum likelihood* (PLIML) approach to the estimation of dynamic panel structural equation models. It is actually an extension of the traditional limited information maximum likelihood (LIML) method, which was originally developed by Anderson and Rubin (1949, 1950). We intend to apply and extend the LIML method to the estimation of dynamic panel structural models when there are dynamic effects and endogenous variables with individual effects at the same time. However, we need to modify the LIML method to handle the dynamic panel models with individual effects and possibly *many orthogonal conditions*. It is 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 need to apply the filtering procedure to remove individual effects in data sets. The PLIML estimation method proposed in this paper gives a consistent estimator and it often attains the asymptotic efficiency bound for general dynamic panel structural equation models, which have the Panel VARs as the reduced form even when the relative ratio T/N (where T is the time-horizon and N is the number of individuals) can not be negligible. 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. By using panel dimensions (N, T) and the number of instruments K, 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 asymptotics* developed by Kunitomo (1980) as an early study and it has been recently examined by Anderson, Kunitomo and Matsushita (2005, 2008a,b).

In our framework of study we shall consider different ways of filtering procedures before estimation systematically, namely, the forward-filtering explained by Alvarez and Arellano (2003) and the backward-filtering explained by Hayakawa

(2006, 2008). 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 we shall show that 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 this situation the *fixed-N asymptotics* to the approximation of the limiting distribution as the first-order approximation is useful for practical applications.

In Section 2 we state the formulation of models and alternative estimation methods of unknown parameters in the dynamic panel structural equations with possibly many instruments and the filtering procedures. Then in Section 3 we give the results on the asymptotic properties of the PLIML estimation method and its asymptotic optimality. In Section 4 we shall discuss some finite sample properties of the GMM and LIML estimators based on a set of Monte Carlo simulations. Some concluding remarks will be given in Section 5. The proofs of our theorems will be given in Section 6.

2 PLIML Approach to Dynamic Panel Structural Equations

2.1 Model

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

$$y_{it}^{(1)} = \sum_{j=2}^{1+G_2} \beta_{2j} y_{it}^{(j)} + \sum_{j=1}^{1+G_2} \sum_{p_j=1}^{Q_j} \gamma_{1jp_j} y_{it-p_j}^{(j)} + \sum_{l=1}^{L} \gamma_{1l} x_{itl}^{(1)} + \eta_i + u_{it} \ , \qquad (2.1)$$

where $y_{it}^{(j)}$ $(j = 1, ..., 1 + G_2)$ are the endogenous variables in the system at period t, $x_{itl}^{(1)}$ $(l = 1, ..., L)$ are the included exogenous variables, $\beta_{2j}, \gamma_{1jp_j}, \gamma_{1l}$ $(j = 1, ..., 1 + G_2;\ p_j = 1, ..., Q_j;\ l = 1, ..., L)$ are the unknown coefficients of the right-hand side variables, η_i $(i = 1, ..., N)$ are individual effects and u_{it} are mutually independent (over individuals and periods) disturbance terms with $\mathcal{E}(u_{it}) = 0$ and $\mathcal{E}(u_{it}^2) = \sigma^2$. In (2.1) we allow some coefficients can be zeros, and the original sample size is $NT\ (= n)$ for $i = 1, \cdots, N; t = 1, \cdots, T$.

We rewrite the dynamic panel structural equation as

$$y_{it}^{(1)} = \boldsymbol{\beta}_2' \boldsymbol{y}_{it}^{(2)} + \boldsymbol{\gamma}_1' \mathbf{z}_{it-1}^{(1)} + \eta_i + u_{it} \ , \qquad (2.2)$$

where $y_{it}^{(1)}$ and $\mathbf{y}_{it}^{(2)}$ $(G_2 \times 1)$ are $1 + G_2$ endogenous variables, $\mathbf{z}_{it-1}^{(1)}$ is the $K_1(= \sum_{j=1}^{1+G_2} Q_j + L)$ vector of the included predetermined variables in (2.1), then $\boldsymbol{\gamma}_1$ and

$\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters. We use the notation such that the vector $\mathbf{z}_{it-1}^{(1)}$ consists of $x_{itl}^{(1)}$ $(l = 1, ..., L)$ and possibly the lagged endogenous variables $y_{it-p_j}^{(j)}$ in this representation.

We assume that the reduced form is written as

$$\mathbf{y}_{it} = \mathbf{\Pi}\mathbf{z}_{it-1} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \ , \tag{2.3}$$

where $\mathcal{E}(\mathbf{v}_{it}) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_{it}\mathbf{v}_{it}^{'}) = \mathbf{\Omega} > 0$ (a positive definite matrix). It can be rewritten as the extended reduced form or the vectored AR(1) representation of the reduced form

$$\mathbf{z}_{it}^{*} = \mathbf{\Pi}^{*}\mathbf{z}_{it-1}^{*} + \boldsymbol{\pi}_i^{*} + \mathbf{v}_{it}^{*} \ , \tag{2.4}$$

$$\mathbf{y}_{it} = \mathbf{J}_{1+G_2}^{'}\mathbf{z}_{it}^{*} \ , \ \mathbf{z}_{it-1}^{(1)} = \mathbf{J}_{K_1}^{'}\mathbf{z}_{it-1} \ , \ \mathbf{z}_{it-1} = \mathbf{J}_{K}^{'}\mathbf{z}_{it-1}^{*} \ , \tag{2.5}$$

where $\mathbf{y}_{it} = (y_{it}^{(1)}, \boldsymbol{y}_{it}^{(2)'})'$ is the $(1+G_2)$ vector of endogenous variables, $\mathbf{z}_{it-1}$ is the $K \times 1$ $(K = K_1 + K_2)$ vector of predetermined variables at t which includes the K_1 exogenous variables and lagged endogenous variables, $\mathbf{\Pi}$ and $\boldsymbol{\pi}_i$ are a $(1+G_2) \times K$ coefficients matrix and a $(1+G_2) \times 1$ individual effect vector, respectively. For the equation (2.5), $\mathbf{\Pi}^{*}$ is the $K^{*} \times K^{*}$ autoregressive coefficients $(K^{*} = K + K_3)$, $\boldsymbol{\pi}_i^{*}$ and $\mathbf{v}_{it}^{*}$ are also $K^{*} \times 1$ individual effects and disturbances vector, respectively, the K_3-variables are excluded from $(1+G_2)$ reduced form equations. In our formulation $\mathbf{J}_{1+G_2}^{'}$ is an $(1+G_2) \times K^{*}$ selection matrix whose each element is one or zero, thus the selection matrix $\mathbf{J}_{K_1}^{'}$ and $\mathbf{J}_{K}^{'}$ are defined in the same way. Also we prepare the notation K_{*}, which means the number of the distinct autoregressive variables in $\mathbf{z}_{it-1}$, therefore

$$K_{*} \leq K \leq K^{*} \ . \tag{2.6}$$

We 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 $\{\mathbf{v}_{it-h}^{*}, \boldsymbol{\pi}_i^{*}\}_{h=1}^{\infty}$, then we shall use the notation $\mathcal{E}_t[\ .\ ] = \mathcal{E}[\ .\ |\mathcal{F}_{t-1}]$ for the conditional expectation operator. The relation between the coefficients in (2.2) and (2.3) gives the condition $(1, -\boldsymbol{\beta}_2^{'})\mathbf{\Pi} = (\boldsymbol{\gamma}_1^{'}, \mathbf{0}^{'})$ and $\boldsymbol{\pi}_{21} = \mathbf{\Pi}_{22}\boldsymbol{\beta}_2$, where $\mathbf{\Pi}_1^{'} = (\boldsymbol{\pi}_{11}, \mathbf{\Pi}_{12})$ is a $K_1 \times (1+G_2)$ matrix, $\mathbf{\Pi}_2^{'} = (\boldsymbol{\pi}_{21}, \mathbf{\Pi}_{22})$ is a $K_2 \times (1+G_2)$ matrix and the $(K_1+K_2) \times (1+G_2)$ matrix of coefficients is partitioned as

$$\mathbf{\Pi}' = \begin{bmatrix} \boldsymbol{\pi}_{11} & \mathbf{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \mathbf{\Pi}_{22} \end{bmatrix} = \left[\mathbf{J}_{1+G_2}^{'}\mathbf{\Pi}^{*}\mathbf{J}_{K_1,K_2}\right]^{'} \ , \tag{2.7}$$

where $\mathbf{J}_{K_1,K_2}^{'}$ is a $K \times K$ selection matrix for reordering columns of the corresponding matrix $\mathbf{\Pi}^{*}$ which is slightly different from $\mathbf{J}_{K}^{'}$.

Although we may call (2.3) and (2.4) as the *reduced form*, the predetermined variables in $\mathbf{z}_{it-1}$ are correlated with unobserved variables $(\boldsymbol{\pi}_i + \mathbf{v}_{it})$ since

$$\mathcal{E}[\mathbf{z}_{it-1}\boldsymbol{\pi}_i'] \neq \mathbf{O} \tag{2.8}$$

in the general case, and this aspect makes the panel model consisting of (2.2) and (2.3) different from the standard simultaneous equation models. We give two examples of dynamic panel structural equations known in the econometric literatures.

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

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \eta_i + u_{it} && (2.9) \\ y_{it}^{(2)} &= \gamma_2 y_{it-1}^{(2)} + \delta\eta_i + v_{it} \ , && (2.10) \end{aligned}$$

where the disturbance terms u_{it} and v_{it} are correlated. In this example $K = K^* = K_* = 2, K_1 = 1$ and $G_2 = 1$. We notice that the equation (2.10) can be regarded as a reduced form equation and the estimation problem of γ_2 was considered by Alvarez and Arellano (2003). They applied the forward-filter 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 $T(T-1)/2$ $(= r_n)$. On the other hand, Hayakawa (2006, 2007) has suggested to use the backward-filter to instruments, which will be defined shortly, for estimation problem of (2.10) and its variant.

Example 2 : Our formulation includes the Panel Vector Autoregressive (Panel VARs) model as the reduced form, which was suggested by Holtz-Eakin, Newey and Rosen (1988). An example can be written as

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} x_{it} + \eta_i + u_{it} \ , \tag{2.11}$$

and the extended reduced form is defined by

$$\begin{pmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ y_{it-1}^{(2)} \\ x_{it+1} \\ x_{it} \end{pmatrix} = \begin{pmatrix} \pi_{11}^* & \pi_{12}^* & \pi_{13}^* & \pi_{14}^* & 0 \\ 0 & \pi_{21}^* & \pi_{22}^* & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{31}^* & \pi_{32}^* \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ y_{it-2}^{(2)} \\ x_{it} \\ x_{it-1} \end{pmatrix} + \begin{pmatrix} \pi_i^{*(1)} \\ \pi_i^{*(2)} \\ 0 \\ \pi_i^* \\ 0 \end{pmatrix} + \begin{pmatrix} v_{it}^{*(1)} \\ v_{it}^{*(2)} \\ 0 \\ \epsilon_{it+1}^* \\ 0 \end{pmatrix} \tag{2.12}$$

where the first two rows of (2.12) are the Panel VARs model $(1 + G_2 = 2)$, and x_{it} is the included independent variable. The number of instruments are assigned such that $K = 4, K^* = 5, K_* = 3, K_1 = 2$, where x_{it-1} is the K_3-variable and $K_* = 3$ follows from $\{y_{i.}^{(1)}, y_{i.}^{(2)}, x_{i.}\}$.

There are 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 $(i = 1, \cdots, N)$. In order to deal with this problem, there have been several statistical procedures developed for the estimating equations with individual effects. (See Hsiao (2003), Arellano (2003) and Baltagi (2005) for the details.) Second, some of the known 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 for a large T. Third, although many previous studies has focused on specific reduced models, when we have endogenous variables in the structural equations of interest, the standard estimation methods have serious drawbacks as Akashi and Kunitomo (2010) have discussed, for instance. Since we can sweep out the source of correlations among the lagged endogenous variables and heterogeneity of individual by using the filtering procedure, however, wet cannot remove the simultaneity at period t by that procedure.

Instead of refining the traditional estimation methods, we shall develop a new estimation procedure which may overcome these problems at the same time by applying the panel limited information maximum likelihood (PLIML) estimation method. The asymptotic properties of the LIML estimation method for estimating structural equations including its asymptotic optimality have been recently investigated by Anderson et al. (2008a,b) when there are *many instruments*. We shall extend their analysis to the PLIML estimation method in the dynamic panel structural equations 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 estimation method, however, first we shall propose to use the filtering procedure for our over-identified model, which is the data transformation. There are two filtering procedures in both forward or/and backward directions of time and remove their individual effects before estimation. We shall focus on both the forward-filtering procedure and the double-filtering procedure in our analysis.

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 Alvarez and Arellano (2003) such that $\mathbf{A}_f\mathbf{A}_f^{'} = \mathbf{I}_{T-1}$, $\boldsymbol{\iota} = (1,...,1)^{'}$ 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 to the random variables of $\mathbf{y}_i^{(1)}$, $\mathbf{Y}_i^{(2)}$, and $\mathbf{Z}_{i(-1)}^{(1)}$, and 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 $\mathbf{Z}_i^{(1,f)} = (\mathbf{z}_{it-1}^{(1,f)'})$. Then, for example, we have

$$\mathbf{y}_{it}^{(2,f)} = c_t\Big[\mathbf{y}_{it}^{(2)} - \frac{1}{T-t}(\mathbf{y}_{it+1}^{(2)} + \cdots + \mathbf{y}_{iT}^{(2)})\Big] \tag{2.13}$$

where $c_t^2 = (T-t)/(T-t+1)$ for $t = 1,...,T-1,\ T\geq 2$.

By using the forward-filtered variables, we re-write for $t = 1,\cdots,T-1$ as

$$y_{it}^{(1,f)} = \boldsymbol{\beta}_2^{'}\boldsymbol{y}_{it}^{(2,f)} + \boldsymbol{\gamma}_1^{'}\boldsymbol{z}_{it-1}^{(1,f)} + u_{it}^{(f)}\ , \tag{2.14}$$

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 $\mathbf{u}_i = (u_{it})$, but also we have the relation that $\mathcal{E}[\boldsymbol{z}_{it}^{(1,f)}u_{it}^{(f)}] \neq \mathbf{0}$, consequently.

On the other hand, we can also apply the backward operator $\mathbf{A}_b$, which is the $(T-1)\times T$ lower triangular matrix as used for Hayakawa (2006). The procedure removes the individual effects from the instrumental variables. Then we denote the transformed instrumental variables as $\mathbf{Z}_{i(-1)}^{(b)} = (\mathbf{z}_{it-1}^{(b)'})$ and we set

$$\mathbf{z}_{it-1}^{(1,b)} = b_t\Big[\mathbf{z}_{it-1}^{(1)} - \frac{1}{t}(\mathbf{z}_{it-2}^{(1)} + \cdots + \mathbf{z}_{i0}^{(1)} + \mathbf{z}_{i(-1)}^{(1)})\Big]\ , \tag{2.15}$$

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

We notice that the forward-filtering enables us to make the orthogonal conditions and keeps the homogeneity of second-moments of the disturbances. The backward-filtering removes the individual effects exactly 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

$$\mathbf{Z}_t^{(a)} = \begin{pmatrix} \mathbf{z}_{1(t-1)}^{(a)} & \cdots & \mathbf{z}_{N(t-1)}^{(a)} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_{10}^{(a)} & \cdots & \mathbf{z}_{N0}^{(a)} \end{pmatrix}^{'}\ ,\ \mathbf{Z}_t^{(b)} = \Big(\mathbf{z}_{1(t-1)}^{(b)},\cdots,\mathbf{z}_{N(t-1)}^{(b)}\Big)^{'}\ , \tag{2.16}$$

where $\mathbf{z}_{it-1}^{(a)}$ is the $K_*\times 1$ vector such that $\mathbf{z}_{it-1}^{(a)} = \mathbf{J}_{K_*}\mathbf{z}_{it-1}$, where the selection matrix $\mathbf{J}_{K_*}$ chooses the nearest lagged variables to $t-1$ as each autoregressive

variable. The reduction K to K_* is needed to be full rank of $(\mathbf{Z}_t^{(a)\prime}\mathbf{Z}_t^{(a)})$. Hence $\mathbf{Z}_t^{(a)}$ is the $N \times (K_* t)$ and $\mathbf{Z}_t^{(b)}$ is the $N \times K$ matrix, we consider that these instrumental choices correspond to the following methods,
(a) At period t we use all available lagged variables after applying the forward-filtering to the structural equation as suggested by Alvarez and Arellano (2003),
(b) At period t we use the only lagged variables included in the reduced form after applying the backward-filtering to all instruments.
In this formulation the orthogonal conditions at period t can be written as

$$\mathcal{E}\left[\mathbf{z}_{is}^{(a)} u_{it}^{(f)}\right] = \mathbf{0} \ (0 \le s < t) \ , \ \mathcal{E}\left[\mathbf{z}_{it-1}^{(b)} u_{it}^{(f)}\right] = \mathbf{0} \ . \tag{2.17}$$

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

$$(a) \ \frac{K_* T(T-1)}{2NT} \underset{N,T\to\infty}{\rightarrow} c_a = (\frac{K_*}{2}) \lim_{N,T\to\infty} (\frac{T}{N}) \ . \tag{2.18}$$

$$(b) \ \frac{K(T-1)}{N_0 T} \underset{T\to\infty}{\rightarrow} c_b = \frac{K}{N_0} \ , \tag{2.19}$$

where we use the notation N_0 to be a fixed integer. Then we shall consider the asymptotic behaviors of estimators when these sequences of ratios can be reasonable approximations as the large-K asymptotics under panel structural equation models provided $K < \infty$, $N_0 < \infty$. When the order of instruments 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 can be regarded as fixed $N_0 < \infty$. The double asymptotics could worsen some approximations on the limiting distributions of estimators, since it is constructed as a further approximation from the fixed T or the fixed N asymptotics.

