CIRJE-F-780

# An Aysmptotically Optimal Modification of the Panel LIML Estimation for Individual Heteroscedasticity

Naoto Kunitomo
University of Tokyo

Kentaro Akashi
Institute of Statistical Mathematics

December 2010

# An Aysmptotically Optimal Modification of the Panel LIML Estimation for Individual Heteroscedasticity *

Naoto Kunitomo †

and

Kentaro Akashi ‡

December 19, 2010

**Abstract**

We consider the estimation of coefficients of a dynamic panel structural equation in the simultaneous equation models. As a semi-parametric method, we introduce a class of modifications of the limited information maximum likelihood (LIML) estimator to improve its asymptotic properties as well as the small sample properties when we have individual heteroscedasticities. We shall show that an asymptotically optimal modification of the LIML estimator, which is called AOM-LIML, removes the asymptotic bias caused by the forward-filtering and improves the LIML and other estimation methods with individual heteroscedasticities.

**Keywords**

Dynamic Panel Structural Equation, Individual Heteroscedasticity, AOM-LIML Estimation, Forward-filtering, Asymptotic Optimality.

---

*KA10-12-19. This is a preliminary manuscript.

†Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1 Tokyo, JAPAN, kunitomo@e.u-tokyo.ac.jp

‡Institute of Statistical Mathematics, Midori-cho 10-3, Tachikawa-shi, Tokyo 190-8562, JAPAN

# 1 Introduction

Recently there has been a growing interest on dynamic panel models, which are indispensable tools for econometric analysis. As a pioneering work Alvarez and Arellano (2003) have investigated the asymptotic properties of alternative estimation methods including the WG (With-in-Groups) estimator, the GMM (generalized method of moments) estimator and the LIML (limited information maximum likelihood) estimator for a simple reduced form equation when the time length is large(i.e. the long panel asymptotics). The asymptotic properties of estimators in the long panel model are closely related to the corresponding methods for the structural equation estimation with many instruments. There has been a considerable literatures on this problem and Anderson, Kunitomo and Matsushita (2010) have found that the LIML estimator, originally developed by Anderson and Rubin (1949, 1950), has some optimal properties as the estimation of structural equation when there are many instruments. In dynamic panel context, Alvarez and Arellano (2003) have developed the use of forward-filtered estimation and obtained the asymptotic properties of alternative estimators (when both the numbers of individuals and the time length are large) while Hayakawa (2006, 2007) have investigated the backward filtered instruments for estimating dynamic panel reduced form models. For estimating the panel structural equation in the simultaneous equation models, Akashi and Kunitomo (2010a, b) have investigated the asymptotic and the finite sample properties of alternative estimators (the WG, the GMM and LIML estimators) when we use the forward-filtered panel data and the backward-filtered instruments. Contrary to some known results in the econometric literature, they have found that the asymptotic biases of the WG estimator and the GMM estimator are rather significant and they should not be ignored while the distribution function of the LIML estimator is almost median-unbiased under a set of reasonable conditions.

One remaining issue, however, is the fact that the asymptotic bias of the LIML estimator can be significant when we have heteroscedastic disturbances of the structural equation in some situations. In the panel econometric analysis the homoscedasticity assumption of disturbances has been often made, but in practical panel data we often have the situation when the heteroscedasticities beyond the standard formulation of the individual effects are present. The effects of the similar bias problem has been investigated independently by Chao, Swanson, Hausman, Newey and Woutersen (2009), and Kunitomo (2008) for the estimation problem of structural equations. Some modifications of the instrumental variables (IV) estimation and the LIML estimation have been considered. For the problem of estimating the structural equation with many instruments, Kunitomo (2008) has shown that a particular modification of the LIML estimation has some asymptotic optimality among a class of estimators when the disturbances have the *persistent heteroscedasticities* in the classical structural equations and the reduced form equations. In the practical panel equation problem it may be often natural to have heteroscedastic disturbances over different individuals and then it may be important to investigate the effects of individual heteroscedasticities on the alternative estimation methods.

In this paper we shall investigate the asymptotic and finite sample properties of the modified LIML estimators in the dynamic panel structural equation models. In particular, we shall investigate an asymptotically optimal modification of the LIML (AOM-LIML) estimator proposed by Kunitomo (2008) and extend it to the problem of the dynamic panel structural equation models. We shall show that the AOM-LIML estimation reduces the possible asymptotic bias caused by the presence of many instruments and the heteroscedasticities of disturbances. We also show that the AOM-LIML estimator is free from the asymptotic bias caused by the forward-filtering of panel data, which was proposed by Alvarez and Arellano (2003). In addition to the asymptotic bias problem we shall show that the AOM-LIML estimator attains the asymptotic bound among a class of estimators including the corresponding the fixed-effects instrumental variables estimation methods. Therefore, when the time length of panel data is not small with *long panels*, the modified panel LIML estimation we shall introduce can solve the problems of individual effects, the dynamic effects, the endogeneity problem and the individual heteroscedastic disturbances at the same time. The AOL-LIML estimator proposed in this paper will be a promising estimation method for practical panel applications.

In Section 2 we formulate the dynamic panel structural equation models and define the alternative estimation methods for unknown parameters in the panel structural equations with the (possible) heteroscedastic disturbances. Then in Section 3 we give the main results on the asymptotic properties of the modified Panel LIML estimation method. In Section 4 we shall discuss some finite sample properties of the modified GMM and the LIML estimators with the original GMM and LIML estimators, which have been investigated by Akashi and Kunitomo (2010a, b) based on a set of Monte Carlo simulations. Concluding remarks will be given in Section 5. The proofs of our theorems will be given in Section 6. Some figures of the distribution functions of the normalized estimators will be given in Appendix.

# 2 Dynamic Panel Structural Equations with Individual Heteroscedasticity

## 2.1 Individual Heteroscedasticity

We consider the dynamic panel structural equation for $i = 1, ..., N;\ t = 1, ..., T$ given by

$$y_{it}^{(1)} = \boldsymbol{\beta}_2' \boldsymbol{y}_{it}^{(2)} + \boldsymbol{\gamma}_1' \mathbf{z}_{it-1}^{(1)} + \eta_i + u_{it}\ , \tag{2.1}$$

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 $K_1$ vector of the included predetermined variables, 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 that the vector $\mathbf{z}_{it-1}^{(1)}$ consists of the lagged endogenous variables $y_{it-p_g}^{(g)}$ $(0 < p_g < \infty\ ;\ g = 1, \cdots, K_*)$ and the original sample saize is $NT(= n)$. Also $\eta_i$ $(i = 1, \cdots, N)$ are individual effects and we assume that $u_{it}$ are mutually independent over individual and periods with $\mathcal{E}[u_{it}] = 0$ and $\mathcal{E}[u_{it}^2] = \sigma^2$ (if the disturbances are homoscedastic).

We shall investigate the case when the reduced form equation with (2.1) can be represented as

$$\mathbf{y}_{it} = \boldsymbol{\pi}_i + \boldsymbol{\Pi} \mathbf{z}_{it-1} + \mathbf{v}_{it}\ , \tag{2.2}$$

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, $\boldsymbol{\Pi}$ and $\boldsymbol{\pi}_i$ are a $(1+G_2) \times K$ coefficients matrix and a $(1 + G_2) \times 1$ individual effects vector. Also $\mathcal{E}_t(\mathbf{v}_{it}) = 0$, $\mathcal{E}_t(\mathbf{v}_{it}\mathbf{v}_{it}') = \boldsymbol{\Omega}_i$ and $\mathcal{E}_t[\ .\ ]$ is the conditional expectation given the $\sigma$-field $\mathcal{F}_{t-1}$, which is generated by the set of random variables $\boldsymbol{\pi}_i$, $\mathbf{z}_{it-h}$ and $\mathbf{v}_{it-h}$ $(h > 0)$. We assume the condition

$$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Omega}_i \xrightarrow{p} \boldsymbol{\Omega} \tag{2.3}$$

and $\boldsymbol{\Omega}$ is a positive definite (constant) matrix.
The *extended reduced form*, or the vector AR(1) representation satisfies

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

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

where $\boldsymbol{\Pi}^*$ is the $K^* \times K^*$ autoregressive coefficients $(K^* = K + K_3)$, $\boldsymbol{\pi}_i^*$ and $\mathbf{v}_{it}^*$ represent the $K^* \times 1$ individual effects and the disturbances, respectively, and the $K_3$-variables are excluded from the $(1 + G_2)$ reduced form equations. Also $\mathbf{J}_{1+G_2}'$

is an $(1+G_2) \times K^*$ appropriate selection matrix whose each element is one or zero, and then the selection matrices $\mathbf{J}'_{K_1}$ and $\mathbf{J}'_K$ are accordingly defined. We define $K_*$ as the number of the distinct autoregressive variables in $\mathbf{z}_{it-1}$ and then $K_* \leq K \leq K^*$. These notations of this paper agree with Akashi and Kunitomo (2010b).

We consider the effects of the *individual (conditional) heteroscedastic disturbances* in the panel structural equation model, for which we write

$$\mathbf{v}^*_{it} = \boldsymbol{h}_i \circ \mathbf{v}^{**}_{it} = (h_i^{(1)} v_{it}^{**(1)}, \cdots, h_i^{(K^*)} v_{it}^{**(K^*)})' \ , \tag{2.6}$$

where $\circ$ denotes the Hadamard product, $\boldsymbol{h}_i = (h_i^{(1)}, ..., h^{(K^*)})'$ and $\mathbf{v}^{**}_{it}$ are the $K^* \times 1$ individual heteroscedasticities and the disturbance terms in the extended reduced form, respectively. Then the reduced form (2.4) can be re-written as

$$\mathbf{z}^*_{it} = \boldsymbol{\pi}^*_i + \boldsymbol{\Pi}^* \mathbf{z}^*_{it-1} + \boldsymbol{h}_i \circ \mathbf{v}^{**}_{it} \ , \tag{2.7}$$

where we have $\mathcal{E}_t[\boldsymbol{h}_i \circ \mathbf{v}^{**}_{it}] = \mathbf{0}$ and

$$\boldsymbol{\Omega}^*_i \ = \ \mathcal{E}_t[(\boldsymbol{h}_i \circ \mathbf{v}^{**}_{it})(\boldsymbol{h}_i \circ \mathbf{v}^{**}_{it})'] = \boldsymbol{\Omega}^{**}_v \circ \boldsymbol{h}_i \boldsymbol{h}'_i \ , \tag{2.8}$$

$$\sigma^2_i \ = \ \mathcal{E}_t[u^2_i] = \boldsymbol{\beta}'[\mathbf{J}'_{1+G_2} \boldsymbol{\Omega}^*_i \mathbf{J}_{1+G_2}] \boldsymbol{\beta} \ . \tag{2.9}$$

and $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}'_2)'$.

Since we have (2.1) and (2.2), the relation on the coefficients gives the condition $(1, -\boldsymbol{\beta}'_2)\boldsymbol{\Pi} = (\boldsymbol{\gamma}'_1, \mathbf{0}')$ and $\boldsymbol{\pi}'_{12} = \boldsymbol{\beta}'_2 \boldsymbol{\Pi}_{22}$, where $\boldsymbol{\Pi}'_1 = (\boldsymbol{\pi}_{11}, \boldsymbol{\Pi}'_{21})$ is a $K_1 \times (1+G_2)$ matrix, $\boldsymbol{\Pi}'_2 = (\boldsymbol{\pi}_{12}, \boldsymbol{\Pi}'_{22})$ is a $K_2 \times (1+G_2)$ matrix. Then we repress the the $(1+G_2) \times (K_1 + K_2)$ partitioned matrix of coefficients as

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\pi}'_{11} & \boldsymbol{\pi}'_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} = [\mathbf{J}'_{1+G_2} \boldsymbol{\Pi}^* \mathbf{J}_{K,K_3}]_{(1+G_2) \times K} \ , \tag{2.10}$$

where $\mathbf{J}'_{K,K_3}$ is a $K^* \times K^*$ selection matrix for re-ordering rows of $\mathbf{z}^*_{it-1}$ as ($K$-variables and $K_3$-variables), and $[\ .\ ]_{(1+G_2) \times K}$ stands for the $(1+G_2) \times K$ sub-matrix of the corresponding matrix.

We notice that although we may call (2.4) and (2.5) as the *reduced form*, the predetermined variables in $\mathbf{z}_{it-1}$ satisfy $\mathcal{E}[\mathbf{z}_{it-1} \mathbf{v}'_{it}] = \mathbf{O}$ and they are correlated with the unobserved variables $(\boldsymbol{\pi}_i + \mathbf{v}_{it})$ since $\mathcal{E}[\mathbf{z}_{it-1} \boldsymbol{\pi}'_i] \neq \mathbf{O}$ in the general case. Hence this aspect makes the panel structural equation model of (2.1) and (2.4) different from the structural equation in the classical simultaneous equation models.

## 2.2 Filtering, Instruments and the modified LIML estimation

### 2.2.1 The Forward and Backward Filters

We define the *forward-filter* by the transformation that for any variables $x_{it}$ $(i = 1, \cdots, N; t = 1, \cdots, T)$ the transformed variables $x_{it}^{(f)} (i = 1, \cdots, N; t = 1, \cdots, T-1)$ have the form

$$x_{it}^{(f)} = c_t[x_{it} - (\frac{1}{T-t})(x_{it+1} + \cdots + x_{iT})] \tag{2.11}$$

with $c_t^2 = (T-t)/(T-t+1)$.
When we apply the forward-filter to the structural equation (2.1) for $t = 1, \cdots, T-1$, the transformed structural equation does not have individual effects as

$$y_{it}^{(1,f)} = \boldsymbol{\beta}_2' \boldsymbol{y}_{it}^{(2,f)} + \boldsymbol{\gamma}_1' \boldsymbol{z}_{it-1}^{(1,f)} + u_{it}^{(f)} \ , \tag{2.12}$$

but consequently we have the relation that $\mathcal{E}[\boldsymbol{z}_{it-1}^{(1,f)} u_{it}^{(f)}] \neq \mathbf{0}$.
We also define the *backward-filter* to the $K \times 1$ vector of instrumental variables $\mathbf{z}_{it-1}$, which is the transformation of

$$\mathbf{z}_{it-1}^{(b)} = b_t\Big[\mathbf{z}_{it-1} - \frac{1}{t}(\mathbf{z}_{it-2} + \cdots + \mathbf{z}_{i0} + \mathbf{z}_{i(-1)})\Big] \ , \tag{2.13}$$

where $b_t^2 = t/(t+1)$ for $t = 1, \cdots, T-1$.
If we use the backward-filtered instruments of $\mathbf{z}_{it}$, then we have the orthogonal conditions $\mathcal{E}[\boldsymbol{z}_{it-1}^{(b)} u_{it}^{(f)}] = \mathbf{0}$.