2.3 The LIML and GMM Estimation

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

$$\mathbf{Y}_t^{(f)\prime} = \left(\mathbf{y}_{1t}^{(f)}, \cdots, \mathbf{y}_{Nt}^{(f)}\right), \ \mathbf{Z}_t^{(1,f)\prime} = \left(\mathbf{z}_{1t}^{(1,f)}, \cdots, \mathbf{z}_{Nt}^{(1,f)}\right),$$

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

$$\mathbf{G}^{(f)} = \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)\prime} \\ \mathbf{Z}_{t-1}^{(1,f)\prime} \end{pmatrix} \mathbf{M}_t \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{2.20}$$

and

$$\mathbf{H}^{(f)}=\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{pmatrix}[\mathbf{I}_N-\mathbf{M}_t]\left(\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)}\right)\ , \tag{2.21}$$

where $\mathbf{M}_t=\mathbf{M}_t^{(a)}$ or $\mathbf{M}_t^{(b)}$,

$$\mathbf{M}_t^{(a)}=\mathbf{Z}_t^{(a)}(\mathbf{Z}_t^{(a)'}\mathbf{Z}_t^{(a)})^{-1}\mathbf{Z}_t^{(a)'}\ \ ,\ \ \mathbf{M}_t^{(b)}=\mathbf{Z}_t^{(b)}(\mathbf{Z}_t^{(b)'}\mathbf{Z}_t^{(b)})^{-1}\mathbf{Z}_t^{(b)'} \tag{2.22}$$

Then the LIML estimator $\hat{\boldsymbol{\theta}}_{LI}^{(.)}=(\hat{\boldsymbol{\beta}}_{2.LI}^{'},\hat{\boldsymbol{\gamma}}_{1.LI}^{'})'$ of $(1,-\boldsymbol{\beta}_2^{'},-\boldsymbol{\gamma}_1^{'})'=(1,-\boldsymbol{\theta}')'$ is defined by

$$\left[\frac{1}{n}\mathbf{G}^{(f)}-\lambda_n\frac{1}{q_n}\mathbf{H}^{(f)}\right]\begin{bmatrix}1\\ -\hat{\boldsymbol{\theta}}_{LI}^{(.)}\end{bmatrix}=\mathbf{0}\ , \tag{2.23}$$

where $n=N(T-1)$, $q_n=n-r_n$ and λ_n is the smallest root of

$$\left|\frac{1}{n}\mathbf{G}^{(f)}-l\frac{1}{q_n}\mathbf{H}^{(f)}\right|=0\ . \tag{2.24}$$

In the above definition we have used the notation $\hat{\boldsymbol{\theta}}_{LI}^{(.)}=\hat{\boldsymbol{\theta}}_{LI}^{(a)}$ in the case of using $\mathbf{M}_t^{(a)}$ and $\hat{\boldsymbol{\theta}}_{LI}^{(b)}$ in the case of using $\mathbf{M}_t^{(b)}$.
The solution to (2.23) gives the minimum of the variance ratio

$$\mathrm{VR}_n=\frac{\left[1,-\boldsymbol{\theta}'\right]\mathbf{G}^{(f)}\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}}{\left[1,-\boldsymbol{\theta}'\right]\mathbf{H}^{(f)}\begin{bmatrix}1\\ -\boldsymbol{\theta}\end{bmatrix}}\ . \tag{2.25}$$

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

$$[\mathbf{0},\mathbf{I}_{G_2+K_1}]\sum_{t=1}^{T-1}\begin{bmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{bmatrix}\mathbf{M}_t\left[\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)}\right]\begin{bmatrix}1\\ -\hat{\boldsymbol{\theta}}_{GM}^{(.)}\end{bmatrix}=\mathbf{0}. \tag{2.26}$$

and define $\hat{\boldsymbol{\theta}}_{GM}^{(a)}$ and $\hat{\boldsymbol{\theta}}_{GM}^{(b)}$ accordingly. It minimizes the numerator of the variance ratio (2.25). 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 models with individual effects.

3 Asymptotic Properties of the LIML and GMM Estimators

3.1 Asymptotic Distributions

We shall derive the limiting distributions of the LIML and the GMM estimators when we have two sequences on N, T, K and r_n. In order to do that we make a set of assumptions on the moments of disturbsnces and the dynamics of the underlying process.

(A1) $\{\mathbf{v}_{it}^*\}$ $(i = 1, ..., N;\ t = 1, ..., T)$ 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 $\mathcal{E}[\|\mathbf{v}_{it}^*\|^8]$ exsits. The random vectors $\boldsymbol{\pi}_i^*$ are i.i.d. across individuals.

(A2) The initial observation satisfies

$$\mathbf{z}_{i0} = (\mathbf{I}_{K^*} - \boldsymbol{\Pi}^*)^{-1} \boldsymbol{\pi}_i^* + \mathbf{w}_{i0} \qquad (i = 1, ..., N),$$

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 we can represent $\mathbf{w}_{i0} = \sum_{j=0}^{\infty} \boldsymbol{\Pi}^{*j} \mathbf{v}_{i(0-j)}^*$. All roots λ_k of

$$|\boldsymbol{\Pi}^* - \lambda \mathbf{I}_{K^*}| = 0 \tag{3.27}$$

satisfy the stationarity condition $|\lambda_k| < 1$ $(k = 1, ..., K^*)$.

The assumptions (A1) and (A2) are analogue to the conditions used by Alvarez and Alrellano (2003). They imply that the underlying processes for $\{\mathbf{y}_{it}\}$ are stationary and we have sufficient moment conditions. To state our main theoretical results in a concise way, we prepare some notations 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 $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$ and

$$\mathbf{u}_{it}^{\perp} = [\mathbf{0}, \mathbf{I}_{G_2}] \left[\mathbf{I}_{1+G_2} - Cov(\mathbf{v}_{it}, u_{it}) \frac{u_{it}}{\sigma^2} \right], \tag{3.28}$$

$$\boldsymbol{\Phi}^* = \mathbf{D}' \mathbf{J}_K' \mathcal{E}[\mathbf{w}_{i(t-1)} \mathbf{w}_{i(t-1)}'] \mathbf{J}_K \mathbf{D}, \tag{3.29}$$

$$\mathbf{D} = \mathbf{J}_{K_1, K_2} \left[\boldsymbol{\Pi}_2, \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix} \right], \ \mathbf{J}_K' = [\mathbf{I}_K, \mathbf{O}_{K \times K_3}]. \tag{3.30}$$

By defining the underlying stationary process $\{\mathbf{w}_{it}\}$ which satisfies

$$\mathbf{w}_{it} = \boldsymbol{\Pi}^* \mathbf{w}_{it-1} + \mathbf{v}_{it}^*, \tag{3.31}$$

then the conditions of (A1) and (A2) imply that it has a solution of stationary vector process.

First, we consider the case (a) when we take the forward-filtering procedure and then apply the LIML and the GMM estimation under the sequence of (a). We denote $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and in this case we have the next result whose proof will be in Section 6.

Theorem 3.1 : Suppose Assumptions (A1) and (A2) hold. Consider the double asymptotics $N, T \to \infty$ and assume that $0 \leq K_* \lim_{N,T\to\infty}(T/N) < 1$.
(i) For $c_a = 0,\ 0 \leq \lim_{N,T\to\infty}(T^3/N) = d_a < \infty$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{b}_0^{(a)}, \sigma^2\boldsymbol{\Phi}^{*-1}) \ , \tag{3.32}$$

where

$$\mathbf{b}_0^{(a)} = (\frac{d_a^{1/2}K_*}{2})\boldsymbol{\Phi}^{*-1}\begin{pmatrix} \mathbf{J}_{G_2}^{'}\boldsymbol{\Omega\beta} \\ \mathbf{0} \end{pmatrix} \ , \ \mathbf{J}_{G_2}^{'} = [\mathbf{0}, \mathbf{I}_{G_2}] \ . \tag{3.33}$$

(ii) For $c_a = 0$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{*-1}) \ . \tag{3.34}$$

(iii) For $0 < c_a \leq 1/2$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(a)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{b}_c^{(a)}, \boldsymbol{\Psi}^{*(a)}) \tag{3.35}$$

and

$$\boldsymbol{\Phi}^{*-1}[\sigma^2\boldsymbol{\Phi}^* + \begin{pmatrix} \mathbf{I}_{G_2} \\ \mathbf{O} \end{pmatrix}(c_{*a}[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega\beta\beta}'\boldsymbol{\Omega}]_{22} + \boldsymbol{\Xi}_4)(\mathbf{I}_{G_2}, \mathbf{O}) + \boldsymbol{\Xi}_3^{(a)} + \boldsymbol{\Xi}_3^{(a)'}]\boldsymbol{\Phi}^{*-1},$$

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

$$\boldsymbol{\Xi}_3^{(a)'} = \begin{pmatrix} \frac{1}{1-c_a}E[u_{it}^2\mathbf{u}_{it}^{\perp}]\lim_{N,T\to\infty}\frac{1}{NT}\sum_{t=2}^{T-1}\mathcal{E}[\mathbf{d}_t^{(a)'}\mathbf{W}_{t-1}]\mathbf{JD} \\ \mathbf{O} \end{pmatrix} , \tag{3.36}$$

$$\boldsymbol{\Xi}_4^{(a)} = (\frac{1}{1-c_a})^2\mathcal{E}[(u_{it}^2 - \sigma^2)\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}](\lim_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{d}_t^{(a)'}\mathbf{d}_t^{(a)}] - c_a^2), \tag{3.37}$$

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

$$\mathbf{b}_c^{(a)} = -(\frac{K_*}{2})^{1/2}\frac{c_a^{1/2}}{(1-c_a)}\boldsymbol{\Phi}^{*-1}\mathbf{D}'\mathbf{J}_K^{'}(\mathbf{I}_{K^*} - \boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*u_{it}] \ , \tag{3.38}$$

provided that $\boldsymbol{\Xi}_3^{(a)}$ and $\boldsymbol{\Xi}_4^{(a)}$ are well-defined.

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 due to the presence of the endonenous variables. This result agree with the one by Anderson et al. (2008b) for a linear structural equation model with many instruments. The asymptotic bias 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 apply the backward-filtering procedure to the set of instrumental variables including the lagged endogenous variables and reduce the number of orthogonal conditions as the sequence of (b). 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 3.2 : Suppose Assumptions (A1) and (A2) hold. Let $T \to \infty$ and $K/N = c_b$.
(i) For $c_b = 0$ or $N \to \infty$, $0 \leq \lim_{N,T\to\infty}(T/N) = d_b < \infty$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{GM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{b}_0^{(b)}, \sigma^2\boldsymbol{\Phi}^{*-1}) , \tag{3.39}$$

where

$$\mathbf{b}_0^{(b)} = (d_b^{1/2}K)\boldsymbol{\Phi}^{*-1}\begin{pmatrix} \mathbf{J}_{G_2}\boldsymbol{\Omega\beta} \\ \mathbf{0} \end{pmatrix} . \tag{3.40}$$

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

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{LM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{*-1}) . \tag{3.41}$$

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

$$\sqrt{N_0T}\left(\hat{\boldsymbol{\theta}}_{LM}^{(b)} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^{*(b)}) \tag{3.42}$$

and

$$\boldsymbol{\Phi}^{*-1}[\sigma^2\boldsymbol{\Phi}^* + \begin{pmatrix} \mathbf{I}_{G_2} \\ \mathbf{O} \end{pmatrix} (c_{*b}[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega\beta\beta'\Omega}]_{22} + \boldsymbol{\Xi}_4)(\mathbf{I}_{G_2}, \mathbf{O}) + \boldsymbol{\Xi}_3^{(b)} + \boldsymbol{\Xi}_3^{(b)'}]\boldsymbol{\Phi}^{*-1},$$

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

$$\boldsymbol{\Xi}_3^{(b)'} = \begin{pmatrix} \frac{1}{1-c_b}\mathcal{E}[u_{it}^2\mathbf{u}_{it}^{\perp}]\lim_{T\to\infty}\frac{1}{N_0T}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{d}_t^{(b)'}\mathbf{W}_{t-1}]\mathbf{J}_K\mathbf{D} \\ \mathbf{O} \end{pmatrix} , \tag{3.43}$$

$$\boldsymbol{\Xi}_4^{(b)} = (\frac{1}{1-c_b})^2\mathcal{E}[(u_{it}^2 - \sigma^2)\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}](\lim_{T\to\infty}\frac{1}{N_0T}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{d}_t^{(b)'}\mathbf{d}_t^{(b)}] - c_b^2) , \tag{3.44}$$

and $\mathbf{d}_t^{(b)} = diag(\mathbf{M}_t^{(b)})\iota_N$, $\mathbf{W}_{t-1} = (\mathbf{w}_{1(t-1)}, ..., \mathbf{w}_{N(t-1)})'$ is the $N \times K^*$ matrix consisting of $\{\mathbf{w}_{it}\}$, provided that $\Xi_3^{(b)}$ and $\Xi_4^{(b)}$ are well-defined.

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 it has the asymptotic normality while the GMM estimator is inconsistent.

We notice that $\mathbf{\Phi}^*$ are 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 asymptotics holds, then the second term of the asymptotic covariance becomes large, so that the *large*-K 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 and attains the asymptotic efficiency bound $\sigma^2 \mathbf{\Phi}^{*-1}$, which is the standard bound when $\boldsymbol{\pi}_i^* = \mathbf{0}$ $(i = 1, \cdots, N)$ and T is a fixed integer.

In the general case, the asymptotic covariance of the LIML estimator depends on the third and fourth order moments of disturbance terms $\mathbf{v}_{it}$. When the random vectors are followed by the class of elliptically contoured distribution EC($\mathbf{\Omega}$) (see Section 2.7 of Anderson (2003)), for instance, we could simplify the explicit formula considerably because the third order moments are zeros and there is a simple expression on the fourth order moments. When the disturbances are normally distributed in particular, $\Xi_3 = \mathbf{O}$ and $\Xi_4 = \mathbf{O}$. In the more general cases we could expect that the contributions from these terms are often negligible numerically.

If the third and fourth order components are negligible, we may compare the asymptotic covariance by the magnitude of c_{*a} and c_{*b}. Although the relation of $c_a > c_b$ holds in the general cases, the relative efficiency of $\hat{\boldsymbol{\theta}}_{LI}^{(b)}$ to $\hat{\boldsymbol{\theta}}_{LI}^{(a)}$ depends on the correct knowledge of the reduced form lag structure. In this sense $\hat{\boldsymbol{\theta}}_{LI}^{(a)}$ may be regarded as the most conservative estimation method as to the choice of instrumental variables.

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 $\mathbf{H}^{(f)}$, and the bias corrected estimator caused by the forward filtering such as the one proposed by Hahn and Kuersteiner (2002). We shall consider a class of estimators which are some functions of these two matrices and then we have some results on the asymptotic optimality under a set of assumptions.

Theorem 3.3 : In the panel structural equations model of (2.2) and (2.3), define the class of consistent estimators for $\boldsymbol{\theta} = (1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})'$ by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\gamma}}_1 \end{pmatrix} = \phi(\mathbf{G}^{(f)}\ ,\ \mathbf{H}^{(f)})\ , \tag{3.45}$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices $(1/n)\mathbf{G}^{(f)}/n$ and $(1/q_n)\mathbf{H}^{(f)}$.
(i) Then either under the conditions of *Theorem 3.1* or *Theorem 3.2*, as $T \to \infty$ with $c_a = 0$ or $c_b = 0$,

$$\sqrt{NT}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{pmatrix}\right] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})\ , \tag{3.46}$$

where

$$\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^* \tag{3.47}$$

and $\boldsymbol{\Psi}^*$ is given in *Theorem 3.1* and *Theorem 3.2*. The LIML estimator and the bias-adjusted GMM estimator attain the asymptotic bound.
(ii) When $0 < c_a < 1$ or $0 < c_b < 1$ in *Theorem 3.1* or *Theorem 3.2*, assume $\boldsymbol{\Xi}_3^{(\cdot)} = \mathbf{O}$ and $\boldsymbol{\Xi}_4^{(\cdot)} = \mathbf{O}$ in addition to their conditions. Then

$$\sqrt{NT}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{pmatrix} - \frac{1}{\sqrt{NT}}\mathbf{b}^{(f)}\right] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})\ , \tag{3.48}$$

where the asympptotic bias $\mathbf{b}^{(f)}$ 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 equations. It can be regarded as an extension of *Theorem 4* of Anderson et al. (2008b) for the linear structural equations of the simultaneous equation systems. The simple sufficient condition for $\boldsymbol{\Xi}_3^{(\cdot)} = \mathbf{O}$ and $\boldsymbol{\Xi}_4^{(\cdot)} = \mathbf{O}$ is the Gaussianity of disturbances. These conditions in *Theorem 3.3* can be further relaxed to the Elliptically Contours (EC) distributions with an additional notation. Because of individual effects in the panel structural equations and the filtering problem, there are some complications on the asymptotic optimality of estimators beyond the results of Anderson et al. (2008b).

3.3 An Extension of PLIML with Heterocedasticity

One of important problems in panel econometric studies has been the heterogeneity among a large number of individuals in data sets. Then it is important to investigate the effects of persistently heteroscedastic disturbances over individuals [4]. Kunitomo (2008) has extended the LIML estimation to the case of heteroscadastic disturbances in structural equation econometric models under the condition

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Omega}_i \xrightarrow{p} \mathbf{\Omega}\ , \tag{3.49}$$

where $\mathbf{\Omega}_i$ is the covariance matrix of $\mathbf{v}_{it}$ $(i = 1, \cdots, N; t = 1, \cdots, T)$ and $\mathbf{\Omega}$ is a positive definite (constant) matrix. Hence we have

$$\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \xrightarrow{p} \sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} > 0\ . \tag{3.50}$$

In the present situation an asymptotically optimal modification of LIML (AOM-LIML) estimation can be constructed as follows. For $N \times N$ matrices $\mathbf{M}_t = (m_{t.ij}) = \mathbf{Z}_t(\mathbf{Z}_t^{'}\mathbf{Z}_t)^{-1}\mathbf{Z}_t^{'}$, we construct $\mathbf{M}_{t.m} = (m_{t.ij}^*)$ and $\mathbf{Q}_{t.m} = (q_{t.ij}^*) = \mathbf{I}_N - \mathbf{M}_{t.m}$ such that $m_{t.ij}^* = m_{t.ij}^*$ $(i \neq j)$ and $m_{t.ii}^* - c = o_p(1)$ $(i, j = 1, \cdots, N)$ for $c = c_a$ or $c = c_b$. Then we define two $(K_1 + 1 + G_2) \times (K_1 + 1 + G_2)$ matrices [5] by

$$\mathbf{G}^{(f.m)} = \sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{pmatrix}\mathbf{M}_{t.m}\left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{3.51}$$

and

$$\mathbf{H}^{(f.m)} = \sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{pmatrix}[\mathbf{I}_N - \mathbf{M}_{t.m}]\left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right)\ , \tag{3.52}$$

where $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ or $\mathbf{M}_t^{(b)}$.

By using $\mathbf{G}^{(f.m)}$ and $\mathbf{H}^{(f.m)}$, we define a class of asymptotically optimal modifications of the PLIML estimator (we call it as AOM-PLIML) such that $\hat{\boldsymbol{\theta}}_{MLI}$ $(= (\hat{\boldsymbol{\beta}}_{2.MLI}^{'})^{'}, \hat{\boldsymbol{\gamma}}_{1.MLI}^{'})^{'})$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}_2^{'}, \boldsymbol{\gamma}_1^{'})^{'}$ is the solution of

$$\left[\frac{1}{n}\mathbf{G}^{(f.m)} - \frac{1}{q_n}\lambda_n\mathbf{H}^{(f.m)}\right]\begin{bmatrix}1\\ -\hat{\boldsymbol{\theta}}_{MLI}\end{bmatrix} = \mathbf{0}\ , \tag{3.53}$$

[4] The definitions of *Weak Heteroscedasticity* and *Persistent Heteroscedasticity* are given in Kunitomo (2008).

[5] We impose the condition that $\mathbf{G}^{(f.m)}$ is a positive definite matrix. If it were not positive definite, we need to modify $\mathbf{G}^{(f)}$ further although it would rarely occur. See Kunitomo (2008) for the detail.

where $q_n = n - r_n$ (> 0) and λ_n is the (non-negative) smallest root of

$$|\frac{1}{n}\mathbf{G}^{(f.m)} - l\frac{1}{q_n}\mathbf{H}^{(f.m)}| = 0\ . \tag{3.54}$$

When N and T are large, the AOM-PLIML estimator is consistent and it has the asymptotic normality under a set of assumptions. There are two important consequences of this modification. First, the AOM-PLIML estimator may have less bias than the LIML estimator. Second, the covariance matrix of the limiting distribution of the LIML estimator has the form

$$\boldsymbol{\Phi}^{**-1}\left[\boldsymbol{\Psi}_1^* + \begin{pmatrix}\mathbf{I}_{G_2}\\ \mathbf{O}\end{pmatrix}\boldsymbol{\Psi}_2^*(c)(\mathbf{I}_{G_2}, \mathbf{O})\right]\boldsymbol{\Phi}^{**-1}, \tag{3.55}$$

where $\boldsymbol{\Phi}^{**}$, $\boldsymbol{\Psi}_1^*$ and $\boldsymbol{\Psi}_2^*(c)$ are defined as in *Theorem 1* of Kunitomo (2008).
It is important to notice that the quantities used for its limiting distribution need more complex notations than the homoscedastic situation due to the possible (persistent) heteroscedasticity while the resulting expressions are free from the third and fourth moments of disturbance terms. Thus it may be useful to use the AOM-PLIML estimation in some applications. Also *Theorem 3.3* implies that it attains an asymptotic bound in a class of estimators and it is not possible to improve the AOM-LIML estimation. Since it may be straightforward to investigate the asymptotic properties of the AOM-PLIML estimation as Kunitomo (2008), we have omitted the detail. The finite sample properties of the AOM-LIML estimator are currently under investigation.

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 analogues.) Hence we need to investigate the distribution functions of several estimators in a systematic way.

In our experiments we took *Example 2* ($K = 4, K_* = 3, K^* = 5, K_1 = 2, G_2 = 1$) in Section 2 as a typical example [6]. In *Example 2* we first set the unknown parameters such as $(\beta_2, \gamma_{11}) = (.5, .5), \gamma_{12} = .3$, and $(\omega_{11}, \omega_{12}, \omega_{22}) = (1.0, .3, 1.0), (1.5, 1.0, 1.0)$. 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 (2010).

[6] We have used *Example 1* in Akashi and Kunitomo (2010) to investigate the case of (a) in more details. *Example 1* can be regarded as a simple case of *Example 2*.

Then we generate large number of normal random variables by simulations and calculate the empirical distribution functions of the GMM and LIML estimators in the normalized form. 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, Kunitomo and Matsushita (2005, 2008a). We shall report only the results for $(N, T) = (100, 25), (100, 50)$ and $(200, 50)$ as the typical cases among a large number of our simulations.

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

$$\frac{\sqrt{NT}}{\sigma}\begin{bmatrix} 1/\sqrt{\phi^{11}} & 0 \\ 0 & 1/\sqrt{\phi^{22}} \end{bmatrix}\begin{bmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{\gamma}_1 - \gamma_1 \end{bmatrix}, \tag{4.56}$$

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

$$\sqrt{NT}\begin{bmatrix} \psi_{11}^{-1/2} & 0 \\ 0 & \psi_{22}^{-1/2} \end{bmatrix}\left[\begin{pmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{\gamma}_1 - \gamma_1 \end{pmatrix} - \frac{1}{\sqrt{NT}}\mathbf{b}\right], \tag{4.57}$$

where $\mathbf{b}$ is the asymptotic bias term, ψ_{11} and ψ_{22} are the (1,1)-th element and the (2,2)-th element of $\mathbf{\Psi}^{*}$, respectively. We have chosen these standardizations because of the forms for the limiting distribution of the LIML estimator in *Theorem 3.1* and *Theorem 3.2*. We may call the classical case when $c = 0$ ($c = c_a$ or c_b) and $c \neq 0$ as the general case.

Since Akashi and Kunitomo (2010) have given many figures on case of (a) with the forward-filtering procedure, we only give some cases as Figures 9-12. We have shown the estimated distribution functions of the GMM and the LIML estimators of (β_2, γ_1) and we have confirmed the findings of Akashi and Kunitomo (2010) in a more simple case. That is, the GMM estimator is badly biased when N and T are large while the LIML estimator is almost median-unbiased. However, the normalization by the limiting covariance matrix of the LIML estimator when $c = 0$ is not appropriate. This aspect can be easily observed because the circles in figures are the standard normal distribution function N(0,1).

For the case of (b) with the backward-filtering procedure, we have shown the estimated distribution functions of the GMM and LIML estimators of β_2 and γ_1 as Figures 1-8 among many results. Form these figures first we can observe that the GMM estimator is often biased when N and T are large while the LIML estimator is almost median-unbiased. Then it may be important to notice that the bias correction of the GMM estimator sometimes works well, but it is not always the case. Secondly, the normalization by the limiting covariance matrix of the LIML estimator when $c = 0$ is often not appropriate and we can see it because of the circles in figures as the standard normal distribution function N(0,1). Since the normal approximations based on the general case $c \neq 0$, it is important to use the variance formulas in Section 3.

From these figures we have shown, we have confirmed that the limiting normal distributions approximate the finite sample distribution functions of the LIML estimator quite well as *Theorem 3.1* and *Theorem 3.2* we have derived.

5 Conclusions

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

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. Since the approximation of its limiting distribution embodies the influence of the number of instrumental variables automatically, our method gives an unified approach for solving practical problems with panel data consisting of various combination of N, T and K.

Furthermore, we have suggested a class of asymptotically optimal modification of the PLIML estimator. Since it may improve the asymptotic properties of the LIML estimator, we are currently investigating its finite sample properties.

In this paper we have examined the effects of possible filtering procedures. When we use the forward-filtering, the GMM estimator is badly biased while the LIML estimator is almost median-unbiased. If we use the backward-filtering to instruments, the GMM estimator is often biased, but its magnitude can be significantly reduced. This finding may lead to an interpretation that we should not use many instruments and just use the GMM estimator with the backward-filtered instruments in practical situations. However, it is the case only when we had known the true lag-structure in advance. Since we often do not know the precise form of lag structures in the simultaneous equations, it may be fair to conclude that the LIML estimation has the asymptotic robustness in both cases of (a) and (b) while the GMM estimation does not have such robustness.

Finally, as we have mentioned, in a companion paper to the present one Akashi and Kunitomo (2010) have investigated the finite sample properties of alternative estimation methods, the WG (Within Groups), the GMM and the PLIML estimators in a simpler setting, based on a large set of Monte Carlo experiments. Although they have used a particular case of dynamic panel simultaneous equations model and the formulation of forward filtering procedure, we have confirmed that their results are quite relevant for more general panel structural equations as we have referred in Section 4. Thus we conclude that the traditional LIML estimation method is quite useful and relevant in dynamic panel econometric modeling.

6 Mathematical Details

In this section we give the proofs of *Theorems* in Section 3. The method of proofs are similar to those used in Alvarez and Arellano (2003), Anderson, Kunitomo and Matsushita (2008) and Akashi and Kunitomo (2010). When we use the generic notations of $(\mathbf{M}_t, \mathbf{N}_t, c, c_*)$, the relevant derivation is valid for the each case of $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$ under the corresponding asymptotics of *Theorem 3.1* and *3.2*, respectively. We shall use $\mathbf{J}'$ for $\mathbf{J}_K'$ below for the sake of convenience.

Some derivations of the asymptotic properties of estimators have been given by Akashi and Kunitomo (2010) when $G_2 = 1$. Since it is straight-forward to extend their analysis to the general cases, we shall freely refer to their results. The derivation of the asymptotic distribution of the GMM estimator is an example.

Derivations of Theorem 3.1 and 3.2 :

Since the derivations of our results are rather lengthy, we shall divide them into several steps.

[**Step 1**] : We drive the probability limit in Step 1 and then the limiting distribution of the LIML estimator at the next step. Substitution of (2.14) into (2.20) yields

$$\mathbf{G}^{(f)} = \mathbf{G}^{(f,1)} + \mathbf{G}^{(f,2)} + \mathbf{G}^{(f,2)'} + \mathbf{G}^{(f,3)} \ , \tag{6.58}$$

where

$$\begin{aligned}
\mathbf{G}^{(f,1)} &= \mathbf{D}^{*'} \sum_{t=2}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{M}_t \mathbf{Z}_{t-1}^{(f)} \mathbf{D}^* \ , \\
\mathbf{G}^{(f,2)} &= \mathbf{D}^{*'} \sum_{t=2}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{M}_t (\mathbf{V}_t^{(f)}, \mathbf{O}) \ , \\
\mathbf{G}^{(f,3)} &= \sum_{t=2}^{T-1} \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{pmatrix} \mathbf{M}_t (\mathbf{V}_t^{(f)}, \mathbf{O}) \ ,
\end{aligned}$$

$\mathbf{V}_t^{(f)'} = (\mathbf{v}_{t1}^{(f)}, \cdots, \mathbf{v}_{tN}^{(f)})$, $\mathbf{v}_{tj}^{(f)}$ $(j = 1, \cdots, N)$ are the corresponding forward-filtered disturbances of $\mathbf{v}_{tj}$, and a $K \times (1 + G_1 + K_1)$ matrix

$$\mathbf{D}^* = \mathbf{D} \left[\ \boldsymbol{\theta}, \mathbf{I}_{G_2+K_1} \ \right] \ , \tag{6.59}$$

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

$$\frac{1}{n}\mathbf{G}^{(f)} \xrightarrow{p} \mathbf{G}_0 = \begin{bmatrix} \boldsymbol{\theta}' \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \boldsymbol{\Phi}^* \left[\ \boldsymbol{\theta}, \mathbf{I}_{G_2+K_1} \ \right] + c \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \tag{6.60}$$

and

$$\frac{1}{q_n}\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0 = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \ , \tag{6.61}$$

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

$$\begin{aligned} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\theta}' \\ \mathbf{I}_{G_2+K_1} \end{pmatrix} \mathbf{D}'\mathbf{Z}_{t-1}^{(f)'} + \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O}' \end{pmatrix} \\ &= \mathbf{D}^{*'}\mathbf{Z}_{t-1}^{(f)'} + \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O}' \end{pmatrix} \ (, say). \end{aligned} \tag{6.62}$$

Then we can show that

$$\frac{1}{n}\mathbf{G}^{(f,2)} = \frac{1}{NT}\mathbf{D}^{*'} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t(\mathbf{V}_t^{(f)}, \mathbf{O}) \xrightarrow{p} \mathbf{O}_{G+K_1}. \tag{6.63}$$

It is because $(1/\sqrt{NT}) \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)} \xrightarrow{p} O_p(1) + O(1)$ by the same arguments as used for $(1/\sqrt{NT}) \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{u}_t^{(f)} \xrightarrow{p} O_p(1) + O(1)$ in Kunitomo and Akashi (2010).

We write

$$\begin{aligned} \mathbf{Z}_{t-1}^{(f)'} &= c_t[\mathbf{I}_K - \frac{1}{T-t}(\sum_{j=1}^{T-t} \boldsymbol{\Pi}^{*j})]\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}' \\ &= \boldsymbol{\Psi}_t\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}' \ (, say), \end{aligned} \tag{6.64}$$

and we further decompose $(1/n)\mathbf{G}^{(f,1)} = (1/n)\mathbf{D}^{*'} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{Z}_{t-1}^{(f)}\mathbf{D}^*$ as

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t\mathbf{Z}_{t-1}^{(f)} &= \frac{1}{n}\sum_{t=1}^{T-1} \boldsymbol{\Psi}_t\mathbf{W}_{t-1}'\mathbf{M}_t\mathbf{W}_{t-1}\boldsymbol{\Psi}_t' - \frac{1}{n}\sum_{t=1}^{T-1} c_t\boldsymbol{\Psi}_t\mathbf{W}_{t-1}'\mathbf{M}_t\tilde{\mathbf{V}}_{tT} \\ &\quad - \frac{1}{n}\sum_{t=1}^{T-1} c_t\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t\mathbf{W}_{t-1}\boldsymbol{\Psi}_t' + \frac{1}{n}\sum_{t=1}^{T-1} c_t^2\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t\tilde{\mathbf{V}}_{tT}. \end{aligned} \tag{6.65}$$

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

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{T-1} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} \mathbf{\Psi}_t^{'} \qquad (6.66) \\
= & \frac{1}{n}\sum_{t=1}^{T-1} c_t^2 \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} \\
& -\frac{1}{n}\sum_{t=1}^{T-1} \frac{c_t^2}{T-t} \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} (\sum_{j=1}^{T-t} \mathbf{\Pi}^{*j})^{'} - \frac{1}{n}\sum_{t=1}^{T-1} \frac{c_t^2}{T-t} (\sum_{j=1}^{T-t} \mathbf{\Pi}^{*j}) \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} \\
& +\frac{1}{n}\sum_{t=1}^{T-1} (\frac{c_t}{T-t})^2 (\sum_{j=1}^{T-t} \mathbf{\Pi}^{*j}) \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} (\sum_{j=1}^{T-t} \mathbf{\Pi}^{*j})^{'} \\
\xrightarrow{p} & \mathcal{E}[\mathbf{w}_{i(t-1)} \mathbf{w}_{i(t-1)}^{'}] \, .
\end{aligned}
$$

The second and third terms of (6.65) have zero means and their variances to tend to zeros. It is because

$$
\begin{aligned}
& Var[\frac{1}{n}\sum_{t=1}^{T-1} c_t \mathbf{e}_j^{'} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k] \qquad (6.67) \\
= & \frac{1}{N^2T^2} | \sum_{t=1}^{T-1}\sum_{s=1}^{T-1} c_t c_s \mathcal{E}[(\mathbf{e}_j^{'} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k)(\mathbf{e}_k^{'} \tilde{\mathbf{V}}_{sT} \mathbf{M}_s \mathbf{W}_{s-1} \mathbf{\Psi}_s^{'} \mathbf{e}_j)]| \\
\leq & \frac{1}{N^2T^2} \sum_{t=1}^{T-1}\sum_{s=1}^{T-1} \sqrt{c_t^2 \mathcal{E}[(\mathbf{e}_j^{'} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k)^2]} \sqrt{c_s^2 \mathcal{E}[(\mathbf{e}_k^{'} \tilde{\mathbf{V}}_{sT} \mathbf{M}_s \mathbf{W}_{s-1} \mathbf{\Psi}_s^{'} \mathbf{e}_j)^2]},
\end{aligned}
$$

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

$$
\begin{aligned}
& c_t^2 \mathcal{E}[(\mathbf{e}_j^{'} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \tilde{\mathbf{V}}_{tT} \mathbf{e}_k)^2] \qquad (6.68) \\
= & c_t^2 [\mathbf{e}_k^{'} \mathcal{E}[\tilde{\mathbf{v}}_{itT} \tilde{\mathbf{v}}_{itT}^{'}] \mathbf{e}_k] \mathcal{E}[\mathbf{e}_j^{'} \mathbf{\Psi}_t \mathbf{W}_{t-1}^{'} \mathbf{M}_t \mathbf{W}_{t-1} \mathbf{\Psi}_t^{'} \mathbf{e}_j] \\
\leq & c_t^2 [\frac{1}{(T-t)^2} \mathbf{e}_k^{'} \sum_{h=1}^{T-t} \mathbf{\Phi}_h \mathcal{E}[\mathbf{v}_{i0}^{*} \mathbf{v}_{i0}^{*'}] \mathbf{\Phi}_h^{'} \mathbf{e}_k][\mathbf{e}_j^{'} \mathbf{\Psi}_t \mathcal{E}[\mathbf{W}_{t-1}^{'} \mathbf{W}_{t-1}] \mathbf{\Psi}_t^{'} \mathbf{e}_j] \\
= & N (\frac{c_t^2}{T-t})^2 (\mathbf{e}_k^{'} \sum_{h=1}^{T-t} \mathbf{\Phi}_h \mathcal{E}[\mathbf{v}_{i0}^{*} \mathbf{v}_{i0}^{*'}] \mathbf{\Phi}_h^{'} \mathbf{e}_k) \\
& \times (\mathbf{e}_j^{'} [\mathbf{I}_k - \frac{1}{T-t} (\sum_{h=1}^{T-t} \mathbf{\Pi}^{*h})] \mathcal{E}[\mathbf{w}_{i0} \mathbf{w}_{i0}^{'}] [\mathbf{I}_k - \frac{1}{T-t} (\sum_{h=1}^{T-t} \mathbf{\Pi}^{*h})^{\prime}] \mathbf{e}_j) \\
= & O(\frac{N}{T-t}),
\end{aligned}
$$

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

Hence

$$Var[\frac{1}{n_0}\sum_{t=1}^{T-1} c_t \mathbf{e}_j^{'}\mathbf{\Psi}_t\mathbf{W}_{t-1}^{'}\mathbf{M}_t\tilde{\mathbf{V}}_{tT}\mathbf{e}_k] \leq \frac{1}{N_0^2T^2}\sum_{t=1}^{T-1}\sqrt{O(\frac{N_0}{T-t})}\sum_{s=1}^{T-1}\sqrt{O(\frac{N_0}{T-s})} = O(\frac{(\sqrt{T})^2}{N_0T^2}) . \tag{6.69}$$