As the initial condition we include $\mathbf{z}_{i(-1)}$ in order to simplify the notation of the index range. By using the notations used in Akashi and Kunitomo (2010b), we shall consider two types of the sets of instrumental variables. For each case the matrix of instrumental variables at period $t$ can be expressed as

$$\mathbf{Z}_t^{(a)} = \begin{pmatrix} \mathbf{z}_{1(t-1)}^{(a)} & \cdots & \mathbf{z}_{N(t-1)}^{(a)} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_{10}^{(a)} & \cdots & \mathbf{z}_{N0}^{(a)} \end{pmatrix}' \ , \ \mathbf{Z}_t^{(b)} = \left(\mathbf{z}_{1(t-1)}^{(b)}, \cdots, \mathbf{z}_{N(t-1)}^{(b)}\right)' \ , \tag{2.14}$$

where $\mathbf{z}_{it-1}^{(a)}$ is the $K_* \times 1$ vector such that $\mathbf{z}_{it-1}^{(a)} = \mathbf{J}_{K_*}\mathbf{z}_{it-1}$ and the selection matrix $\mathbf{J}_{K_*}$ chooses the nearest lagged variables to $t-1$ as autoregressive variables. The reduction $K^*$ to $K_*$ is needed to be of full rank for $(\mathbf{Z}_t^{(a)'}\mathbf{Z}_t^{(a)})$. The first matrix $\mathbf{Z}_t^{(a)}$ is the $N \times (K_* t)$ and the second matrix $\mathbf{Z}_t^{(b)}$ is the $N \times K$ matrix, and then we shall investigate two possible use of the instrumental variables, which correspond to two methods :

($a$) At period $t$ we use all available lagged variables,

($b$) At period $t$ we use the only lagged variables included in the reduced form after the backward filtering.

The order of the number of instruments for the case ($a$) becomes $O(T^2)$ and the dynamic GMM method with this filtering has been used in many empirical applications. The case ($b$) is equivalent to use an optimal instrumetal variable method as pointed out by Hayakawa (2007). The asymptotic theory, which corresponds to Case (a) and it will be called the large-$K$ asymptotics, explicitly depends on the following ratio in the case of homogenous disturbances and implicitly the resulting asymptotics depends on it in the case of heteroscedastic disturbance. Under the double asymptotics $N,T \to \infty$, the ratios of the number of orthogonal conditions to the total sample $NT$ in two cases are given by

$$(a)\ \frac{K_* T(T-1)}{2N(T-1)} \underset{N,T\to\infty}{\to} c_a = (\frac{K_*}{2}) \lim_{N,T\to\infty} (\frac{T}{N}) \ , \tag{2.15}$$

$$(b)\ \frac{K(T-1)}{N(T-1)} \underset{N,T\to\infty}{\to} c_b = \frac{K}{N} = 0 \ , \tag{2.16}$$

respectively.

#### 2.2.2 The Modified Panel LIML and GMM Estimators

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

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

be $(1+G_2)\times N$, and $K_1 \times N$ matrices of the forward-filtered variables, respectively. In the present situation an asymptotically optimal modification of LIML (AOM-LIML) estimation proposed by Kunitomo (2008) 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 [1] by

$$\mathbf{G}^{(f.m)} = \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} \mathbf{M}_{t.m} \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{2.17}$$

and

$$\mathbf{H}^{(f.m)} = \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{pmatrix} [\mathbf{I}_N - \mathbf{M}_{t.m}] \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right), \tag{2.18}$$

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

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 panel LIML estimator (we call it as AOM-LIML) 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

$$\left[\frac{1}{n}\mathbf{G}^{(f.m)} - \frac{1}{q_n}\lambda_n \mathbf{H}^{(f.m)}\right]\left[\begin{array}{c} 1 \\ -\hat{\boldsymbol{\theta}}_{MLI} \end{array}\right] = \mathbf{0}\ , \tag{2.19}$$

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

$$|\frac{1}{n}\mathbf{G}^{(f.m)} - l\frac{1}{q_n}\mathbf{H}^{(f.m)}| = 0\ . \tag{2.20}$$

By using these notations, we shall denote $\hat{\boldsymbol{\theta}}_{MLI}^{(.)} = \hat{\boldsymbol{\theta}}_{MLI}^{(a)}$ for the case of the forward-filter with $(a)$ and $\hat{\boldsymbol{\theta}}_{MLI}^{(b)}$ for the backward-filter with $(b)$, respectively. Also for both the forward-filter and the backward-filter cases we define the modified GMM estimator, $\hat{\boldsymbol{\theta}}_{MGM}^{(.)}\ = (\hat{\boldsymbol{\beta}}_{2.MGM}^{'}, \hat{\boldsymbol{\gamma}}_{1.MGM}^{'})'$ of $(1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})' = (1, -\boldsymbol{\theta}')'$ by

$$[\mathbf{0}, \mathbf{I}_{G_2+K_1}]\sum_{t=1}^{T-1}\left[\begin{array}{c} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_{t-1}^{(1,f)'} \end{array}\right]\mathbf{M}_{t,m}\left[\mathbf{Y}_t^{(f)}, \mathbf{Z}_{t-1}^{(1,f)}\right]\left[\begin{array}{c} 1 \\ -\hat{\boldsymbol{\theta}}_{MGM}^{(.)} \end{array}\right] = \mathbf{0} \tag{2.21}$$

and we denote $\hat{\boldsymbol{\theta}}_{MGM}^{(a)}$ and $\hat{\boldsymbol{\theta}}_{MGM}^{(b)}$, respectively.

The LIML estimation method was originally developed by Anderson and Rubin (1949), and here we modify it slightly to define the modified LIML method and GMM for the dynamic panel structural equation models with individual heteroscedasticities. We notice that the Panel LIML estimation can be derived without assuming any particular distribution such as the normality for disturbances, which is different from the original formulation.

In order to motivate the method developed in this section, we shall illustrate why the particular modification of the LIML estimation proposed here can work well. By utilizing the notations in Section 3 and Section 6 below, we can evaluate the *key* asymptotic bias as the first order asymptotic behavior of the modified panel LIML estimator. For instance, we take the case (a) and then the expected value of the leading term of the stochastic expansion for $\hat{\boldsymbol{\theta}}_{MLI}^{(a)} - \boldsymbol{\theta}$ with the modified LIML estimator is approximately represented as

$$\mathcal{AE}\left[\hat{\boldsymbol{\theta}}_{MLI}^{(a)} - \boldsymbol{\theta}\right] \propto \boldsymbol{\Phi}^{*-1}\left[\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}(\mathbf{D}'\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{u}_t^{(f)}) + \frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}(\mathbf{U}_t^{(\perp,f)'}\mathbf{M}_t^D\mathbf{u}_t^{(f)})\right] \tag{2.22}$$

and

$$\mathbf{M}_t^D = \mathbf{M}_t^{(a)} - \mathbf{D}_t^{(a)}, \tag{2.23}$$

where $\mathcal{AE}[\,\cdot\,]$ is the expectation operator with respect to the limiting distribution of the modified LIML estimator, $\mathbf{M}_t^D$ corresponds to $\mathbf{M}_{t,m}$ with $\mathbf{M}_t = \mathbf{M}_t^{(a)}$, $\mathbf{D}_t^{(a)} = \mathbf{I}_N \circ \mathbf{M}_t^{(a)}$ (we shall use the notation $\mathbf{D}_t^{(b)} = \mathbf{I}_N \circ \mathbf{M}_t^{(b)}$ in the same way), $\mathbf{\Phi}^*$ and $\mathbf{D}$ are some constant matrices (which will be given by (3.5) and (3.12) in Theorem 3.1 and Theorem 3.2). The random matrices $\mathbf{Z}_{t-1}^{(f)} = (z_{it-1}^{(f)})$ and $\mathbf{u}_t^{(f)} = (u_{it}^{(f)})$ are the forward-filtered random matrices of $\mathbf{Z}_{t-1}$ and $\mathbf{U}_t^{\perp,f}$ $(= (u_{it}^{\perp,f}))$ is the filtered random matrix with each row vectors being orthogonal to $\mathbf{u}_t$ in the case of homogeneous disturbances. (These notations will be precisely defined by (3.8) and (3.9) in Section 3 and used in Section 6.) It is important to notice that the LIML estimator based on $\mathbf{M}_t^D$ instead of $\mathbf{M}_{t,m}$ in Equations (2.16) and (2.17) is equivalent to the AOL-LIML estimator.