For the fourth term of (6.65), it expected value is given by

$$\begin{aligned}\mathcal{E}[\frac{1}{n}\sum_{t=1}^{T-1} c_t^2\mathbf{e}_j^{'}\tilde{\mathbf{V}}_{tT}^{'}\mathbf{M}_t\tilde{\mathbf{V}}_{tT}\mathbf{e}_k] &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2\mathcal{E}[tr(\mathbf{M}_t\mathcal{E}_t[\tilde{\mathbf{V}}_{tT}\mathbf{e}_k\mathbf{e}_j^{'}\tilde{\mathbf{V}}_{tT}^{'}])] \\ &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^2 tr(\mathbf{M}_t)\mathcal{E}[\mathbf{e}_j^{'}\tilde{\mathbf{v}}_{itT}\tilde{\mathbf{v}}_{itT}^{'}\mathbf{e}_k] \\ &= O(\frac{1}{NT}\sum_t \frac{tr(\mathbf{M}_t)}{T-t+1}) \end{aligned} \tag{6.70}$$

and it converges to zero in probability. Also 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 $(1/n)\mathbf{G}^{(f,3)}$. By using the fact that $\mathcal{E}_t[\mathbf{v}_{it}^{(f)}\mathbf{v}_{it}^{(f)'}] = \mathbf{\Omega}$, we have

$$\begin{aligned}\mathcal{E}[\frac{1}{n}\sum_{t=1}^{T-1}\mathbf{e}_g^{'}\mathbf{V}_t^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)}\mathbf{e}_h] &= \frac{1}{NT}\mathcal{E}[tr(\mathbf{M}_t\mathbf{V}_t^{(f)}\mathbf{e}_h\mathbf{e}_g^{'}\mathbf{V}_t^{(f)'})] \\ &= \frac{\mathbf{e}_g^{'}\mathbf{\Omega}\mathbf{e}_h}{NT}\sum_{t=1}^{T-1} tr(\mathbf{M_t}) \\ &\to c(\mathbf{e}_g^{'}\mathbf{\Omega}\mathbf{e}_h) \end{aligned} \tag{6.71}$$

as $n \to \infty$.
Moreover, by using $\mathbf{V}_t^{(f)} = (\mathbf{V}_t - \bar{\mathbf{V}}_{tT})/c_t$, we have

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{V}_t^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)} &= \frac{1}{NT}\sum_{t=1}^{T-1} c_t^{-2}\mathbf{V}_t^{'}\mathbf{M}_t\mathbf{V}_t - \frac{1}{NT}\sum_{t=1}^{T-1} c_t^{-2}\mathbf{V}_t^{'}\mathbf{M}_t\bar{\mathbf{V}}_t \\ &\quad -\frac{1}{NT}\sum_{t=1}^{T-1} c_t^{-2}\bar{\mathbf{V}}_t^{'}\mathbf{M}_t\mathbf{V}_t + \frac{1}{NT}\sum_{t=1}^{T-1} c_t^{-2}\bar{\mathbf{V}}_t^{'}\mathbf{M}_t\bar{\mathbf{V}}_t. \end{aligned} \tag{6.72}$$

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

$$Var[\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{e}_g^{'}\mathbf{V}_t^{(f)'}\mathbf{M}_t\mathbf{V}_t^{(f)}\mathbf{e}_h] = \frac{1}{N^2T^2}\sum_{t=1}^{T-1}(1+\frac{1}{T-t})^2 \times O(t) , \tag{6.73}$$

which converges to zero.

The second and third terms of the right-hand side of (6.72) can be evaluated analogously as $\Upsilon_{21n}^{(k)}$ and $\Upsilon_{22n}^{(k)}$, and their variances tend to zeros by using the similar arguments.

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

$$\frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)\prime}\\ \mathbf{Z}_{t-1}^{(1,f)\prime}\end{pmatrix}(\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)}) \tag{6.74}$$

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

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

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

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

Hence

$$\frac{1}{q_n}\mathbf{H}^{(f)} \xrightarrow{p} \frac{1}{1-c}\left[\text{plim}\frac{1}{n}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)\prime}\\ \mathbf{Z}_{t-1}^{(1,f)\prime}\end{pmatrix}(\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)}) - \mathbf{G}_0\right] = \mathbf{H}_0. \tag{6.75}$$

Therefore we have established that $(1/n)\mathbf{G}^{(f)} \xrightarrow{p} \mathbf{G}_0$ and $(1/q_n)\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0$.

[**Step 2**] : By using the convergence results in Step 1, we have

$$\left|\mathbf{\Phi}_\theta + [c-(\text{plim}_{n\to\infty}\lambda_n)]\begin{bmatrix}\mathbf{\Omega} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\right| = 0\ , \tag{6.76}$$

where

$$\mathbf{\Phi}_\theta = \begin{bmatrix}\boldsymbol{\theta}'\\ \mathbf{I}_{G_2+K_1}\end{bmatrix}\mathbf{\Phi}^*\left[\boldsymbol{\theta},\mathbf{I}_{G_2+K_1}\right]. \tag{6.77}$$

By the positive-definiteness of $\mathbf{\Phi}^*$, $\lambda_n \xrightarrow{p} c$, and we have that $\hat{\boldsymbol{\theta}}_{LI} \xrightarrow{p} \boldsymbol{\theta}$ because (2.23) gives

$$\mathbf{\Phi}^*(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0} + o_p(1) \ . \tag{6.78}$$

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 (2.23), it is asymptotically equivalent to

$$[\mathbf{G}_0 - c\mathbf{H}_0]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + \frac{1}{\sqrt{n}}[\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + \frac{1}{\sqrt{n}}[\mathbf{G}_0 - c\mathbf{H}_0]\mathbf{b}_1$$

$$-\frac{1}{\sqrt{q_n}}[c\mathbf{H}_1^{(f)}]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} = o_p(\frac{1}{\sqrt{n}}) \ . \tag{6.79}$$

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

$$\mathbf{\Phi}_\theta \mathbf{b}_1 = [\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1^{(f)}]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + o_p(1) \ . \tag{6.80}$$

Multiplication of (6.79) from the left by $(1, -\boldsymbol{\theta})$ yields

$$\lambda_{1n}^{(f)} = \frac{(1, -\boldsymbol{\theta}')[\mathbf{G}_1^{(f)} - \sqrt{cc_*}\mathbf{H}_1^{(f)}](1, -\boldsymbol{\theta}')'}{(1, -\boldsymbol{\theta}')\mathbf{H}_0(1, -\boldsymbol{\theta}')'} + o_p(1) \ . \tag{6.81}$$

Also the multiplication of (6.80) from the left by $(0, \mathbf{I}_{G_2+K_1})$ and substitution for $\lambda_{1n}^{(f)}$ for (6.80) yields

$$\mathbf{\Phi}^*\sqrt{n}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{2LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1LI} - \boldsymbol{\gamma}_1 \end{bmatrix} \tag{6.82}$$

$$= [\mathbf{0}, \mathbf{I}_{G_2+K_1}]\left[\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1^{(f)}\right]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + o_p(1)$$

$$= [\mathbf{0}, \mathbf{I}_{G_2+K_1}]\left[\mathbf{I}_{1+G_2+K_1} - \frac{1}{\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}}\begin{pmatrix} \mathbf{\Omega}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}(1, -\boldsymbol{\theta}')\right][\mathbf{G}_1^{(f)} - \sqrt{cc_*}\mathbf{H}_1^{(f)}]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix}$$

$$+ o_p(1) \ .$$

Using the relations of (6.58), we have

$$
\begin{aligned}
& [\mathbf{G}_1^{(f)} - \sqrt{cc_*}\mathbf{H}_1^{(f)}] \begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} \qquad (6.83) \\
= \quad & \frac{1}{\sqrt{n}}\mathbf{D}^{*\prime}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}\mathbf{M}_t\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\left[\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)\prime} \\ \mathbf{O}\end{pmatrix}\mathbf{M}_t\mathbf{u}_t^{(f)} - r_n\begin{pmatrix}\boldsymbol{\Omega\beta} \\ \mathbf{0}\end{pmatrix}\right] \\
& -\frac{\sqrt{cc_*}}{q_n}\mathbf{D}^{*\prime}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}[\mathbf{I}_N - \mathbf{M}_t]\mathbf{u}_t^{(f)} \\
& -\frac{\sqrt{cc_*}}{q_n}\left[\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)\prime} \\ \mathbf{O}\end{pmatrix}[\mathbf{I}_N - \mathbf{M}_t]\mathbf{u}_t^{(f)} - q_n\begin{pmatrix}\boldsymbol{\Omega\beta} \\ \mathbf{0}\end{pmatrix}\right] .
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{\Phi}^*\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{2LI} - \boldsymbol{\beta}_2 \\ \hat{\gamma}_{1LI} - \boldsymbol{\gamma}_1\end{pmatrix} \quad = \quad & \frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)\prime}\mathbf{N}_t\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{U}_t^{(\perp,f)\prime} \\ \mathbf{O}\end{pmatrix}\mathbf{N}_t\mathbf{u}_t^{(f)} \\
& + o_p(1) \ . \qquad (6.84)
\end{aligned}
$$

where

$$
\mathbf{N}_t \;=\; \mathbf{M}_t - c_*(\mathbf{I}_N - \mathbf{M}_t) = \frac{1}{1-c}[\mathbf{M}_t - c\mathbf{I}_N] \ , \qquad (6.85)
$$

$$
\mathbf{U}_t^{(\perp,f)\prime} \;=\; [\mathbf{0}, \mathbf{I}_{G_2}]\left[\mathbf{I}_{1+G_2} - \frac{\boldsymbol{\Omega\beta\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega\beta}}\right]\mathbf{V}_t^{(f)\prime} = (\mathbf{u}_{1t}^{(\perp,f)}, ..., \mathbf{u}_{Nt}^{(\perp,f)}) \ . \qquad (6.86)
$$

[**Step 3**] : We evaluate the effects of the forward-filtering at this step and first consider the case of $\mathbf{M}_t = \mathbf{M}_t^{(a)}$. Set the k-th unit vector $\mathbf{e}_k = (0, ..., 0, 1, 0, ..., 0)'$, then using the relations of (2.5), (A2), and $\mathbf{u}_t^{(f)} = (\mathbf{u}_t - \mathbf{u}_{tT})/c_t$, we decompose the first and second terms of (6.84) as follows, for $k = 1, ..., K (= K_1 + K_2)$ and

$g = 1, ..., G_2$,

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{Z}_{t-1}^{(f)\prime} \mathbf{N}_t^{(a)} \mathbf{u}_t^{(f)} \\
= \;& \frac{1}{1-c_a} \left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{J}' \mathbf{W}_{t-1}' \mathbf{M}_t \mathbf{u}_t - \Upsilon_{11n}^{(k,a)} - \Upsilon_{12n}^{(k,a)} \right) - \left(\Upsilon_{21n}^{(k,a)} - \Upsilon_{22n}^{(k,a)} \right) \right] \\
& - c_{*a} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{J}' \mathbf{W}_{t-1}' \mathbf{u}_t - \Upsilon_{3n}^{(k)} \right) , \\
& \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{(\perp,f)\prime} \mathbf{N}_t^{(a)} \mathbf{u}_t^{(f)} \\
= \;& \frac{1}{1-c_a} \Big[\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{M}_t^{(a)} \mathbf{u}_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \Big(\frac{1}{T-t} \Big) \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{M}_t^{(a)} \mathbf{u}_t \Big) \\
& - c_t^{-2} \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT} - c_t^{-2} \mathbf{e}_g' \bar{\mathbf{U}}_{tT}^{\perp\prime} \mathbf{M}_t^{(a)} \mathbf{u}_t + c_t^{-2} \mathbf{e}_g' \bar{\mathbf{U}}_{tT}^{\perp\prime} \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT} \Big] \\
& - c_{*a} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{u}_t - \sqrt{\frac{T}{N}} \sum_{i=1}^{N} \mathbf{e}_g' \bar{\mathbf{u}}_i^{\perp} \bar{u}_i \right) \\
= \;& \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{N}_t^{(a)} \mathbf{u}_t - \Upsilon_{4n}^{(g,a)} , \qquad (6.87)
\end{aligned}
$$

where

$$\Upsilon_{11n}^{(k,a)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{J}' \mathbf{W}_{t-1}' \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT}, \qquad (6.88)$$

$$\Upsilon_{12n}^{(k,a)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \frac{c_t}{T-t} \mathbf{e}_k' \mathbf{J}' \tilde{\mathbf{W}}_{t-1}' \mathbf{M}_t^{(a)} \mathbf{u}_t^{(f)}, \qquad (6.89)$$

$$\Upsilon_{21n}^{(k,a)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{J}' \tilde{\mathbf{V}}_{tT}' \mathbf{M}_t^{(a)} \mathbf{u}_t, \qquad (6.90)$$

$$\Upsilon_{22n}^{(k,a)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_k' \mathbf{J}' \tilde{\mathbf{V}}_{tT}' \mathbf{M}_t^{(a)} \bar{\mathbf{u}}_{tT}, \qquad (6.91)$$

$$\Upsilon_{3n}^{(k)} = \sqrt{\frac{T}{N}} \sum_{i=1}^{N} \mathbf{e}_k' \mathbf{J}' \bar{\mathbf{w}}_{i(-1)} \bar{u}_i, \qquad (6.92)$$

$$\Upsilon_{4n}^{(g,a)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{(\perp,f)\prime} \mathbf{N}_t^{(a)} \mathbf{u}_t^{(f)} - \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{\perp\prime} \mathbf{N}_t^{(a)} \mathbf{u}_t , \qquad (6.93)$$

and

$$\bar{\mathbf{u}}_{tT} = (\mathbf{u}_t + \cdots + \mathbf{u}_T)/(T-t+1), \tag{6.94}$$

$$\tilde{\mathbf{W}}'_{t-1} = (\sum_{j=1}^{T-t} \mathbf{\Pi}^{*h})\mathbf{W}'_{t-1}\ ,\quad \tilde{\mathbf{V}}'_{tT} = \frac{1}{T-t}\sum_{j=1}^{T-t} \mathbf{\Phi}_h \mathbf{V}^{*'}_{T-h}\ , \tag{6.95}$$

$$\mathbf{V}^{*'}_h = (\mathbf{v}^*_{1h}, ..., \mathbf{v}^*_{Nh}) = (\mathbf{v}^{*(1)}_h, ..., \mathbf{v}^{*(K)}_h)', \tag{6.96}$$

$$\mathbf{\Phi}_h = (\mathbf{I}_{K^*} - \mathbf{\Pi}^*)^{-1}(\mathbf{I}_{K^*} - \mathbf{\Pi}^{*h}), \tag{6.97}$$

$$\bar{\mathbf{w}}_{i(-1)} = \frac{1}{T}\sum_{t=1}^{T-1} \mathbf{w}_{i(t-1)}\ ,\quad \bar{u}_i = \frac{1}{T}\sum_{t=1}^{T-1} u_{it}\ , \tag{6.98}$$

$$\mathbf{U}^{\perp'}_t = [\mathbf{0}, \mathbf{I}_{G_2}]\left[\mathbf{I}_{1+G_2} - \frac{\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}}\right]\mathbf{V}'_t = (\mathbf{u}^{\perp}_{1t}, ..., \mathbf{u}^{\perp}_{Nt})\ , \tag{6.99}$$

$$\bar{\mathbf{U}}^{\perp}_t = (\mathbf{U}^{\perp}_t + \cdots + \mathbf{U}^{\perp}_T)/(T-t+1)\ ,\quad \bar{\mathbf{u}}^{\perp}_i = \frac{1}{T}\sum_{t=1}^{T-1} \mathbf{u}^{\perp}_{it}\ , \tag{6.100}$$

We shall show that the variances of (6.88) to (6.91) go to zeros. (The variances of the terms $\Upsilon^{(g,a)}_{4n}$ and $\Upsilon^{(g,b)}_{4n}$ can be shown by the same argument of Akashi and Kunitomo (2010).) For this purpose, prepare two lemmas. The proof of the first one has been given in Akashi and Kunitomo (2010).

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 $tr(\mathbf{M}_t) = \mathbf{d}'_t\boldsymbol{\iota}_N,\ tr(\mathbf{M}_s) = \mathbf{d}'_s\boldsymbol{\iota}_N,\ \mathbf{d}'_t\mathbf{d}_s \le \max\{tr(\mathbf{M}_t), tr(\mathbf{M}_s)\}$ and $tr(\mathbf{M}_t\mathbf{M}_s) \le \max\{tr(\mathbf{M}_t), tr(\mathbf{M}_s)\}$. Then, for $l \ge r \ge t,\ p \ge q \ge s,\ t \ge s$,

$$Cov[\boldsymbol{\epsilon}^{*'}_l\mathbf{M}_t\boldsymbol{\epsilon}^{**}_r, \boldsymbol{\epsilon}^{*'}_p\mathbf{M}_s\boldsymbol{\epsilon}^{**}_q] \tag{6.101}$$
$$= \begin{cases} (m^{(3)} + m^{(2)})tr(\mathbf{M}_t\mathbf{M}_s) + m^{(0)}\mathcal{E}[\mathbf{d}'_t\mathbf{d}_s] & \text{if } l = r = p = q, \\ \mathcal{E}[\epsilon^{*2}_{it}\epsilon^{**}_{it}]\mathcal{E}[\mathbf{d}'_t\mathbf{M}_s\boldsymbol{\epsilon}^{**}_q] & \text{if } l = r = p \ne q < t, \\ m^{(3)}tr(\mathbf{M}_t\mathbf{M}_s) & \text{if } l = p \ne r = q, \\ 0 & \text{otherwise,} \end{cases}$$

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

$$\begin{aligned} m^{(1)} &= m^{(1)}(\boldsymbol{\epsilon}^*_t, \boldsymbol{\epsilon}^{**}_t) = \mathcal{E}[\boldsymbol{\epsilon}^{*2}_{it}\boldsymbol{\epsilon}^{**2}_{it}], \\ m^{(2)} &= m^{(2)}(\boldsymbol{\epsilon}^*_t, \boldsymbol{\epsilon}^{**}_t) = (\mathcal{E}[\boldsymbol{\epsilon}^*_{it}\boldsymbol{\epsilon}^{**}_{it}])^2, \\ m^{(3)} &= m^{(3)}(\boldsymbol{\epsilon}^*_t, \boldsymbol{\epsilon}^{**}_t) = \mathcal{E}[\boldsymbol{\epsilon}^{*2}_{it}]\mathcal{E}[\boldsymbol{\epsilon}^{**2}_{it}], \\ m^{(0)} &= m^{(0)}(\boldsymbol{\epsilon}^*_t, \boldsymbol{\epsilon}^{**}_t) = m^{(1)} - 2m^{(2)} - m^{(3)}. \end{aligned} \tag{6.102}$$

Lemma 2 : For any k, j, T, we have

$$\sum_{t=1}^{T} |\mathbf{e}_k'\mathbf{\Pi}^{*t}\mathbf{e}_j| \leq O(\max_{l,i} |\mathbf{e}_k'\mathbf{C}\mathbf{e}_l||\mathbf{e}_i'\mathbf{C}^{-1}\mathbf{e}_j|) \times \max_{m',m''} J_{m'm''T} < +\infty, \qquad (6.103)$$

where $\mathbf{C}$ satisfies $\mathbf{\Pi}^* = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^{-1}$, $J_{m'm''T}$ is a bounded positive constant and $\mathbf{\Lambda}$ denotes a Jordan matrix.

We have the second lemma because for any multiplicity $m(= m')$ a corresponding diagonal block's element of $\mathbf{\Lambda}^t$ has the form $\binom{t}{m''}\lambda_{m'}^{t-m''}$ in each position m'' above the main diagonal, $m'' = 0, 1, ..., m-1$, and for any m', m'', $J_{m'm''T} = \sum_{t=m''}^{T} \binom{t}{m''}|\lambda_{m'}|^{t-m''}$ converges to a positive value as $T \to \infty$.

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

$$Var[\Upsilon_{11n}^{(k,a)}] = \frac{1}{NT}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathcal{E}[\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(a)}\mathbf{W}_{t-1}\mathbf{J}\mathbf{e}_k]. \qquad (6.104)$$

For $t \geq s$,

$$\begin{aligned}
\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}] &= \frac{\mathcal{E}[u_{it}^2]}{(T-s+1)}\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(a)}\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}] \qquad (6.105)\\
&= \frac{\sigma^2}{(T-s+1)}\mathcal{E}[\mathcal{E}_s[\mathbf{w}_{t-1}^{(k)'}]\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}]\\
&= \frac{\sigma^2}{(T-s+1)}\mathcal{E}[\sum_{j=1}^{K'}(\mathbf{e}_{kJ}'\mathbf{\Pi}^{*t-s}\mathbf{e}_j)\mathbf{w}_{s-1}^{(j)'}\mathbf{M}_s^{(a)}\mathbf{w}_{s-1}^{(k)}],
\end{aligned}$$

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 also an unit k-th vector. The second equality of (6.105) is due to the fact that $\mathbf{M}_t^{(a)}\mathbf{M}_s^{(a)} = \mathbf{M}_s^{(a)}$. By using the relations 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 $(\mathcal{E}[(\mathbf{w}_0^{(j)'}\mathbf{w}_0^{(j)})(\mathbf{w}_0^{(k)'}\mathbf{w}_0^{(k)})])^{1/2} = O(N)$,

$$\begin{aligned} Var[\Upsilon_{11n}^{(k,a)}] &\leq \frac{\sigma^2 \max_j\{(\mathcal{E}[(\mathbf{w}_0^{(j)'}\mathbf{w}_0^{(j)})(\mathbf{w}_0^{(k)'}\mathbf{w}_0^{(k)})])^{1/2}\}}{N} \\ &\quad \times \frac{1}{T}\left[\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\frac{1}{T-s+1}\sum_{j=1}^{K^*}|\mathbf{e}_{kJ}'\mathbf{\Pi}^{*|t-s|}\mathbf{e}_j|\right] \\ &= \frac{O(1)}{T}[(\frac{1}{T}+\cdots+\frac{1}{2})+2\sum_{j=1}^{K^*}S_T^{(k,j)}] \\ &= \frac{O(\log T)}{T}, \end{aligned} \tag{6.106}$$

where

$$\begin{aligned} S_T^{(k,j)} &= \frac{1}{T}(|\mathbf{e}_{kJ}'\mathbf{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\mathbf{\Pi}^{*T-2}\mathbf{e}_j|) \\ &\quad + \frac{1}{T-1}(|\mathbf{e}_{kJ}'\mathbf{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\mathbf{\Pi}^{*T-3}\mathbf{e}_j|) + \cdots + \frac{1}{3}|\mathbf{e}_{kJ}'\mathbf{\Pi}^*\mathbf{e}_j| \\ &\quad \leq (\frac{1}{3}+\cdots+\frac{1}{T})(|\mathbf{e}_{kJ}'\mathbf{\Pi}^*\mathbf{e}_j| + \cdots + |\mathbf{e}_{kJ}'\mathbf{\Pi}^{*T-2}\mathbf{e}_j|) = O(\log T), \end{aligned}$$

since (6.103). Next,

$$\begin{aligned} Var[\Upsilon_{12n}^{(k,a)}] &= \frac{1}{NT}Var[\sum_{t=1}^{T-1}\frac{c_t}{T-t}\tilde{\mathbf{w}}_{t-1}^{(k)'}\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}] \qquad (6.107) \\ &= \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}[\tilde{\mathbf{w}}_{t-1}^{(k)'}\mathbf{M}_t^{(a)}\tilde{\mathbf{w}}_{t-1}^{(k)}] \\ &\leq \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}[\tilde{\mathbf{w}}_{t-1}^{(k)'}\tilde{\mathbf{w}}_{t-1}^{(k)}] \\ &= \frac{\sigma^2}{NT}\sum_{t=1}^{T-1}\frac{c_t^2 N}{(T-t)^2}\left[\mathbf{e}_k'\mathbf{J}'(\sum_{h=1}^{T-t}\mathbf{\Pi}^{*h})\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}'](\sum_{h=1}^{T-t}\mathbf{\Pi}^{*h})'\mathbf{J}\mathbf{e}_k\right] \\ &\leq \frac{N}{N}\frac{\sigma_2\lambda_{\max}\{\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}']\}}{T}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}\mathbf{e}_k'\mathbf{J}'(\sum_{h=1}^{T-t}\mathbf{\Pi}^{*h})(\sum_{h=1}^{T-t}\mathbf{\Pi}^{*h})'\mathbf{J}\mathbf{e}_k \\ &= \frac{O(1)}{T}. \end{aligned}$$

where $\tilde{\mathbf{w}}_{t-1}^{(k)'} = \mathbf{e}_k'\mathbf{J}'\tilde{\mathbf{W}}_{t-1}'$, and λ_{max} stands for the largest eigenvalue of $\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}']$. The last inequality follows from that $c_t^2 < 1$ and the boundness of $(\sum_{h=1}^{T-t}\mathbf{\Pi}^{*h})$ for any t, T.

Turning to evluate 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 $a_{1n}^{(k,j,a)}$ $(j = 1, ..., K')$

which are defined by

$$
\begin{aligned}
Var[\Upsilon_{21n}^{(k,a)}] &= \frac{1}{NT}Var[\sum_{t=1}^{T-1}\frac{1}{T-t}\sum_{h=1}^{T-t}\sum_{j=1}^{K^*}(\mathbf{e}_{kJ}'\mathbf{\Phi}_h\mathbf{e}_j)\mathbf{e}_j'\mathbf{V}_{T-h}^{*\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t] \qquad (6.108)\\
&= \frac{1}{NT}\Big[\sum_{j=1}^{K^*}Var[\sum_{t=1}^{T-1}\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t]\\
&\quad +\sum_{i,j}^{K^*}Cov[\sum_{t=1}^{T-1}\tilde{\mathbf{v}}_{tT}^{*(k,i)\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t,\sum_{t=1}^{T-1}\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^{(a)}\mathbf{u}_t]\Big]\\
&= \sum_{j=1}^{K^*}(a_{0n}^{(k,j,a)}+a_{1n}^{(k,j,a)})+\frac{1}{NT}\sum_{i,j}^{K^*}Cov[.,.]\ ,
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime} &= \frac{1}{T-t}\sum_{h=1}^{T-t}(\mathbf{e}_{kJ}'\mathbf{\Phi}_h\mathbf{e}_j)\mathbf{e}_j'\mathbf{V}_{T-h}^{*\prime}\ , \qquad (6.109)\\
a_{0n}^{(k,j,a)} &= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{1}{(T-t)^2}\Big[(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t}\mathbf{e}_j)^2Var[\mathbf{u}_t'\mathbf{M}_t^{(a)}\mathbf{v}_t^{*(j)}]+\\
&\quad \cdots+(\mathbf{e}_{kJ}'\mathbf{\Phi}_1\mathbf{e}_j)^2Var[\mathbf{u}_t'\mathbf{M}_t^{(a)}\mathbf{v}_{T-1}^{*(j)}]\Big]\ ,\\
a_{1n}^{(k,j,a)} &= \frac{2}{NT}\sum_{t=1}^{T-2}\Big[\frac{(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t-1}\mathbf{e}_j)^2Cov[\mathbf{u}_t'\mathbf{M}_t^{(a)}\mathbf{v}_{t+1}^{*(j)},\mathbf{u}_{t+1}'\mathbf{M}_{t+1}^{(a)}\mathbf{v}_{t+1}^{*(j)}]}{(T-t)(T-t-1)}+\\
&\quad \cdots+\frac{(\mathbf{e}_{kJ}'\mathbf{\Phi}_1\mathbf{e}_j)^2Cov[\mathbf{u}_t'\mathbf{M}_t^{(a)}\mathbf{v}_{T-1}^{*(j)},\mathbf{u}_{T-1}'\mathbf{M}_{T-1}^{(a)}\mathbf{v}_{T-1}^{*(j)}]}{(T-t)}\Big]\ .
\end{aligned}
$$

By using *Lemma 1* and the boundness of $(\mathbf{e}_{kJ}'\mathbf{\Phi}_h\mathbf{e}_j)^2$, we have

$$
\begin{aligned}
a_{0n}^{(k,j,a)} &\leq \frac{1}{NT}\sum_{t=1}^{T-1}\frac{tr(\mathbf{M}_t^{(a)})}{(T-t)^2}\Big[(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t}\mathbf{e}_j)^2\Big[m^{(3)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})+m^{(2)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})\\
&\quad +|m^{(0)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})|\Big]+\Big[(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t-1}\mathbf{e}_j)^2+\cdots+(\mathbf{e}_{kJ}'\mathbf{\Phi}_1\mathbf{e}_j)^2\Big]m^{(3)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})\Big]\\
&\leq \frac{O(1)}{NT}\sum_{t=1}^{T-1}\frac{t}{(T-t)^2}\Big[\Big[m^{(3)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})+m^{(2)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})\\
&\quad +|m^{(0)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})|\Big]+(T-t-1)m^{(3)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})\Big]\\
&= O(\frac{1}{NT}\sum_t\frac{t}{T-t})=O(\frac{\log T}{N})\ ,
\end{aligned}
$$

where $m^{(l)}(\mathbf{u}_t,\mathbf{v}_t^{*(j)})$ $(l=0,2,3)$ are defined in the same way to (6.102).

Moreover, from 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

$$
\begin{aligned}
|a_{1n}^{(k,j,a)}| &= \frac{2}{NT}\Big|\sum_{t=1}^{T-2}\frac{(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t-1}\mathbf{e}_j)^2 O(\mathcal{E}[\mathbf{d}_{t+1}'\mathbf{M}_t^{(a)}\mathbf{u}_t])}{(T-t)(T-t-1)} \\
&\quad + \cdots + \frac{(\mathbf{e}_{kJ}'\mathbf{\Phi}_1\mathbf{e}_j)^2 O(\mathcal{E}[\mathbf{d}_{T-1}'\mathbf{M}_t^{(a)}\mathbf{u}_t])}{(T-t)}\Big| \\
&\leq \frac{O(1)}{NT}\sum_{t=1}^{T-2}\frac{1}{(T-t)}(\frac{t+1}{T-t-1}+\cdots+\frac{T-1}{1}) \\
&\leq \frac{O(1)}{NT}(\frac{1}{2}+\cdots+\frac{1}{T-1})[T(\frac{1}{2}+\cdots+\frac{1}{T-1})+1] = O(\frac{(\log T)^2}{N}) \ .
\end{aligned}
$$

since $(\log T)^2/N \sim c(\log T)^2/T$. Finally, we consider the variance of $\Upsilon_{22n}^{(k,a)}$,

$$
\begin{aligned}
Var[\Upsilon_{22n}^{(k,a)}] &= \frac{1}{NT}Var[\sum_{t=1}^{T-1}\frac{1}{T-t}\sum_{h=1}^{T-t}\sum_{j=1}^{K^*}(\mathbf{e}_{kJ}'\mathbf{\Phi}_h\mathbf{e}_j)\mathbf{e}_j'\mathbf{V}_{T-h}^{*'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}] \\
&= \frac{1}{NT}\Big[\sum_{j=1}^{K^*}Var[\sum_{t=1}^{T-1}\tilde{\mathbf{v}}_{tT}^{*(k,j)'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}] + \sum_{i,j}^{K^*}Cov[.,.]\Big] \ .
\end{aligned}
$$

By the same arguments as used for the derivations of *Lemma 1*, we have

$$
\begin{aligned}
Var[\tilde{\mathbf{v}}_{tT}^{*(k,j)'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}] &= \Big[m^{(3)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) + m^{(2)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT})\Big]tr(\mathbf{M}_t^{(a)2}) \\
&\quad + m^{(0)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT})\ \mathcal{E}[\mathbf{d}_t^{(a)'}\mathbf{d}_t^{(a)}] \\
&\leq tr(\mathbf{M}_t^{(a)})\Big[m^{(1)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) + m^{(3)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT})\Big] \ ,
\end{aligned}
$$

where $\mathbf{d}_t^{(a)'}\boldsymbol{\iota}_N = tr(\mathbf{M}_t^{(a)})$. The inequality follows from that $E[\mathbf{d}_t^{(a)'}\mathbf{d}_t^{(a)}] \leq tr(\mathbf{M}_t^{(a)})$ and $m^{(1)}(\tilde{\mathbf{v}}_{tT}^{(k,j)},\bar{\mathbf{u}}_{tT}) - m^{(2)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) \geq 0$.

Then,

$$
\begin{aligned}
m^{(3)}(\tilde{\mathbf{v}}_{tT}^{*(k,j)},\bar{\mathbf{u}}_{tT}) &= Var[\frac{1}{T-t}(\mathbf{e}_{kJ}'\mathbf{\Phi}_{T-t}\mathbf{e}_j v_{it}^{*(j)} + \cdots + \mathbf{e}_{kJ}'\mathbf{\Phi}_1\mathbf{e}_j v_{iT-1}^{*(j)})] \\
&\quad \times Var[\frac{1}{T-t+1}(u_{it}+\cdots+u_{iT})] \\
&= O((\frac{1}{T-t})^2) \ ,
\end{aligned}
$$

since $(v_{it}^{*(j)}, u_{it})$ is independent from $(v_{is}^{*(j)}, u_{is})$, if $s \neq t$. Similarly,

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

Therefore, for any j,

$$Var[\tilde{\mathbf{v}}_{tT}^{*(k,j)'}\mathbf{M}_t^{(a)}\bar{\mathbf{u}}_{tT}] = O(\frac{t}{(T-t)^2}) \ . \tag{6.110}$$

From this result and againg by using the arguments as Alvarez and Arellano (2003), we conclude that

$$Var[\Upsilon_{22n}^{(k,a)}] = O(\frac{(\log T)^2}{N}) \ . \tag{6.111}$$

[**Step 4**] : Now we turn to evaluate the limiting distribution of the LIML estimator in the case of using the backward-filtered instruments. We replace $\mathbf{M}_t^{(b)}$ for $\mathbf{M}_t^{(a)}$, then 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 $Var[\Upsilon_{12n}^{(k,.)}]$ can be free with $\mathbf{M}_t$, and those of $\Upsilon_{21n}^{(k,b)}$ and $\Upsilon_{22n}^{(k,b)}$ are reduced by the fact that $tr(\mathbf{M}_t^{(b)}) = O(1)$. For instance, $Var[\Upsilon_{12n}^{(k,b)}] = O(\frac{1}{T})$,

$$Var[\Upsilon_{21n}^{(k,b)}] = O(\frac{(\log T)^2}{N_0 T}), Var[\Upsilon_{22n}^{(k,b)}] = O(\frac{(\log T)^2}{N_0 T}) \ . \tag{6.112}$$

In order to evaluate $Var[\Upsilon_{11n}^{(k,b)}]$, we prepare the next lemma, which is a generalization of the corresponding one by Hayakawa (2006).