By evaluating the conditional expectations of (2.22), we find that

$$\mathcal{AE}\left[\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MLI}^{(a)} - \boldsymbol{\theta})\right] = o(1)\ . \tag{2.24}$$

The most important aspect is the finding that the asymptotic bias of the LIML estimator obtained in Theorems 3.1 and Theorem 3.2 in Akashi and Kunitomo (2010b) disappears if we modify the original LIML estimation slightly. It should be noted that the above phenomenon is true not only for the case (a), but also for the case (b), which has some implication for applications.

In the case of homoscedasticity, the panel LIML estimator is consistent under many instruments mainly because $\mathcal{E}_t[\mathbf{u}_{it}^{\perp} u_{it}] = \mathbf{0}$. However, under the heteroscedasticities we generally observe $\mathcal{E}_t[\mathbf{u}_{it}^{\perp} u_{it}] \neq \mathbf{0}$. For the LIML estimator, the corresponding first-term to (2.21) has an asymptotic bias term $O(1/\sqrt{NT})$ due to the forward-filtering, and the corresponding second-term has the bias term $O(1)$ due to the heteroscedasticities of disturbances. However, the AOM-LIML estimator does not have such asymptotic bias terms. It is because (i) off-diagonal elements of $\mathcal{E}_t[\mathbf{u}_t^{(f)} \mathbf{z}_{t-1}^{(f)\prime}]$ and $\mathcal{E}_t[\mathbf{u}_t^{(f)} \mathbf{u}_t^{(\perp,f)\prime}]$ are 0 by i.i.d., (ii) the diagonal elements of $\mathbf{M}_t^D$ are $o_p(1)$ by construction. It is essential to remove the asymptotic bias in this method that the off-diagonal elements of $\mathcal{E}_t[\mathbf{u}_t^{(f)} \mathbf{z}_{t-1}^{(f)\prime}]$ and $\mathcal{E}_t[\mathbf{u}_t^{(f)} \mathbf{u}_t^{(\perp,f)\prime}]$ are $o_p(1)$.

# 3 Asymptotic Distribution of the Panel AOM-LIML Estimator

In order to investigate the asymptotic properties of the class of modified LIML and GMM estimators, we make a set of assumptions. They may be standard in the panel structural equation except the fact we shall consider the situation when the individual disturbances have heteroscedasticities.

**(A-I)** The random vectors $\{\mathbf{v}_{it}^{**}\}$ $(i=1,\cdots,N;\ t=1,\cdots,T)$ are i.i.d. across time and individuals, which are independent of the indivisual effects $\bar{\boldsymbol{\pi}}_i^*$ ($\bar{\boldsymbol{\pi}}_i^*$ are i.i.d. random variables across individuals) and $\mathbf{z}_{i0}^*$, and $\mathcal{E}[\mathbf{v}_{it}^{**}]=\mathbf{0}$, $\mathcal{E}[\mathbf{v}_{it}^{**}\mathbf{v}_{it}^{**'}]=\boldsymbol{\Omega}_v^{**}$ with $\mathbf{J}_{1+G_2}'\boldsymbol{\Omega}_v^{**}\mathbf{J}_{1+G_2}>0$ (positive definite) and $\mathcal{E}[\|\mathbf{v}_{it}^{**}\|^8]$ exists.

**(A-II)** (i) The random vectors $\mathbf{v}_{it}^*$ $(t=1,\cdots,T)$ are conditionally independent given $\boldsymbol{h}_i$, (ii) $\|\boldsymbol{h}_i\|$ are bounded and there exist $M_1,M_2$ $(M_2>M_1>0)$ such that $0<M_1<h_i^{(k)}<M_2$, $(i=1,\cdots,N;k=1,\cdots,K^*)$ and $(1/N)\sum_{i=1}^N\boldsymbol{h}_i\boldsymbol{h}_i'\xrightarrow{p}\boldsymbol{\Omega}_h$ (positive definite) as $N\to\infty$.

**(A-III)** (i) The initial observations satisfy

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

where $\mathbf{w}_{i0}$ is independent of $\boldsymbol{\pi}_i^*$ and i.i.d. with the steady state distribution of the homogenous process so that $\mathbf{w}_{i0}=\sum_{j=0}^{\infty}\boldsymbol{\Pi}^{*j}\mathbf{v}_{i(0-j)}^*$ with $\mathbf{J}_K'\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}']\mathbf{J}_K>0$ (positive definite). (ii) All eigenvalues $\lambda_k$ of

$$|\boldsymbol{\Pi}^*-\lambda\mathbf{I}_{K^*}|=0\tag{3.1}$$

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

The assumptions (A-I) and (A-III) are analogous to the conditions used by Alvarez and Arellano (2003), and Akashi and Kunitomo (2010b). They imply that the underlying processes for $\{\mathbf{y}_{it}\}$ after applying the forward filtering are stationary and we impose the sufficient moment conditions. Define the underlying process $\{\mathbf{w}_{it}\}$ for $\mathbf{y}_{it}$ satisfying

$$\mathbf{w}_{it}=\boldsymbol{\Pi}^*\mathbf{w}_{it-1}+\mathbf{v}_{it}^*\ .\tag{3.2}$$

Then (A-III) means that $\mathbf{w}_{it}$ for each $i$ has a stationary solution. For (A-II) we can set $M_2=1$ without loss of generality. The assumption (i) of (A-I) describes the orthogonal conditions for estimation and covariance stationarity of $\mathbf{w}_{it}$. The condition (ii) will be used for evaluating the asymptotic distribution of estimators. We can regard $\{\boldsymbol{h}_i\}_{i=1}^N$ as the additional scaling parameters for the individual heteroscedasticities for keeping the orthogonal conditions and stationarity. Also we note that the conditions (A-I)-(A-III) can be certainly relaxed with some complications of our analysis derived in Section 6. The obvious extension would be the time dependent conditional heteroscedasticities.

We shall discuss the asymptotic properties of the modified estimators. We first state the result for the case $(a)$ under the condition on $K_*T < N$, which ensures the non-singularity of $(\mathbf{Z}_t^{(a)'}\mathbf{Z}_t^{(a)})$. The proof will be given in Section 6.

**Theorem 3.1** : Suppose the conditions of (A-I) to (A-III) hold. We assume that $K_*T < N$ and $c_a = \lim_{N,T\to\infty}(K_*/2)(T/N) < 1/2$ as $N,T \to \infty$. Then $\hat{\boldsymbol{\theta}}_{MLI}^{(a)} \xrightarrow{p} \boldsymbol{\theta}$ and

$$\sqrt{NT}\Big[\ \hat{\boldsymbol{\theta}}_{MLI}^{(a)} - \begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}\ \Big] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^{*(a)})\ , \tag{3.3}$$

where

$$\boldsymbol{\Psi}^{*(a)} = \boldsymbol{\Phi}^{*(a)-1}[\ \boldsymbol{\Psi}_1^{*(a)} + \begin{pmatrix} \mathbf{I}_{G_2} \\ \mathbf{O} \end{pmatrix} \boldsymbol{\Psi}_2^{*(a)}(\mathbf{I}_{G_2}, \mathbf{O})\ ]\boldsymbol{\Phi}^{*(a)-1}\ , \tag{3.4}$$

and $\boldsymbol{\Phi}^{*(a)}$, $\boldsymbol{\Psi}^{*(a)}$ and $\boldsymbol{\Psi}_i^{*(a)}$ $(i = 1, 2)$ are defined by

$$\boldsymbol{\Phi}^{*(a)} = \lim_{N,T\to\infty} \frac{1}{NT}\mathbf{D}'\mathbf{J}_K'\mathcal{E}\Big[\sum_{t=1}^{T-1}\sum_{i=1}^{N}(1 - m_{ii}^{(t,a)})\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\Big]\mathbf{J}_K\mathbf{D}\ , \tag{3.5}$$