Lemma 3 : Define the $N \times 1$ vectors of erros of the population linear projection of $\mathbf{W}_{t-1}\mathbf{J}$ on $\mathbf{Z}_t^{*(b)}\mathbf{J}$,

$$\mathbf{E}_t^{(b)} = [\boldsymbol{\epsilon}_t^{(1,b)}, ..., \boldsymbol{\epsilon}_t^{(K,b)}] = \mathbf{W}_{t-1}\mathbf{J} - \mathbf{Z}_t^{*(b)}\mathbf{J}[\boldsymbol{\gamma}_t^{*(1,b)}, ..., \boldsymbol{\gamma}_t^{*(K,b)}], \tag{6.113}$$

where $\mathbf{Z}_t^{*(b)} = [\mathbf{z}_{1(t-1)}^{*(b)}, ..., \mathbf{z}_{N(t-1)}^{*(b)}]'$, $\mathbf{Z}_t^{*(b)}\mathbf{J} = \mathbf{Z}_t^{(b)}$ and $\boldsymbol{\gamma}_t^{*(k,b)}$ is defined by

$$[\boldsymbol{\gamma}_t^{*(1,b)}, ..., \boldsymbol{\gamma}_t^{*(K,b)}] = (b_t^2 \lim_{t\to\infty} \mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{z}_{it-1}^{*(b)'}]\mathbf{J})^{-1}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{w}_{it-1}^{'}]\mathbf{J}.$$

Then, for $k = 1, ..., K$,

$$\mathcal{E}[\epsilon_{it}^{(k,b)2}] = O(\frac{1}{t}) \ . \tag{6.114}$$

Therefore, for any $\mathbf{M}_t^*$ such that $\mathbf{M}_t^*\mathbf{M}_t^{(b)} = \mathbf{M}_t^{(b)}$,

$$\frac{1}{N_0 T}\sum_{t=1}^{T-1} \mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^*\mathbf{W}_{t-1}\mathbf{J} \xrightarrow{p} \mathbf{J}'\mathcal{E}[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}']\mathbf{J} = \mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J} \ . \tag{6.115}$$

Proof : By using (2.14) and (3.29), we observe

$$\begin{aligned}\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{w}_{it-1}^{'}] &= b_t[\boldsymbol{\Gamma}_0 - \frac{1}{t}\boldsymbol{\Gamma}_0\boldsymbol{\Pi}^{*'}(\mathbf{I}_{K^*} - \boldsymbol{\Pi}^{*'})^{-1}(\mathbf{I}_{K^*} - \boldsymbol{\Pi}^{*'t+1})] \\ &= b_t[\boldsymbol{\Gamma}_0 + O(\frac{1}{t})],\end{aligned} \tag{6.116}$$

and then we can show that

$$\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{z}_{it-1}^{*(b)'}] = b_t^2[\mathbf{\Gamma}_0 + O(\frac{1}{t})] \ . \tag{6.117}$$

Thus we find that $\lim_{t\to\infty} \mathcal{E}[\mathbf{z}_{i(t-1)}^{*(b)}\mathbf{z}_{i(t-1)}^{*(b)'}] = \mathbf{\Gamma}_0$ and then

$$\begin{aligned} & \mathcal{E}[\boldsymbol{\epsilon}_{it}^{(b)}\boldsymbol{\epsilon}_{it}^{(b)'}] \\ = \ & \mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J} - 2\mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{z}_{it-1}^{*(b)'}]\mathbf{J}(b_t^2\mathbf{J}'\mathbf{\Gamma}_0\mathbf{J})^{-1}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{w}_{it-1}']\mathbf{J} \\ & +\mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{z}_{it-1}^{*(b)'}]\mathbf{J}(b_t^2\mathbf{J}'\mathbf{\Gamma}_0\mathbf{J})^{-1}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{z}_{it-1}^{*(b)'}]\mathbf{J}(b_t^2\mathbf{J}'\mathbf{\Gamma}_0\mathbf{J})^{-1}\mathbf{J}'\mathcal{E}[\mathbf{z}_{it-1}^{*(b)}\mathbf{w}_{it-1}']\mathbf{J} \\ = \ & \mathbf{J}'\mathbf{\Gamma}_0\mathbf{J} - \mathbf{J}'\mathbf{\Gamma}_0\mathbf{J} + O(\frac{1}{t}) \ , \end{aligned}$$

where $\boldsymbol{\epsilon}_{it}^{(b)} = (\epsilon_{it}^{(1,b)}, ..., \epsilon_{it}^{(K,b)})'$. By using the fact that $(\mathbf{I}_N - \mathbf{M}_t^*)\mathbf{Z}_t^{*(b)}\mathbf{J} = \mathbf{O}$ and $\mathbf{W}_{t-1}'\mathbf{M}_t^* \ \mathbf{W}_{t-1} = \mathbf{W}_{t-1}'\mathbf{W}_{t-1} - \mathbf{E}_t^{(b)'}(\mathbf{I}_N - \mathbf{M}_t^*)\mathbf{E}_t^{(b)}$, we have

$$\frac{1}{N_0T}\sum_{t=1}^{T-1}\mathcal{E}[\boldsymbol{\epsilon}_t^{(k,b)'}(\mathbf{I}_{N_0} - \mathbf{M}_t^*)\boldsymbol{\epsilon}_t^{(k,b)}] \le \frac{1}{T}\sum_{t=1}^{T-1}\mathcal{E}[\boldsymbol{\epsilon}_{it}^{(k,b)2}] = \frac{O(\log T)}{T} \ . \tag{6.118}$$

Then the convergence in probability of (6.118) is valid by the Markov inequality. For $j \neq k$, we apply the Cauchy-Schwarz inequality and we have that $(1/N_0T)\sum_t \mathbf{W}_{t-1}'\mathbf{W}_{t-1} \xrightarrow{p} \mathbf{\Gamma}_0$ as $T \to \infty$. **Q.E.D.**

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

$$Var[\Upsilon_{11n}^{(k,b)}] = \frac{1}{N_0T}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathcal{E}[\mathbf{e}_k'\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(b)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(b)}\mathbf{W}_{t-1}\mathbf{J}\mathbf{e}_k]. \tag{6.119}$$

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

$$\begin{aligned} & \mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(b)}\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}'\mathbf{M}_s^{(b)}\mathbf{w}_{s-1}^{(k)}] \\ = \ & \frac{\mathcal{E}[u_{it}^2]}{(T-s+1)}\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{M}_t^{(b)}\mathbf{M}_s^{(b)}\mathbf{w}_{s-1}^{(k)}] \\ = \ & \frac{\sigma^2}{(T-s+1)}\Big[\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}(\mathbf{I}_{N_0} - \mathbf{M}_s^{(b)})\boldsymbol{\epsilon}_{s-1}^{(k,b)}] - \mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{w}_{s-1}^{(k)}] \\ & -\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}(\mathbf{I}_{N_0} - \mathbf{M}_t^{(b)})(\mathbf{I}_{N_0} - \mathbf{M}_s^{(b)})\boldsymbol{\epsilon}_{s-1}^{(k,b)}] + \mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}(\mathbf{I}_{N_0} - \mathbf{M}_t^{(b)})\mathbf{w}_{s-1}^{(k)}]\Big] \ , \end{aligned} \tag{6.120}$$

where we have used 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 have $\mathcal{E}[\mathbf{w}_{t-1}^{(k)'}\mathbf{w}_{s-1}^{(k)}] = \mathcal{E}[\mathcal{E}_s[\mathbf{w}_{t-1}^{(k)'}]\mathbf{w}_{s-1}^{(k)}]$. Thus the corresponding order is equal to $O(Var[\Upsilon_{11n}^{(k,a)}]) = O(\log T/T)$. Hence for the first term we have the same result. As for the third term

$$
\begin{aligned}
& |\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)})(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)})\boldsymbol{\epsilon}_{s-1}^{(k,b)}]| \leq (\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}\boldsymbol{\epsilon}_{t-1}^{(k,b)}\boldsymbol{\epsilon}_{s-1}^{(k,b)'}\boldsymbol{\epsilon}_{s-1}^{(k,b)}])^{1/2} \\
= & \Big(\sum_{i=1}^{N_0}\mathcal{E}[\epsilon_{i(t-1)}^{(k,b)2}\epsilon_{i(s-1)}^{(k,b)2}] + \sum_{i,j,i\neq j}^{N_0}\mathcal{E}[\epsilon_{i(t-1)}^{(k,b)2}][\epsilon_{j(s-1)}^{(k,b)2}]\Big)^{1/2} \\
\leq & [O(\frac{N_0}{ts})+O(\frac{N_0(N_0-1)}{ts})]^{1/2} \\
= & O(N_0\frac{1}{\sqrt{t}}\frac{1}{\sqrt{s}}),
\end{aligned}
$$

where the first equality is due to independence of random variables $\epsilon_{i(t-1)}^{(k,b)2}$. Then at the second inequality we have applied *Lemma 3* and the Cauchy-Schwarz inequality as

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

Thus

$$
\begin{aligned}
& \frac{1}{N_0T}\sum_{s=1}^{T-1}2\sum_{t\geq s}^{T-1}\frac{\sigma^2}{T-s+1}|\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k,b)'}(\mathbf{I}_{N_0}-\mathbf{M}_t^{(b)})(\mathbf{I}_{N_0}-\mathbf{M}_s^{(b)})\boldsymbol{\epsilon}_{s-1}^{(k,b)}]| \\
\leq & \frac{N_0O(1)}{N_0T}\sum_{s=1}^{T-1}2\sum_{t\geq s}^{T-1}\frac{1}{T-s+1}\frac{1}{\sqrt{t}}\frac{1}{\sqrt{s}} \\
\leq & \frac{O(1)}{T}\sum_{s=1}^{T-1}\frac{1}{T-s+1}\sum_{t=1}^{T-1}\frac{1}{\sqrt{t}} \\
= & O(\frac{(\log T)\sqrt{T}}{T}) .
\end{aligned}
$$

For the fourth term of last equality of (6.120), we have the same order by the similar arguments. Hence, we find that

$$Var[\Upsilon_{11n}^{(k,b)}] = O(\frac{\log T}{\sqrt{T}}) . \tag{6.121}$$

[**Step 5**] : We shall drive the relevant asymptotic covariance and bias at this step. First, we prepare the next lemma, which is useful for deriving an explicit asymptotic covariance formula for the case (a).

Lemma 4 : Let $(\mu_i^{(1)},...,\mu_i^{(k)},...,\mu_i^{(K^*)})' = \boldsymbol{\mu}_i = [\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*]^{-1}\boldsymbol{\pi}_i^*$ and $\mathbf{M}_\mu = [\boldsymbol{\mu}_1,...,\boldsymbol{\mu}_N]'$. Define the $N\times 1$ vectors of errors of the population linear projection of $\mathbf{M}_\mu\mathbf{J}$ on $\mathbf{Z}_t^{(a)}$,

$$\mathbf{E}_t^{(a)} = [\boldsymbol{\epsilon}_t^{(1,a)},...,\boldsymbol{\epsilon}_t^{(K,a)}] = \mathbf{M}_\mu\mathbf{J}-\mathbf{Z}_t^{(a)}[\boldsymbol{\gamma}_t^{*(1,a)},...,\boldsymbol{\gamma}_t^{*(K,a)}], \tag{6.122}$$

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

$$\mathcal{E}[\epsilon_{it}^{(k,a)2}] = O(\frac{1}{t}) \ . \tag{6.123}$$

Therefore,

$$\frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(a)}\mathbf{W}_{t-1}\mathbf{J} \xrightarrow{p} \mathbf{J}'\mathcal{E}[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}']\mathbf{J} = \mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J} \ . \tag{6.124}$$

Proof : For $k=1,...,K$, by using the fact $\mathbf{z}_{i(h-1)}^{*}=\mathbf{w}_{i(h-1)}+\boldsymbol{\mu}_i$,

$$\begin{aligned} \epsilon_{it}^{(k,a)} &= \mu_i^{(k)} - \sum_{h=1}^{t}\mathbf{z}_{i(h-1)}^{(a)'}\boldsymbol{\gamma}_{th}^{*(k,a)} \qquad (6.125)\\ &= (\mu_i^{(k)} - \frac{1}{t}\sum_{h=1}^{t}\mu_i^{(k)}) - \frac{1}{t}\sum_{h=1}^{t} w_{i(h-1)}^{[k]} \ , \end{aligned}$$

since by the construction of the K_*-variables, there is one variable in each $\mathbf{z}_{i(h-1)}^{(a)}$ ($h=1,...,t$) such that $l=k$. (For convenience we use the notation that $w_{i(h-1)}^{[k]}$ is the $k-$th element of $\mathbf{w}_{i(h-1)}$.) Therefore, $\mathcal{E}[\epsilon_{it}^{(k,a)2}] = Var[(1/t)\sum_{h=1}^{t} w_{i(h-1)}^{[k]}] = O(1/t)$.

Moreover,

$$\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{M}_t^{(a)}\mathbf{W}_{t-1}\mathbf{J} = \mathbf{J}'\mathbf{W}_{t-1}'\mathbf{W}_{t-1}\mathbf{J} - \mathbf{E}_t^{(a)'}(\mathbf{I}_N-\mathbf{M}_t^{(a)})\mathbf{E}_t^{(a)} \ , \tag{6.126}$$

the equality follows from the facts that $\mathbf{W}_{t-1}\mathbf{J} = \mathbf{Z}_{t-1}^{*}\mathbf{J} - \mathbf{Z}_t^{(a)}[\boldsymbol{\gamma}_t^{*(1,a)},...,\boldsymbol{\gamma}_t^{*(K,a)}] - \mathbf{E}_t^{(a)}$, and $(\mathbf{I}_N-\mathbf{M}_t^{(a)})(\mathbf{Z}_{t-1}^{*}\mathbf{J}-\mathbf{Z}_t^{(a)}[\boldsymbol{\gamma}_t^{*(1,a)},...,\boldsymbol{\gamma}_t^{*(K,a)}]) = \mathbf{O}_{N\times K}$ since $\mathbf{M}_t^{(a)}\mathbf{Z}_{t-1}^{*}\mathbf{J} = \mathbf{Z}_{t-1}^{*}\mathbf{J}$ and we define an $N\times K^*$ matrix $\mathbf{Z}_{t-1}^{*}=(\mathbf{z}_{t-1}^{*'})$. The rest of the proof is established by the same arguments used for *Lemma 3*. **Q.E.D.**

Then we can re-write (6.84) as

$$\begin{aligned} &\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{N}_t\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{U}_t^{(\perp,f)'}\\ \mathbf{O}\end{pmatrix}\mathbf{N}_t\mathbf{u}_t^{(f)}\\ =\ &\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{N}_t\mathbf{u}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{U}_t^{\perp'}\\ \mathbf{O}\end{pmatrix}\mathbf{N}_t\mathbf{u}_t + O(1)+o_p(1)\\ =\ &\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{u}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{U}_t^{\perp'}\\ \mathbf{O}\end{pmatrix}\mathbf{N}_t\mathbf{u}_t + O(1)+o_p(1)\\ =\ &\mathbf{A}_{1n}+\mathbf{A}_{2n}+O(1)+o_p(1)\ ,(say,) \end{aligned} \tag{6.127}$$

where the first equality is due to the result of Step 2. The second equality follows from that $\mathbf{N}_t = \mathbf{I}_N - (1+c_*)(\mathbf{I}_N - \mathbf{M}_t)$ and

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

where $\epsilon_{it}^{(k,.)} = \epsilon_{it}^{(k,a)}$ or $\epsilon_{it}^{(k,a)}$, and the last equality is due to *Lemma 3* and *Lemma 4*.

Then we can evaluate the asymptotic variance-covariance terms of the LIML estimator. We immedeately have

$$
\mathcal{E}[\mathbf{A}_{1n}\mathbf{A}_{1n}'] = \frac{\mathcal{E}_t[u_{it}^2]}{NT}\mathbf{D}'\mathcal{E}\Big[\sum_{t=2}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{W}_{t-1}\mathbf{J}\Big]\mathbf{D} \longrightarrow \sigma^2\boldsymbol{\Phi}^* , \tag{6.128}
$$

By using the *i-th* unit vector $\mathbf{e}_i$ $(i = 1, ..., N)$,

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

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

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

Then the first term converges

$$
\frac{1}{NT}\sum_{t=2}^{T-1}tr(\mathbf{N}_t^2)\sigma^2\mathcal{E}\left[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}\right] \longrightarrow c_*\sigma^2\mathcal{E}\left[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}\right] , \tag{6.129}
$$

because we have $\mathbf{N}_t^2 = \mathbf{M_t} + c_*^2(\mathbf{I}_N - \mathbf{M_t})$ and

$$
\frac{1}{n}\sum_{t=2}^{T-1}tr(\mathbf{M}_t) + c_*^2\frac{1}{n}\sum_{t=2}^{T-1}tr(\mathbf{I}_N - \mathbf{M}_t) = \frac{r_n}{n} + \frac{q_n}{n}c_*^2 \longrightarrow c_* . \tag{6.130}
$$

For any vector $\mathbf{t}$, the second term becomes

$$\mathbf{t}^{'}\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{U}_t^{\perp'}\mathbf{N}_t(\mathbf{u}_t\mathbf{u}_t^{'}-\sigma^2\mathbf{I}_N)\mathbf{N}_t\mathbf{U}_t^{\perp}\right]\mathbf{t} \tag{6.131}$$
$$= \frac{1}{NT}\sum_{t=2}^{T-1}\sum_{j=1}^{N}\mathcal{E}\left[(\mathbf{e}_j^{'}\mathbf{N}_t\mathbf{e}_j)^2\mathcal{E}_t\left[(u_{it}^2-\sigma^2)(\mathbf{u}_{it}^{\perp'}\mathbf{t})^2\right]\right]\longrightarrow \mathbf{t}^{'}\boldsymbol{\Xi}_4\mathbf{t}$$

by using the similar calculations as $\mathcal{E}[\mathbf{A}_{1n}\mathbf{A}_{2n}^{'}]$.

Next, we shall evaluate the asymptotic bias of LIML estimator, and first notice that $\mathcal{E}[\Upsilon_{4n}^{(g,a)}]=\mathcal{E}[\Upsilon_{4n}^{(g,b)}]=0$ in (6.93) by using the fact that for any i,j,s,t, $\mathcal{E}_t[\mathbf{u}_{it}^{\perp}u_{js}]=0$. So that in the case of $\mathbf{M}_t=\mathbf{M}_t^{(a)}$, we can evaluate the asymptotic bias as follow

$$\mathbf{b}^{(a)}=\boldsymbol{\Phi}^{*-1}\mathbf{D}'\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\mathcal{E}\Big[\mathbf{Z}_{t-1}^{(f)'}(\frac{1}{1-c_a})(\mathbf{M}_t^{(a)}-c_a\mathbf{I}_N)\mathbf{u}_t^{(f)}\Big]. \tag{6.132}$$

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

$$\mathcal{E}\Big[\mathbf{W}_{i(-1)}^{'}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{'}\mathbf{u}_i\Big] \tag{6.133}$$
$$=\mathcal{E}\Big[(\mathbf{w}_{i0},\boldsymbol{\Pi}^*\mathbf{w}_{i0}+\mathbf{v}_{i1}^*\ ,...,\ \boldsymbol{\Pi}^{*T-1}\mathbf{w}_{i0}+\cdots+\mathbf{v}_{i(T-1)}^*)\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{'}\mathbf{u}_i\Big]$$
$$=\sum_{h=1}^{T-1}\sum_{j=0}^{T-1-h}\boldsymbol{\Pi}^{*j}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]$$
$$=T(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]$$
$$-(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}[\mathbf{I}_{K^*}+\boldsymbol{\Pi}^*(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-1})]\mathcal{E}[\mathbf{v}_{it}^*u_{it}]\ .$$

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

$$\mathcal{E}\Big[c_t\tilde{\mathbf{V}}_{tT}^{'}\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}\Big] \tag{6.134}$$
$$=\frac{K_*t}{T-t+1}\Big[\boldsymbol{\Phi}_{T-t}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]-\frac{1}{T-t}(\boldsymbol{\Phi}_{T-t-1}+\cdots+\boldsymbol{\Phi}_1)\mathcal{E}[\mathbf{v}_{it}^*u_{it}]\Big]$$
$$=\frac{K_*t}{T-t+1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}\Big[(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t})-\mathbf{I}_{K^*}$$
$$+(\frac{1}{T-t})[\mathbf{I}_{K^*}+\boldsymbol{\Pi}^*(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t-1})]\Big]\mathcal{E}[\mathbf{v}_{it}^*u_{it}]$$
$$=\frac{K_*t}{T-t+1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}\Big[\boldsymbol{\Pi}^{*T-t}+(\frac{1}{T-t})(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t})\Big]\mathcal{E}[\mathbf{v}_{it}^*u_{it}]$$
$$=\frac{K_*t}{T-t+1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-2}\Big[-(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)\boldsymbol{\Pi}^{*T-t}+(\frac{1}{T-t})(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t})\Big]\mathcal{E}[\mathbf{v}_{it}^*u_{it}]$$
$$=tK^*(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-2}$$
$$\times\Big[\frac{1}{T-t}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t})-\frac{1}{T-t+1}(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*T-t+1})\Big]\mathcal{E}[\mathbf{v}_{it}^*u_{it}],$$

thus

$$\sum_{t=1}^{T-1}\mathcal{E}\Big[\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^{(a)}\mathbf{u}_t^{(f)}\Big] \tag{6.135}$$

$$= -K_*(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)^{-2}\Big[(T-1)(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)+O(\log T)\Big]\mathcal{E}[\mathbf{v}_{it}^* u_{it}].$$

Therefore, we obtain

$$\begin{aligned}\mathbf{b}^{(a)} &= -\lim_{N,T\to\infty}\Big(\frac{K_*}{1-c_a}\frac{T}{\sqrt{NT}}-\frac{c_a}{1-c_a}\frac{N}{\sqrt{NT}}\Big)\mathbf{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{K'}-\mathbf{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]\\ &= -\frac{K_*\sqrt{\lim(T/N)}}{2-K^*\lim(T/N)}\mathbf{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]. \end{aligned}\tag{6.136}$$

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

$$\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^{(b)}\mathbf{u}_t^{(f)}] \tag{6.137}$$

$$= -K\mathbf{J}'(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)^{-2}\Big[(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)-(1/T)(\mathbf{I}_{K^*}-\mathbf{\Pi}^{*T})\Big]\mathcal{E}[\mathbf{v}_{it}^*u_{it}].$$

Hence, regardless of whether $N_0\to\infty$ or fixed, we have non bias

$$\begin{aligned}\mathbf{b}_0^{(b)} &= -\lim_{T\to\infty}\Big(\frac{K}{1-c_b}\frac{1}{\sqrt{N_0T}}-\frac{c_b}{1-c_b}\frac{N_0}{\sqrt{N_0T}}\Big)\mathbf{\Phi}^{*-1}\mathbf{D}'\mathbf{J}'(\mathbf{I}_{K^*}-\mathbf{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{it}^*u_{it}]\\ &= \mathbf{0}\,. \end{aligned}\tag{6.138}$$

[**Step 6**] : We now turn to consider the asymptotic covariance matrix and the bias of the GMM estimator in some case. If $c=0$, the normalized GMM estimator are asymptotically equivalent to

$$\begin{aligned}&\mathbf{G}_0\sqrt{n}(\hat{\boldsymbol{\theta}}_{GM}-\boldsymbol{\theta})\\ =\ &\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'\mathbf{u}_t+\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{J}_{G_2}'\mathbf{V}_t'\\ \mathbf{O}\end{pmatrix}\mathbf{M}_t\mathbf{u}_t+o_p(1)\,,\end{aligned}$$

where $\mathbf{J}_{G_2}'=[\mathbf{0},\mathbf{I}_{G_2}]$. For any $G_2\times 1$ vector $\mathbf{t}$ and $\mathbf{t}_G=\mathbf{J}_G\mathbf{t}$, by using *lemma 1*

$$\begin{aligned}Var\Big[\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{t}_G'\mathbf{V}_t'\mathbf{M}_t\mathbf{u}_t\Big] &= \frac{1}{NT}\sum_{t=1}^{T-1}Var[\mathbf{t}_G'\mathbf{V}_t'\mathbf{M}_t\mathbf{u}_t]\\ &= O(c)\,. \end{aligned}\tag{6.139}$$

Thus in both cases $\mathbf{M}_t=\mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$, the asymptotic variance-covariance matrix becomes $\mathbf{G}_0^{-1}(\sigma^2\mathbf{\Phi}^*)\mathbf{G}_0^{-1}=\sigma^2\mathbf{\Phi}^{*-1}$. Also under the condition $\sum_t tr(\mathbf{M}_t)$

$/(\sqrt{NT}) < \infty$, the asymptotic bias is given by

$$\mathbf{b}^{(\cdot)} = \lim_{N,T\to\infty} \Big[\frac{1}{\sqrt{NT}} \sum_{t=1}^{T-1} tr(\mathbf{M}_t)\Big] \mathbf{\Phi}^{*-1} \begin{pmatrix} \mathbf{J}_{G_2}' \mathcal{E}[\mathbf{v}_{it} u_{it}] \\ \mathbf{0} \end{pmatrix} . \tag{6.140}$$

[**Step 7**] : Finally, we consider the asymptotic normality of the LIML estimator. (The asymptotic normality of the GMM estimator in some case can be proven in the same way.) Define the $(G_2 + K_1) \times 1$ martingale difference sequence by

$$\mathbf{A}_{tN} = \mathbf{A}_{1tN} + \mathbf{A}_{2tN} = \frac{1}{\sqrt{N}} \Big[\mathbf{D}'\mathbf{J}' \sum_{i=1}^{N} \mathbf{w}_{i(t-1)} \mathbf{u}_t + \begin{pmatrix} \mathbf{U}_t^{\perp'} \\ \mathbf{O} \end{pmatrix} \mathbf{N}_t \mathbf{u}_t \Big] , \tag{6.141}$$

then $\mathbf{A}_{1n} + \mathbf{A}_{2n} = (1/\sqrt{T}) \sum_t (\mathbf{A}_{1tN} + \mathbf{A}_{2tN})$. Then $(1/n) \sum_t \mathbf{W}_{t-1}' (\mathbf{W}_{t-1}, \iota_N) \xrightarrow{p} (\mathbf{\Gamma}_0, \mathbf{0})$, the independence of u_{it} from $\mathcal{F}_{t-1}$, and the same arguments as used for the asymptotic covariance evaluation, for any vector $\mathbf{t}$ and any N,

$$\frac{1}{T} \sum_{t=1}^{T-1} \mathcal{E}\Big[\mathbf{t}' \mathbf{A}_{tN} \mathbf{A}_{tN}' \mathbf{t} | \mathcal{F}_{t-1}\Big] \xrightarrow{p} \frac{1}{T} \sum_{t=1}^{T-1} \mathcal{E}\Big[\mathbf{t}' \mathbf{A}_{tN} \mathbf{A}_{tN}' \mathbf{t}\Big] , \tag{6.142}$$

as $T \to \infty$. Moreover, for some constant Δ' and any t, N,

$$\mathcal{E}\Big[[\mathbf{t}'(\mathbf{A}_{1tN} + \mathbf{A}_{2tN})]^4\Big] < \Delta' . \tag{6.143}$$

This is so because $\mathcal{E}\Big[[\mathbf{t}'\mathbf{A}_{1tN}]^4\Big] < \infty$ and $\mathcal{E}\Big[[(c_*/\sqrt{N})\mathbf{t}'(\mathbf{U}^{\perp'}(\mathbf{I}_N - \mathbf{M}_t)\mathbf{u}_t)]^4\Big] < \infty$ by the similar arguments as used for the following *Lemma 5*. Thus the Lyapounov conditions hold for both cases $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ and $\mathbf{M}_t^{(b)}$.

Lemma 5 : For any $G_2 \times 1$ vector $\mathbf{t}$ and any t, N, there is a positive constant Δ such that

$$\mathcal{E}\Big[[(\frac{1}{\sqrt{N}})\mathbf{t}'\mathbf{U}_t^{\perp'} \mathbf{M}_t \mathbf{u}_t]^4\Big] < \Delta . \tag{6.144}$$

Proof : Define $t_i^{(t)} = \mathbf{t}'\mathbf{U}_t^{\perp'}\mathbf{e}_i$, $m_{ij}^{(t)} = \mathbf{e}_i'\mathbf{M}_t\mathbf{e}_j$ and re-write $u_j^{(t)} = u_{jt}$, then

$$\begin{aligned} & \mathcal{E}_t\Big[[\mathbf{t}'\mathbf{U}_t^{\perp'}\mathbf{M}_t\mathbf{u}_t]^4\Big] \\ = & \sum_{i,i',i'',i'''}^{N} \sum_{j,j',j'',j''}^{N} m_{ij}^{(t)} m_{i'j'}^{(t)} m_{i''j''}^{(t)} m_{i'''j'''}^{(t)} \mathcal{E}_t[t_i^{(t)} t_{i'}^{(t)} t_{i''}^{(t)} t_{i'''}^{(t)} u_j^{(t)} u_{j'}^{(t)} u_{j''}^{(t)} u_{j'''}^{(t)}] \\ = & \mathcal{E}[t_i t_{i'} t_{i''} t_{i'''} u_j u_{j'} u_{j''} u_{j'''}] \sum_{I_h} m_{ij}^{(t)} m_{i'j'}^{(t)} m_{i''j''}^{(t)} m_{i'''j'''}^{(t)} + O(N^2) \end{aligned} \tag{6.145}$$

where the second equality follows from that the homogeneity of $\mathbf{v}_{it}$ over i, t, and the fact that $\mathcal{E}[t_i u_j] = 0$, $|m_{ij}| \leq 1$ for any i, j, t. Hence we shall check that the summation over the following index set I_h becomes also $O(N^2)$. In order to define I_h, put the terms which has more than three products of the moments

$$\begin{aligned}
\alpha_1 &= \mathcal{E}[t_i t_{i'} u_j u_{j'}]\mathcal{E}[t_{i''} t_{i'''}]\mathcal{E}[u_{j''} u_{j'''}] \ , \ \alpha_2 = \mathcal{E}[t_i t_{i'} t_{i''} t_{i'''}]\mathcal{E}[u_j u_{j'}]\mathcal{E}[u_{j''} u_{j'''}] \ , \\
\alpha_3 &= \mathcal{E}[t_i t_{i'}]\mathcal{E}[t_{i''} t_{i'''}]\mathcal{E}[u_j u_{j'} u_{j''} u_{j'''}] \ , \ \alpha_4 = \mathcal{E}[t_i t_{i'}]\mathcal{E}[t_{i''} t_{i'''}]\mathcal{E}[u_j u_{j'}]\mathcal{E}[u_{j''} u_{j'''}] \ , \\
\alpha_5 &= \mathcal{E}[t_i u_j u_{j'}]\mathcal{E}[t_{i'} t_{i''} t_{i'''}]\mathcal{E}[u_{j''} u_{j'''}] \ , \ \alpha_6 = \mathcal{E}[t_i t_{i'} u_{j'''}]\mathcal{E}[t_{i''} t_{i'''}]\mathcal{E}[u_j u_{j'} u_{j''}] \ , \\
\alpha_7 &= \mathcal{E}[t_i t_{i'} u_j]\mathcal{E}[t_{i''} t_{i'''} u_{j'}]\mathcal{E}[u_{j''} u_{j'''}] \ , \ \alpha_8 = \mathcal{E}[t_i t_{i'}]\mathcal{E}[t_{i''} u_j u_{j'}]\mathcal{E}[t_{i'''} u_{j''} u_{j'''}] \ .
\end{aligned}$$

Thus define the set $I_h = \{\{i, i', i'', i''', j, j', j'', j''\} | \ \mathcal{E}[t_i t_{i'} t_{i''} t_{i'''} u_j u_{j'} u_{j''} u_{j'''}] = \alpha_h\}$. From the fact that $\mathbf{M}_t^2 = \mathbf{M}_t$, we have $m_{ij}^{(t)} = m_{ji}^{(t)}$, and

$$\sum_{j}^{N} m_{ij}^{(t)} m_{ji'}^{(t)} = m_{ii'}^{(t)} \ , \ \sum_{i,j}^{N} m_{i,j}^{(t)} \leq N \ , \tag{6.146}$$

where the inequality can be shown by using the similar arguments as used for Lemma 3 in Anderson et al. (2008b). In effect, using these properties we can obtain three types order $O([tr(\mathbf{M}_t)]^2)$, $O(tr(\mathbf{M}_t)N)$, and $O(N^2)$ for $\sum_{I_h}$. Then the total number of the patterns which belong to some order type is finite, hence we may conclude $\sum_{I_h} = O(N^2)$.

This is so because for $h = 1, ..., 4$, we have the conditions $i = i'$ and $i'' = i'''$ regardless $i = i''$ or $i \neq i''$, and also $j = j'$, $j'' = j'''$. Then $\sum_{I_h}$ is reduced to the double summation

$$\sum_{i=i', i''=i''', j=j', j''=j'''}^{N} m_{ij}^{(t)} m_{i'j'}^{(t)} m_{i''j''}^{(t)} m_{i'''j'''}^{(t)} = \sum_{i=i', i''=i'''}^{N} m_{ii'}^{(t)} m_{i''i'''}^{(t)} = [tr(\mathbf{M}_t)]^2 \tag{6.147}$$

For $h = 5, ..., 8$, we can have the type of conditions that $i = i'$, $i'' = i'''$ or $j = j'$, $j'' = j'''$, and at least $j = j'$ or $i = i'$, thus the summation is reduced to the triple summation. By using (6.146),

$$\sum_{i=i', i''=i''', j=j', j'', j'''}^{N} m_{ij}^{(t)} m_{i'j'}^{(t)} m_{i''j''}^{(t)} m_{i'''j'''}^{(t)} = \sum_{j=j', j'', j'''}^{N} m_{jj'}^{(t)} m_{j''j'''}^{(t)} = N tr(\mathbf{M}_t) . \tag{6.148}$$

For any N, t, it holds that $tr(\mathbf{M}_t)/N < 1$ and then the existence of 8-th order moment ensures (6.144). **Q.E.D.**

Proof of Theorem 3.3 :

The method of proof of Theorem 3.3 is essentially the same as the one used by Anderson et al. (2008b). Hence it should be short and we treat the case when

$K_1 = 0$ and $\boldsymbol{\gamma}_1 = \mathbf{0}$ for the simplicity. We set the vector of true parameters $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2') = (1, \beta_2, \cdots, \beta_{1+G_2})$. Then an estimator of the vector $\boldsymbol{\beta}_2$ is composed of

$$\hat{\beta}_i = \phi_i(\frac{1}{n}\mathbf{G}^{(f)}, \frac{1}{q_n}\mathbf{H}^{(f)}) \ \ (i = 2, \cdots, 1 + G_2) \ . \tag{6.149}$$

For the estimator to be consistent, we need the conditions

$$\beta_i = \phi_i \left[\left(\begin{array}{c} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{array} \right) \boldsymbol{\Phi}^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2} \right) + c\, \boldsymbol{\Omega},\ \boldsymbol{\Omega} \right] \ \ (i = 2, \cdots, 1 + G_2) \tag{6.150}$$

as identities with respect to parameters $\boldsymbol{\beta}_2$, $\boldsymbol{\Phi}^*$, and $\boldsymbol{\Omega}$. Then the proof of Theorem 4 of Anderson et al. (2008b) implies the next result.

Lemma 6 : Let $\hat{\boldsymbol{\beta}}_2$ be a consistent estimator in the class of (3.42). For $\phi = (\phi_k)$, let also

$$\boldsymbol{\tau}_{11} = \left[\begin{array}{c} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{array} \right] \ , \tag{6.151}$$

where $\tau_{11}^{(k)} = \frac{\partial \phi_k}{\partial g_{11}}$ $(k = 2, \cdots, 1 + G_2)$. Then

$$\begin{aligned} \hat{\mathbf{e}} &= \sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \\ &= \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}' + (\mathbf{0}, \boldsymbol{\Phi}^{*-1}) \right] \mathbf{S}^{(f)}\boldsymbol{\beta} + o_p(1) \end{aligned} \tag{6.152}$$

where $\mathbf{S}^{(f)} = \mathbf{G}_1^{(f)} - \sqrt{cc_*}\mathbf{H}_1^{(f)}$.

When $c_a = 0$ or $c_b = 0$, for instance, the asymptotic variance-covariance matrix of $\mathbf{S}^{(f)}\boldsymbol{\beta}$ has been obtained by *Theorem 3.1* and *Theorem 3.2* as the corresponding cases. Then

$$\begin{aligned} & \mathcal{E}\left[\hat{\mathbf{e}}\, \hat{\mathbf{e}}' \right] \\ = & \left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}' + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}) \right] \\ & \times \mathcal{E}[\mathbf{S}^{(f)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{S}^{(f)}] \times \left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}' + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}) \right]' \\ = & \boldsymbol{\Psi}^* + \mathcal{E}\left[(\boldsymbol{\beta}'\mathbf{S}^{(f)}\boldsymbol{\beta})^2 \right] \left[\sigma^2\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta} \right] \left[\sigma^2\boldsymbol{\tau}_{11}' + \boldsymbol{\beta}'\boldsymbol{\Omega} \left(\begin{array}{c} \mathbf{0}' \\ \boldsymbol{\Phi}^{*-1} \end{array} \right) \right] + o(1) \ , \end{aligned}$$

where $\mathbf{\Psi}^*$ has been given by *Theorem 3.1* and *Theorem 3.2*.
This covariance matrix is the sum of a positive semi-definite matrix of rank 1 and a positive definite matrix. It has a minimum if

$$\boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \mathbf{\Phi}^{*-1})\mathbf{\Omega}\boldsymbol{\beta} \ . \tag{6.153}$$

Hence we have completed the proof of *Theorem 3.3* for the case of $c_a = 0$ or $c_b = 0$. Other cases in *Theorem 3.1* and *Theorem 3.2* can be treated in the same way.
Q.E.D.

References

[1] Akashi, K. (2008), "*t*-Test in Dynamic Panel Structural Equations," *Unpublished Manuscript*.

[2] Akashi, K. and N. Kunitomo (2010), "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," Discussion Paper CIRJE-F-707, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2010).

[3] Alvarez, J. and M. Arellano (2003), "The Time Series and Cross Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, Vol. 71, 1121-1159.

[4] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, John-Wiley, 3rd Edition.

[5] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[6] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of

Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[7] Anderson, T.W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, Vol. 76, 598-606.

[8] Anderson, T.W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models with Panel Data," *Journal of Econometrics*, Vol. 18, 47-82.

[9] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations with Possibly Many Instruments," Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo.

[10] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2008a), "On Finite Sample Properties of Alternative Estimators of Coefficients in a Structural Equation with Many Instruments," Discussion Paper CIRJE-F-576, Graduate School of Economics, University of Tokyo, forthcoming in *the Journal of Econometrics*.

[11] Anderson, T.W., N. Kunitomo and Y. Matsushita (2008b), "On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments," Discussion Paper CIRJE-F-542, Graduate School of Economics, University of Tokyo, forthcoming in *the Journal of Econometrics*.

[12] Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press.

[13] Baltagi, Badi H. (2005), *Econometric Analysis of Panel Data*, 3rd Edition, John-Wiley.