$$\boldsymbol{\Psi}_1^{*(a)} = \lim_{N,T\to\infty} \frac{1}{NT}\mathbf{D}'\mathbf{J}_K'\mathcal{E}\left[\sum_{t=1}^{T-1}\sum_{i=1}^{N}\sigma_i^2(1 - m_{ii}^{(t,a)})^2\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\right]\mathbf{J}_K\mathbf{D}\ , \tag{3.6}$$

$$\boldsymbol{\Psi}_2^{*(a)} = \lim_{N,T\to\infty} \frac{1}{NT}\mathcal{E}\left[\sum_{t=1}^{T-1}\sum_{i,j=1}^{N}\Big(\sigma_i^2\mathcal{E}_t[\mathbf{u}_{jt}^{\perp}\mathbf{u}_{jt}^{\perp'}] + \mathcal{E}_t[\mathbf{u}_{it}^{\perp}u_{it}]\mathcal{E}_t[u_{jt}\mathbf{u}_{jt}^{\perp'}])(m_{ij}^{(t,a)}[1 - \delta_{ij}])^2\Big)\right]\ , \tag{3.7}$$

$$\mathbf{u}_{it}^{\perp} = [\ \mathbf{0}, \mathbf{I}_{G_2}\ ]\left[\mathbf{I}_{1+G_2} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\right]\mathbf{v}_{it}\ , \tag{3.8}$$

$$\mathbf{D} = \left[\boldsymbol{\Pi}_2, \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}\right]\ , \tag{3.9}$$

$\sigma_i^2 = \mathcal{E}_t[u_{it}^2]$, $\delta_{ij} = 1$ if $i = j$, $0$ $(i \neq j)$, and $\boldsymbol{\Phi}^{*(a)}$, $\boldsymbol{\Psi}_1^{*(a)}$ and $\boldsymbol{\Psi}_2^{*(a)}$ are well-defined and $\boldsymbol{\Phi}^{*(a)}$ is a positive definite (constant) matrix.

Next, we shall state our result for the case $(b)$ with the backward-filtering procedure for instruments. The proof will be given in Section 6.

**Theorem 3.2** : Suppose the conditions of (A-I) to (A-III) hold. Then as $N, T \to$

$\infty$, $\hat{\boldsymbol{\theta}}_{MLI}^{(b)} \xrightarrow{p} \boldsymbol{\theta}$ and

$$\sqrt{NT}\Big[\ \hat{\boldsymbol{\theta}}_{MLI}^{(b)} - \begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}\ \Big] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^{*(b)}) \ , \tag{3.10}$$

where $\sigma_i^2 = \mathcal{E}_t[u_{it}^2]$,

$$\boldsymbol{\Psi}^{*(b)} = \boldsymbol{\Phi}^{*(b)-1}\ \boldsymbol{\Psi}_1^{*(b)}\ \boldsymbol{\Phi}^{*(b)-1} \ , \tag{3.11}$$

and

$$\boldsymbol{\Phi}^{*(b)} = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{t=1}^{T-1} \sum_{i=1}^{N} \mathbf{D}'\mathbf{J}_K'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J}_K\mathbf{D} \ , \tag{3.12}$$

$$\boldsymbol{\Psi}_1^{*(b)} = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{t=1}^{T-1} \sum_{i=1}^{N} \mathbf{D}'\mathbf{J}_K'\mathcal{E}[\sigma_i^2\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J}_K\mathbf{D} \ , \tag{3.13}$$

are well-defined and $\boldsymbol{\Phi}^{*(b)}$ is a positive definite (constant) matrix.

**Remark 1:** In both theorems the asymptotic distributions do not have asymptotic biases and the asymptotic variances depend on up to the second order moments of disturbances while the panel LIML estimator possibly depends on the fourth moments in the general case. (See Section 3 of Akashi and Kunitomo (2010b).)

**Remark 2:** (i) If $T/N \to 0$ i.e., $c_a = 0$, then the formula for the case $(a)$ is simplified as

$$\boldsymbol{\Psi}^{*(a)} = \boldsymbol{\Phi}^{*(b)-1}\ \boldsymbol{\Psi}_1^{*(b)}\ \boldsymbol{\Phi}^{*(b)-1} \ , \tag{3.14}$$

which does not depend on $\boldsymbol{\pi}_i$.
(ii) When $c_a = 0, c_b = 0$ and the disturbances are homoscedastic ($\sigma_i^2 = \sigma^2$), the asymptotic variances of the LIML estimator, which was derived by Akashi and Kunitomo (2010b), and the modified LIML estimator are same as

$$\boldsymbol{\Psi}^{*(a)} = \boldsymbol{\Psi}^{*(b)} = \sigma^2\boldsymbol{\Phi}^{*(b)-1} \ . \tag{3.15}$$

if we adjust the asymptotic biase for the case (a) and the distributions of the estimators are centered at the vector of the true parameter vector in advance.
(iii) The order conditions for the cases of $c_a = 0$ or $c_b = 0$ are different. In fact we need only $N \to \infty$ for the case of (b). Additionally, the case $(b)$ allows the sequence $N < T$, while the case $(a)$ has the restriction that $K_* T < N$.

**Remark 3:** (i) Under $c_a > 0$ and the homoscedasticity of disturbances, the modified LIML estimator is relatively efficient than the modified GMM estimator. For the simplicity we take the case when $K^* = 1 + G_2$ and $\mathcal{E}_t(\mathbf{v}_{it}\mathbf{v}_{it}^{'} = \boldsymbol{\Omega}$. Then $\mathcal{E}_t[\mathbf{u}_{it}^{\perp}u_{it}]\mathcal{E}_t[u_{it}\mathbf{u}_{it}^{\perp'}] = \mathbf{O}$ and

$$\mathbf{J}_{G_2}'\left[\ \boldsymbol{\Omega} - \frac{\boldsymbol{\Omega\beta\beta'\Omega}}{\boldsymbol{\beta'\Omega\beta}}\ \right]\mathbf{J}_{G_2} = \mathcal{E}_t[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp'}] \leq \mathcal{E}_t[\mathbf{J}_{G_2}'\mathbf{v}_{it}\mathbf{v}'_{it}\mathbf{J}_{G_2}] = \mathbf{J}_{G_2}'\boldsymbol{\Omega}\mathbf{J}_{G_2}\ . \quad (3.16)$$

The corresponding first term of the GMM estimator is the same as $\boldsymbol{\Psi}_1^{*(a)}$ and the second term is given by

$$\boldsymbol{\Psi}_{2,MGM}^{*(a)} = \lim_{N,T\to\infty}\frac{1}{NT}\mathcal{E}\left[\mathbf{J}_{G_2}'[\sigma^2\boldsymbol{\Omega} + \boldsymbol{\Omega\beta\beta'\Omega}]\mathbf{J}_{G_2}\sum_{t=1}^{T-1}\sum_{i,j=1}^{N}(m_{ij}^{(t,a)}[1-\delta_{ij}])^2\right]. (3.17)$$

(ii) In the case of $(b)$ under the homoscedasticity, we do not need double asymptotics $N, T \to \infty$ as mentiond by Akashi and Kunitomo (2010b). The heteroscedasticity setting requires the condition $N \to \infty$ for some second moments convergence such as $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}_i\boldsymbol{h}_i^{'} \xrightarrow{p} \boldsymbol{\Omega}_h$ , as $N \to \infty$ when $\mathbf{h}_i$ are stochastic. However, under the homoscedasticity, the modified LIML estimator also can be apply fixed-$N$ asymptotics. Then, $c_b > 0$ in (2.16) and by the same arguments of (3.16) the modified LIML estimator is also relatively efficient than the modified GMM estimator in the case $(b)$.

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 $\mathbf{H}^{(f)}$. Because we consider the modifications of the LIML and GMM estimators, we shall consider a class of estimators which are some functions of $\mathbf{G}^{(f,m)}$ and $\mathbf{H}^{(f,m)}$ and we have some results on the asymptotic optimality. The proof is quite similar to Theorem 2 of Kunitomo (2008) and so it is omitted.

**Theorem 3.3** : In the panel dynamic structural equation models of (2.1) and (2.2), define the class of consistent estimators for $\boldsymbol{\theta} = (\boldsymbol{\beta}_2^{'}, \boldsymbol{\gamma}_1^{'})'$ by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\gamma}}_1 \end{pmatrix} = \phi(\mathbf{G}^{(f,m)}\ ,\ \mathbf{H}^{(f,m)})\ , \quad (3.18)$$

where $\phi$ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices $(1/n)\mathbf{G}^{(f,m)}$ and $(1/q_n)\mathbf{H}^{(f,m)}$.
Then either under the assumptions of *Theorem 3.1* or *Theorem 3.2*, as $N, T \to \infty$

$$\sqrt{NT}\begin{pmatrix} \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})\ , \quad (3.19)$$

where

$$\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^* \tag{3.20}$$

and $\boldsymbol{\Psi}^*$ ($\boldsymbol{\Psi}^{*(a)}$ or $\boldsymbol{\Psi}^{*(b)}$) is given in *Theorem 3.1* and *Theorem 3.2*. The AOM-LIML estimator attains the asymptotic bound.

This is a new result on the asymptotic efficiency of the LIML estimation for the dynamic panel structural equation models. It could be regarded as the extensions of *Theorem 4* of Anderson et al. (2010) and *Theorem 3.3* of Akashi and Kunitomo (2010b).

# 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 MSE 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* ($G_2 = 1,\ K_1 = 2,\ K_* = 3,\ K = 4,\ K^* = 5$) of Akashi and Kunitomo (2010b). We set the unknown parameters such as $(\beta_2,\ \gamma_{11}) = (.5, .5)$ $\gamma_{12} =\ .3$, and $(\omega_{11},\ \omega_{12},\ \omega_{22}) = (1.0, .3, 1.0)$. Also we control the variance of each components of $\boldsymbol{\pi}_i$ as 1, and generate $\boldsymbol{h}_i$ from uniform distribution $u[0, 2]$ independently over $k = 1, \cdots, K_*$. Then we generate a large number of normal random variables except for $\boldsymbol{h}_i$, and calculate the empirical distribution function of the GMM and LIML estimators in their normalized forms. We repeat 3,000 replications for each case and the smoothing technique to estimate the empirical distribution functions. Then among a large number of simulations, we only report the results for $(N, T) = (75,\ 25)$ and $(150, 50)$.

For both cases $(a)$ and $(b)$, by using true parameters we have set the normalizing factors as the corresponding sample analogues,

$$\boldsymbol{\Phi}^{\dagger} = \frac{1}{N(T-2)}\mathbf{D}'\mathbf{J}_K' \sum_{t=2}^{T-1} \mathbf{W}_{t-1}'[\mathbf{I}_N - \mathbf{D}_t]\mathbf{W}_{t-1}\mathbf{J}_K\mathbf{D}\ , \tag{4.1}$$

$$\boldsymbol{\Psi}_1^{\dagger} = \frac{1}{N(T-2)}\mathbf{D}'\mathbf{J}_K' \sum_{t=2}^{T-1} \mathbf{W}_{t-1}'(\boldsymbol{\Lambda}_N^{(\sigma)}[\mathbf{I}_N - \mathbf{D}_t]^2)\mathbf{W}_{t-1}\mathbf{J}_K\mathbf{D}\ , \tag{4.2}$$

$$\Psi_2^{\dagger} = \frac{1}{N(T-2)} \sum_{t=2}^{T-1} \sum_{i,j=1}^{N} \boldsymbol{\omega}^{\perp\prime}[\sigma_i^2\boldsymbol{\Omega}_j + \boldsymbol{\Omega}_i\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega}_j]\boldsymbol{\omega}^{\perp}[m_{ij}^{(t)}(1-\delta_{ij})]^2\ , \tag{4.3}$$

where $\mathbf{\Lambda}_N^{(\sigma)} = \text{diag}\{\sigma_i^2\}$, $\mathbf{W}_t = (\mathbf{w}_{it})$, $\mathbf{D}_t = \mathbf{D}_t^{(a)}$ or $\mathbf{D}_t^{(b)}$, and $\boldsymbol{\omega}^{\perp\prime} = (0,1)[\mathbf{I}_2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'/\sigma^2]$. Then we investigate the empirical distribution of the normalized estimator [2]

$$\sqrt{N(T-2)}\begin{bmatrix} (\psi_{11}^{\dagger})^{-\frac{1}{2}} & 0 \\ 0 & (\psi_{22}^{\dagger})^{-\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{\gamma}_{11} - \gamma_{11} \end{bmatrix}, \tag{4.4}$$

where $\psi_{11}^{\dagger}$ and $\psi_{22}^{\dagger}$ are the (1,1)-th element and (2,2)-th element of $\mathbf{\Psi}^{\dagger} = \mathbf{\Phi}^{\dagger-1}[\mathbf{\Psi}_1^{\dagger} + (1,0,0)'\Psi_2^{\dagger}(1,0,0)]\mathbf{\Phi}^{\dagger-1}$ in the case $(a)$, $\mathbf{\Psi}^{\dagger} = \mathbf{\Phi}^{\dagger-1}\mathbf{\Psi}_1^{\dagger}\mathbf{\Phi}^{\dagger-1}$ in the case $(b)$.

Figures 1-4 show the results of the forward-filtering case $(a)$, and compare the LIML and GMM estimators by Akashi and Kunitomo (2010b) and the corresponding modified estimators. We first notice that the modification removes the biases caused by the forward filtering and heterogeneity for LIML and GMM estimators. As for the LIML estimator, an investigation of its empirical distribution as (4.4) by increasing the sample size indicates that the order of bias due to the heterogeneity is $O(NT)$. For the simulation of $\beta_2$, the modified LIML estimator seems to be slightly efficient than those of GMM estimator. But we have found that the difference of two estimators become often small. This aspect on the estimation methods confirms the findings of Anderson et al. (2008, 2010) and Kunitomo (2008) because the number of instruments is large in some sense.

Figures 5-8 show the results of the backward-filtering case $(b)$, and we see the similar results as the case of (a). In the case of (b) the LIML estimator is asymptotically median-unbiased for the homogeneous disturbances. However, in the case of heteroscedastic disturbances the non-centered scaling of the distribution of the LIML estimator has some bias so that the order of bias of (4.4) seems to be $O(1)$. The label "$(b)^*$" stands for using the different normalizing factor as $\mathbf{\Psi}^{\dagger} = \mathbf{\Phi}^{\dagger-1}[\mathbf{\Psi}_1^{\dagger} + (1,0,0)'\Psi_2^{\dagger}(1,0,0)]\mathbf{\Phi}^{\dagger-1}$ as if we had the homogeneous disturbances with a fixed-$N$. It seems that the approximation of the finite sample distribution of estimators have been improved.