[14] Blundell, R. and S. Bond (2000), "GMM Estimation with Persistent Panel Data : An Application to Production Function," *Econometric Review*, Vol. 19-3, 321-340.

[15] Hahn, J and G. Kuersteiner (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica*, Vol. 70, 1639-1657.

[16] Hayakawa, K. (2006), "Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present," Hi-Stat Discussion Paper No.130, Hitotsubashi University.

[17] Hayakawa, K. (2007), "A Simple Efficient Instrumental Variable Estimator in Panel AR(p) Models," Hi-Stat Discussion Paper No.213, Hitotsubashi University.

[18] Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, Vol. 56, 1371-1395.

[19] Hsiao, Cheng (2003), *Analysis of Panel Data*, 2nd Edition, Cambridge University Press.

[20] Kunitomo, N. (1980), "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, Vol.75, 693-700.

[21] Kunitomo, N. (2008), "Improving the LIML estimation with many instruments and persistent heteroscedasticity," Discussion Paper CIRJE-F-576, Graduate School of Economics, University of Tokyo, cirje/research/dp/2008), Revised.

APPENDIX : Some Figures

In Figures the distribution functions of the GMM and the LIML estimators are shown with the large sample normalization (i.e. the case of $c = 0$) and the large-K normalization (i.e. the case of $c > 0$). The limiting distributions for the LIML estimator in the large-K asymptotics are $N_2(\mathbf{0}, \mathbf{I}_2)$ and its marginal distributions are $N(0,1)$ as $n \to \infty$, which are denoted as "o". For the sake of comparisons, the distribution functions of the GMM estimator are normalized in the same way and presented in figures. The parameters of our settings and the details of numerical computation method are similar to those explained in Anderson et al. (2005, 2008a), Akashi (2008), Akashi and Kunitomo (2010).

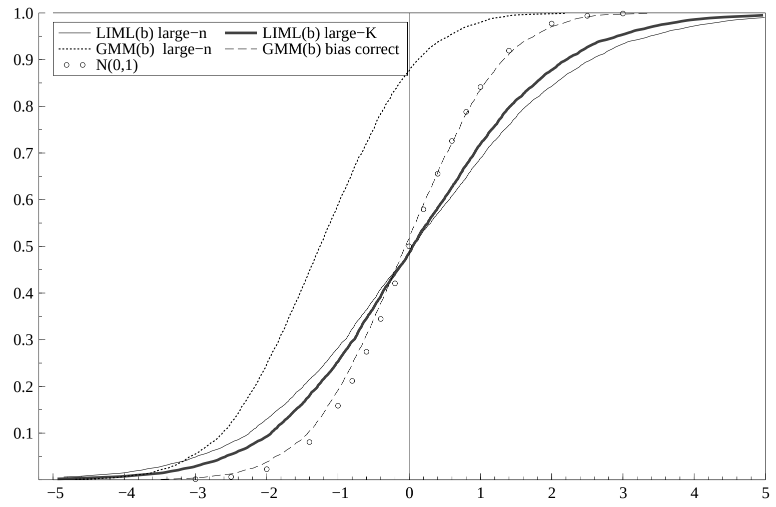

Figure 1: $\beta_2: \ N = 100, \ T = 25, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

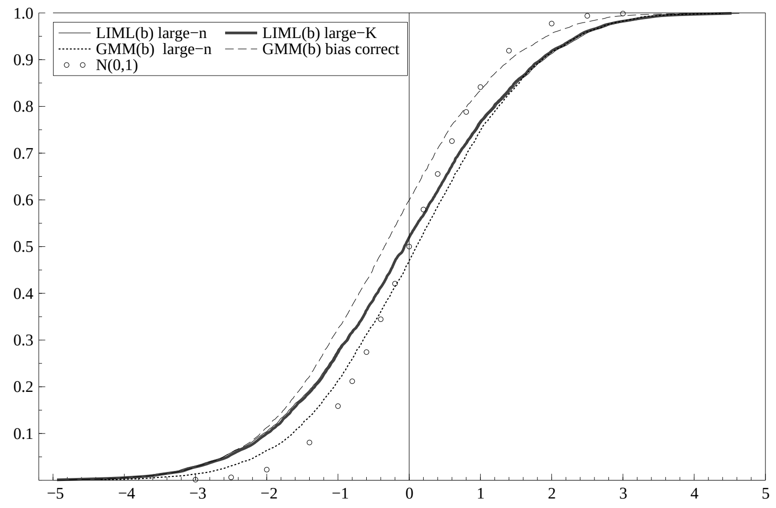

Figure 2: $\gamma_{11}: \ N = 100, \ T = 25, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

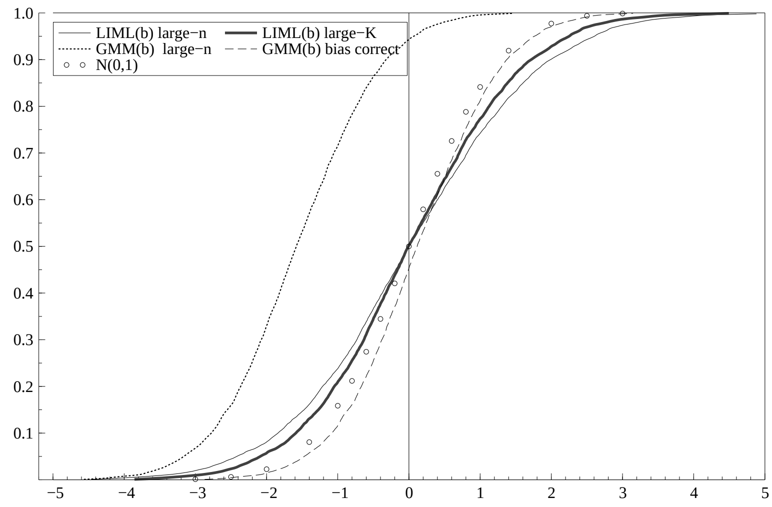

Figure 3: $\beta_2: \ N = 100, \ T = 50, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

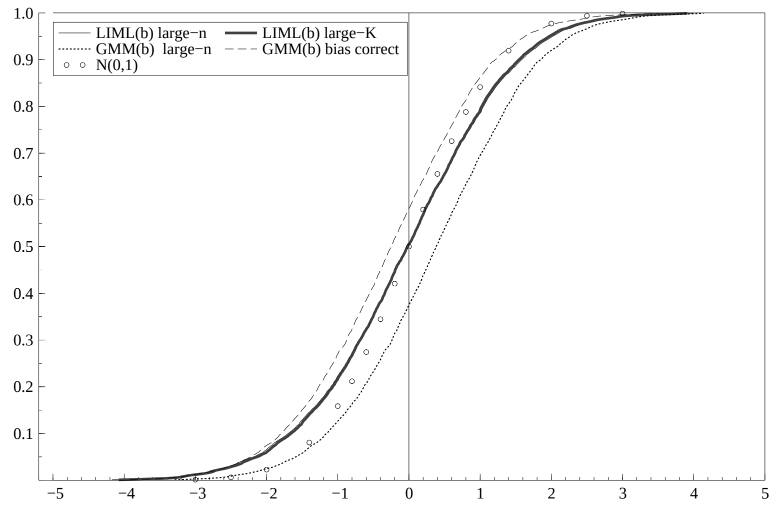

Figure 4: $\gamma_{11}: \ N = 100, \ T = 50, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

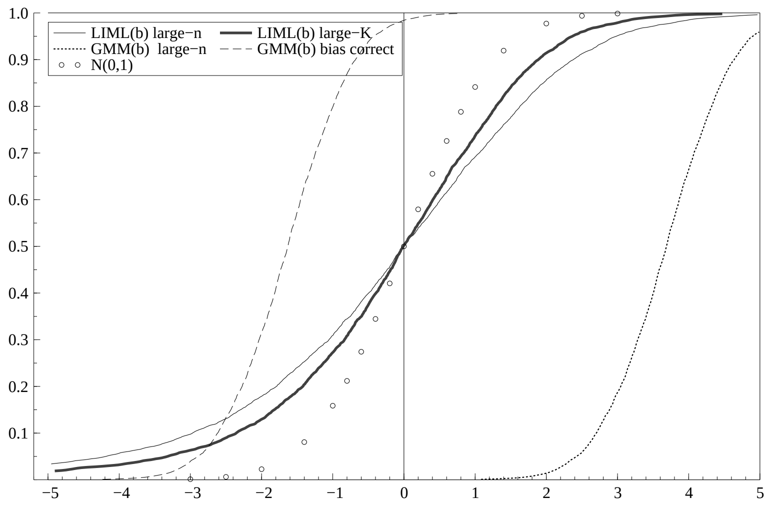

Figure 5: $\beta_2: \ N = 100, \ T = 25, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.5, 1.0)$

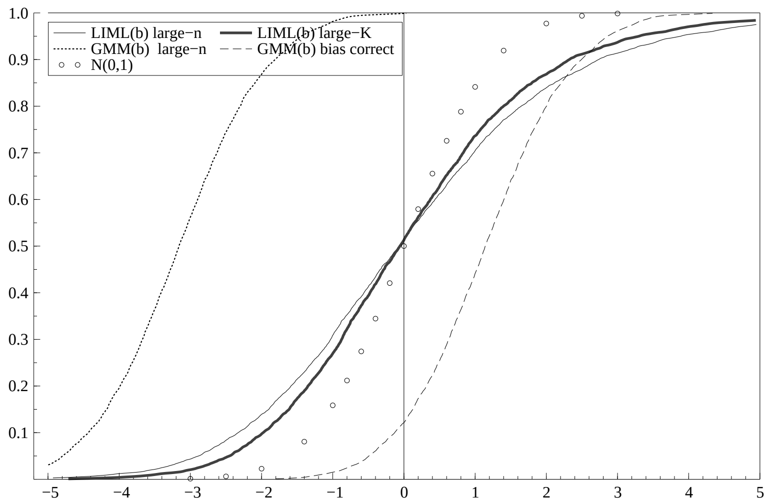

Figure 6: $\gamma_{11}: \ N = 100, \ T = 25, \ c_b = \frac{4}{100}, \ (\omega_{11}, \omega_{12}) = (1.5, 1.0)$

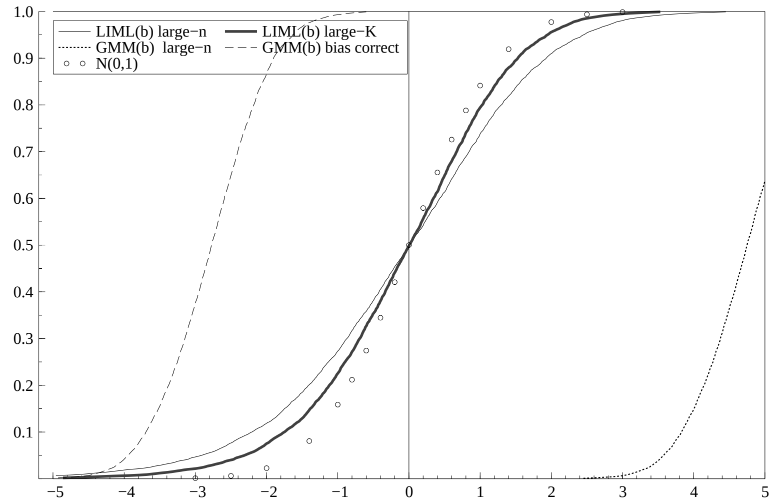

Figure 7: $\beta_2:\ N=100,\ T=50,\ c_b=\frac{4}{100},\ (\omega_{11},\omega_{12})=(1.5,1.0)$

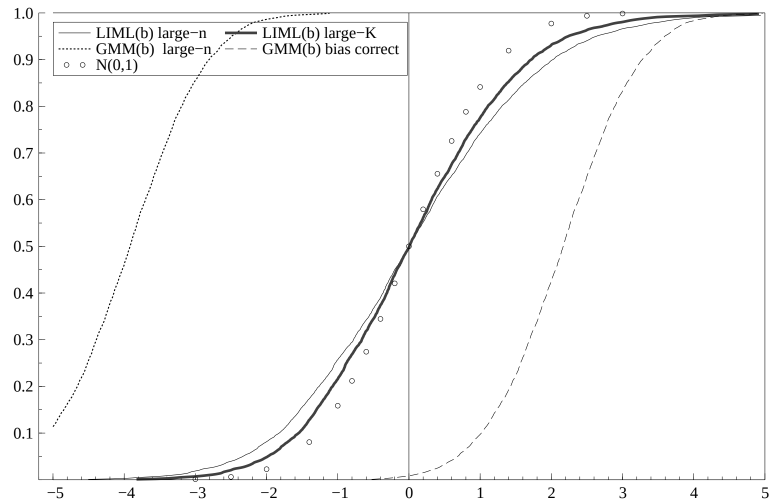

Figure 8: $\gamma_{11}:\ N=100,\ T=50,\ c_b=\frac{4}{100},\ (\omega_{11},\omega_{12})=(1.5,1.0)$

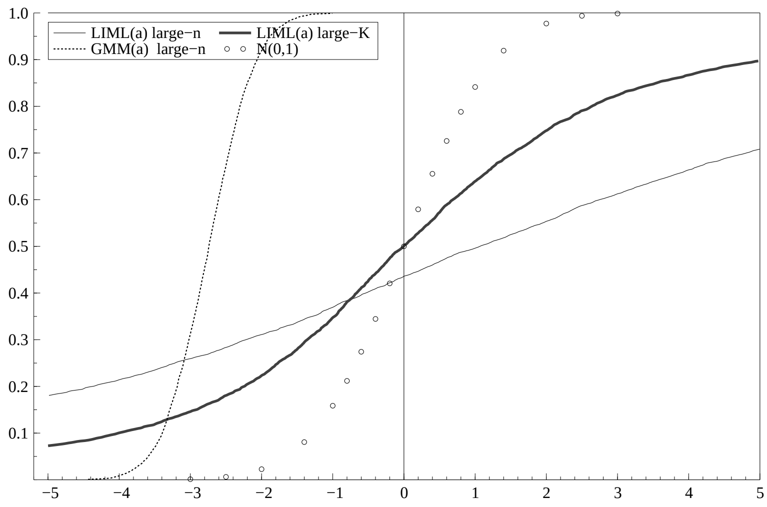

Figure 9: $\beta_2: \ N = 100, \ T = 25, \ c_a = \frac{3}{2}\frac{25}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

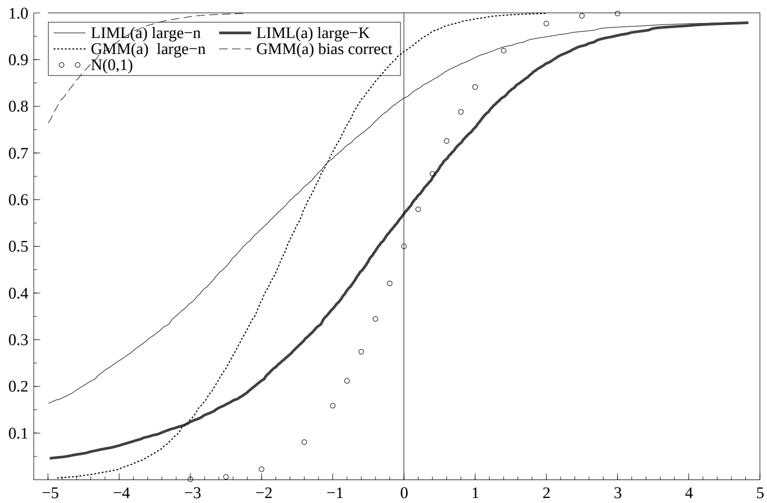

Figure 10: $\gamma_{11}: \ N = 100, \ T = 25, \ c_a = \frac{3}{2}\frac{25}{100}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

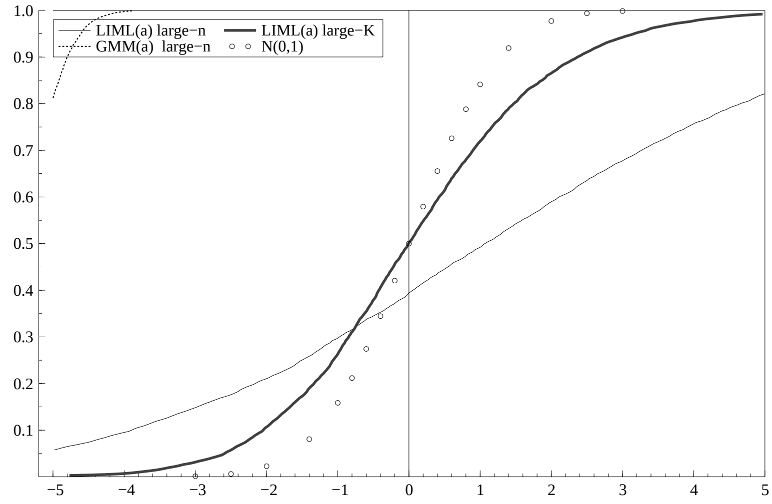

Figure 11: $\beta_2: \ N = 200, \ T = 50, \ c_a = \frac{3}{2}\frac{50}{200}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$

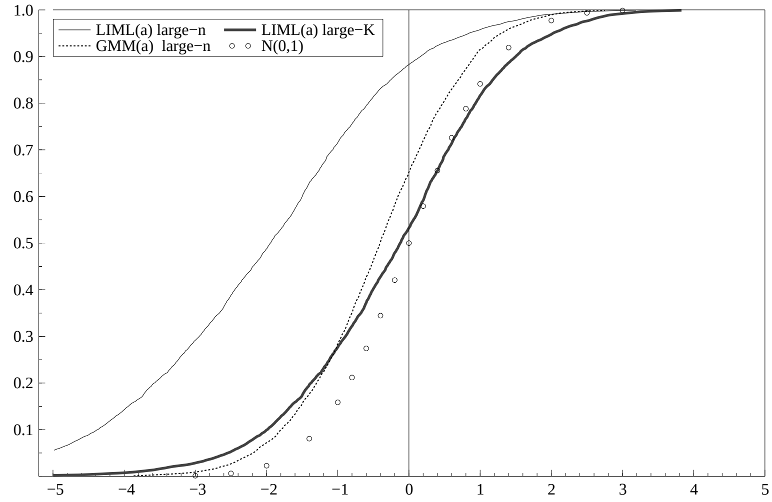

Figure 12: $\gamma_{11}: \ N = 200, \ T = 50, \ c_a = \frac{3}{2}\frac{50}{200}, \ (\omega_{11}, \omega_{12}) = (1.0, 0.3)$