# 5 Conclusions

In this paper we have introduced a class of the modified LIML estimators and investigated their asymptotic properties and finite sample properties in the dynamic panel structural equation models when we have significant individual heteroscedasticities. When we have panel dynamic structural equations, we often need to cope with the individual effects and the use of the forward-filtering panel data has been

[2] Reduction of sample size from $T$ to $T-2$ is due to using the backward filtering in the case $(b)$, and from the result the sample analogue of $c_a$ are actually less than 1/2.

one of standard methods in panel econometric analysis. However, this procedure causes some additional bias in the standard estimation methods such as the GMM estimator, which may complicate the evaluation of estimation methods. In this paper we have shown that if we apply an asymptotically optimal modification of the LIML (AOM-LIML) estimator proposed by Kunitomo (2008) to the panel structural equation, the resulting estimator does not have the asymptotic bias and also its asymptotic covariance does depend on the second order moments. More importantly, the AOM-LIML estimator attains the asymptotic bound in a class of estimators. Because it has also reasonable finite sample properties, it should be useful for many practical applications in the econometric panel data analyses. Because the AOM-LIML estimator has the simple form of the asymptotic variance, it is straightforward to develop the hypothesis testing procedure for coefficients of the structural equation in the dynamic panel models.

There are alternative ways to use the filtering methods, namely the forward-filtering and the backward filtering, to the panel structural equation given a set of panel data. Although the backward-filtering looks an attractive way of handling panel data, their asymptotic results depend on the list of instrumental variables. Often it may be reasonable to have many instrumental variables, the forward-filtering has been one way to solve the problem of handling panel data. Since the resulting estimators have different asymptotic properties and finite sample properties, the comparison of alternative methods are currently under investigation.

# 6 Mathematical Details

We shall derive the limiting distributions of the modified LIML estimators reported as Theorems 3.1 and 3.2 in several steps. Because the arguments used in our derivations are very similar to those developed by Akashi and Kunitomo (2010a, b), we shall try to give the additional arguments to derive the results.

When we use the generic notations of $(\mathbf{M}_t^D, \mathbf{M}_t, \mathbf{D}_t, \boldsymbol{\Upsilon}_n^{(.)})$, the relevant derivation is valid for the each case of $(a)$ and $(b)$, and then $\boldsymbol{\Upsilon}_n^{(.)} = (\Upsilon_{11n}^{(.)}, \Upsilon_{12n}^{(.)}, \Upsilon_{21n}^{(.)}, \Upsilon_{22n}^{(.)}, \Upsilon_{4n}^{(.)})$ will be defined in (6.46)-(6.50) below. We shall use the notation that $\mathbf{e}_i, \mathbf{e}_{k^*}$ $(i = 1, ..., N; k^* = 1, ..., K^*)$ stand for the $i$-th($k^*$-th) unit vectors and $\mathbf{J}_K', \mathbf{J}_{G_2}'$ are $K(G_2) \times K^*$ selection matrices. We shall freely use the results in *Lemma 3* and *Lemma 4* of Akashi and Kunitomo (2010b). Then for the heteroscedastic models we need to change the derivations of the asymptotic properties of estimators from those used by Akashi and Kunitomo (2010b). In the following derivations and the proofs we shall use the representation of $\mathbf{M}_t^D = \mathbf{M}_t - \mathbf{D}_t$ with $\mathbf{M}_t = \mathbf{M}_t^{(a)}$ or $\mathbf{M}_t^{(b)}$ and $\mathbf{D}_t = \mathbf{D}_t^{(a)}$ or $\mathbf{D}_t^{(b)}$ instead of $\boldsymbol{M}_{t,m}$ for the resulting mathematical convenience.

We have omitted the proof of Theorem 3.3 because it is essentially the same as Theorems reported in Anderson and Kunitomo (2010b).

First, we prepare the following lemma, which is an extension of the one of Arvarez and Arellano (2003) and Akashi and Kunitomo (2010b), and the results will be repeatedly used in our derivations.

**Lemma 1 :** Let $\boldsymbol{\epsilon}_t^{(a)} = \boldsymbol{h}_i^{(a)} \circ \boldsymbol{\epsilon}_t^{*(a)} = (h_i^{(a)} \epsilon_{it}^{*(a)})$ be an $N \times 1$ vector and $\mathbf{M}_t^D = \mathbf{M}_t^{(a)} - \mathbf{D}_t^{(a)}$, where $\epsilon_{it}^{(a)}$ $(i = 1, \cdots, N)$ are conditionally independent across $i$ with $\mathcal{E}_t[\epsilon_{it}^{(a)}] = 0$, where $\mathcal{F}_{t-1}$ is the $\sigma-$field generated the random variables given at $t-1$. Then, for $l \geq r \geq t,\ p \geq q \geq s,\ t \geq s$,

$$
\begin{aligned}
& Cov[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)}, \boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_q^{(b)}] \qquad (6.1)\\
\leq & \begin{cases} (m^{(3)} + m^{(2)})\mathrm{tr}(\mathbf{M}_t) & \text{if } l = r = p = q, \\ m^{(3)}\mathrm{tr}(\mathbf{M}_t) & \text{if } l = p \neq r = q, \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
m^{(2)} &= m^{(2)}(\boldsymbol{\epsilon}_t^{*(a)}, \boldsymbol{\epsilon}_t^{*(b)}) = (\mathcal{E}[\epsilon_{it}^{*(a)} \epsilon_{it}^{*(b)}])^2, \\
m^{(3)} &= m^{(3)}(\boldsymbol{\epsilon}_t^{*(a)}, \boldsymbol{\epsilon}_t^{*(b)}) = \mathcal{E}[\epsilon_{it}^{*(a)2}]\mathcal{E}[\epsilon_{it}^{*(b)2}] \ . \qquad (6.2)
\end{aligned}
$$

**Proof of Lemma 1 :** The unconditional covariances are represented by

$$
\begin{aligned}
& Cov[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)}, \boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_q^{(b)}] \qquad (6.3)\\
= \ & \mathcal{E}[Cov_t[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)}, \boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_q^{(b)}]] + \ Cov[\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)}], \mathcal{E}_t[\boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_q^{(b)}]] \\
= \ & \mathcal{E}[\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)} \boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_q^{(b)}]] \ .
\end{aligned}
$$

We notice that if $p < t$ $\boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s \boldsymbol{\epsilon}_q^{(b)}$ is constant and the covariances vanish. The second equality follows from $\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^{(D)} \boldsymbol{\epsilon}_r^{(b)}] = 0$ and $\mathcal{E}_t[\boldsymbol{\epsilon}_p^{(a)\prime} \mathbf{M}_s^{(D)} \boldsymbol{\epsilon}_q^{(b)}] = 0$. In fact,

$$
\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)\prime} \mathbf{M}_t^D \boldsymbol{\epsilon}_r^{(b)}] = \mathrm{tr}(\mathbf{M}_t^D \mathcal{E}_t[\boldsymbol{\epsilon}_r^{(b)} \boldsymbol{\epsilon}_l^{(a)\prime}]) \begin{cases} \mathcal{E}[\epsilon_{it}^{(a)} \epsilon_{it}^{(b)}]\mathrm{tr}(\mathbf{M}_t^D \boldsymbol{\Lambda}_N^{(a)} \boldsymbol{\Lambda}_N^{(b)}) = 0 & \text{if } l = r, \\ 0 & \text{if } l \neq r, \end{cases}
$$

$$
\mathcal{E}_t[\boldsymbol{\epsilon}_q^{(a)\prime} \mathbf{M}_s^D \boldsymbol{\epsilon}_p^{(b)}] = \mathrm{tr}(\mathbf{M}_s^D \mathcal{E}_t[\boldsymbol{\epsilon}_p^{(b)} \boldsymbol{\epsilon}_q^{(a)\prime}]) \begin{cases} \mathcal{E}[\epsilon_{it}^{(a)} \epsilon_{it}^{(b)}]\mathrm{tr}(\mathbf{M}_s^D \boldsymbol{\Lambda}_N^{(a)} \boldsymbol{\Lambda}_N^{(b)}) = 0 & \text{if } q = p, \\ 0 & \text{if } q \neq p. \end{cases}
$$

since we use the notation$\boldsymbol{\Lambda}_N^{(a)} = \mathrm{diag}(h_i^{(a)})$ and $\boldsymbol{\Lambda}_N^{(b)} = \mathrm{diag}(h_i^{(b)})$ is a diagonal matrix, and then the diagonal elements of $\mathbf{M}_t^D, \mathbf{M}_s^D$ are zeros.

As for the leading term of (6.3) we have

$$
\begin{aligned}
&\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_r^{(b)}\boldsymbol{\epsilon}_p^{(a)'}\mathbf{M}_s^D\boldsymbol{\epsilon}_q^{(b)}] \qquad (6.4)\\
= &\begin{cases}
\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)}\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_s^D\boldsymbol{\epsilon}_l^{(b)}] & \text{if } l=r=p=q,\\
\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)}\boldsymbol{\epsilon}_l^{(a)'}]\mathbf{M}_s^D\boldsymbol{\epsilon}_q^{(b)}=0 & \text{if } l=r=p\neq q<t,\\
\mathcal{E}_t[\mathcal{E}_{l\wedge p}(\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)})\mathcal{E}_{l\wedge p}(\boldsymbol{\epsilon}_p^{(a)'}\mathbf{M}_s^D\boldsymbol{\epsilon}_p^{(b)})]=0 & \text{if } l=r\neq p=q,\\
\mathrm{tr}(\mathbf{M}_t^D\mathcal{E}_t[\boldsymbol{\epsilon}_r^{(b)}\boldsymbol{\epsilon}_r^{(b)'}]\mathbf{M}_s^D\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)}\boldsymbol{\epsilon}_l^{(a)'}]) & \text{if } l=p\neq r=q,\\
0 & \text{otherwise.}
\end{cases}
\end{aligned}
$$

For the first type of non-zero terms

$$
\begin{aligned}
&\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)}\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_s^D\boldsymbol{\epsilon}_l^{(b)}] \qquad (6.5)\\
= &\sum_i\sum_j\sum_k\sum_\ell m_{ij}^{(D,t)}m_{k\ell}^{(D,s)}(h_i^{(a)}h_j^{(b)}h_k^{(a)}h_\ell^{(b)})\mathcal{E}_t[\epsilon_{il}^{(a)}\epsilon_{jl}^{(b)}\epsilon_{kl}^{(a)}\epsilon_{\ell l}^{(b)}]\\
= &\mathcal{E}[\epsilon_{it}^{(a)2}\epsilon_{it}^{(b)2}]\sum_i m_{ii}^{(D,t)}m_{ii}^{(D,s)}h_i^{(a)2}h_i^{(b)2}+(\mathcal{E}[\epsilon_{it}^{(a)}\epsilon_{it}^{(b)}])^2\sum_{i,k,k\neq i}m_{ii}^{(D,t)}m_{kk}^{(D,s)}h_i^{(a)}h_i^{(b)}h_k^{(a)}h_k^{(b)}\\
&+\mathcal{E}[\epsilon_{it}^{(a)2}]\mathcal{E}[\epsilon_{it}^{(b)2}]\sum_{i,j,j\neq i}m_{ij}^{(t)}m_{ij}^{(s)}h_i^{(a)2}h_j^{(b)2}+(\mathcal{E}[\epsilon_{it}^{(a)}\epsilon_{it}^{(b)}])^2\sum_{i,j,j\neq i}m_{ij}^{(t)}m_{ji}^{(s)}h_i^{(a)}h_i^{(b)}h_j^{(a)}h_j^{(b)}\\
= &m^{(3)}[\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)2}\mathbf{M}_t\boldsymbol{\Lambda}_N^{(b)2}\mathbf{M}_s)-\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)2}\mathbf{D}_t\boldsymbol{\Lambda}_N^{(b)2}\mathbf{D}_s)]\\
&+m^{(2)}[\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_t\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_s)-\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{D}_t\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{D}_s)]\ ,
\end{aligned}
$$

where $m_{ij}^{(D,t)}$ and $m_{k\ell}^{(D,s)}$ denotes elements of $\mathbf{M}_t^D$ and $\mathbf{M}_s^D$ respectively, and we use the fact that $m_{ii}^{(D,t)}=0$ and $m_{ij}^{(D,t)}=m_{ij}^{(t)}$ $(i\neq j)$.

Let $\mathbf{C}_t$ be an $N\times N$ orthogonal matrix such that $\boldsymbol{\Lambda}_t=\mathbf{C}_t'\mathbf{M}_t\mathbf{C}_t$ is a diagonal matrix, $\mathbf{P}_s=\mathbf{C}_t'\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_s\boldsymbol{\Lambda}_N^{(b)}\boldsymbol{\Lambda}_N^{(a)}\mathbf{C}_t(\geq\mathbf{O})$ and $\mathbf{C}_t\mathbf{C}_t'=\mathbf{I}_N$. Then

$$
\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_s\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_t)=\mathrm{tr}(\mathbf{P}_s\boldsymbol{\Lambda}_t)=\sum_{i=1}^N\lambda_i^{(t)}p_{ii}^{(s)}\leq\sum_{i=1}^N p_{ii}^{(s)}\leq\mathrm{tr}(\mathbf{M}_s) \quad (6.6)
$$

because the first inequality is due to $p_{ii}^{(s)}\geq 0,\ 0\leq\lambda_i^{(t)}\leq 1$ and the second inequality follows from that $\sum_{i=1}^N p_{ii}^{(s)}=\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)}\mathbf{M}_s\boldsymbol{\Lambda}_N^{(a)}\boldsymbol{\Lambda}_N^{(b)})$ and $0<h_i^{(a)}h_i^{(b)}\leq 1$.

For the second type

$$
\begin{aligned}
\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)}\boldsymbol{\epsilon}_l^{(a)'}]\mathbf{M}_s^D\boldsymbol{\epsilon}_q^{(b)} &= \mathcal{E}[\epsilon_{it}^{(a)2}\epsilon_{it}^{(b)}](m_{11}^{(D,t)}h_1^{(a)2}h_1^{(b)},...,m_{NN}^{(D,t)}h_N^{(a)2}h_N^{(b)})\mathbf{M}_s^D\boldsymbol{\epsilon}_q^{(b)}\\
&= 0\ , \qquad (6.7)
\end{aligned}
$$

and third type is also zero since $\mathcal{E}_{l\wedge p}(\boldsymbol{\epsilon}_l^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_l^{(b)})=0$ or $\mathcal{E}_{l\wedge p}(\boldsymbol{\epsilon}_p^{(a)'}\mathbf{M}_t^D\boldsymbol{\epsilon}_p^{(b)})=0$.

For the fourth type

$$
\mathrm{tr}(\mathbf{M}_t^D\mathcal{E}_t[\boldsymbol{\epsilon}_r^{(b)}\boldsymbol{\epsilon}_r^{(b)'}]\mathbf{M}_s^D\mathcal{E}_t[\boldsymbol{\epsilon}_l^{(a)}\boldsymbol{\epsilon}_l^{(a)'}])=m^{(3)}[\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)2}\mathbf{M}_t\boldsymbol{\Lambda}_N^{(b)2}\mathbf{M}_s)-\mathrm{tr}(\boldsymbol{\Lambda}_N^{(a)2}\mathbf{D}_t\boldsymbol{\Lambda}_N^{(b)2}\mathbf{D}_s)]\ .
$$

For the term $\mathrm{tr}(\mathbf{\Lambda}_N^{(a)2}\mathbf{M}_t\mathbf{\Lambda}_N^{(b)2}\mathbf{M}_s)$ in the first and fourth types, we have

$$
\begin{aligned}
|\mathrm{tr}(\mathbf{\Lambda}_N^{(a)2}\mathbf{M}_t\mathbf{\Lambda}_N^{(b)2}\mathbf{M}_s)| &\leq [\ \mathrm{tr}(\mathbf{\Lambda}_N^{(a)2}\mathbf{M}_t^2\mathbf{\Lambda}_N^{(a)2})\mathrm{tr}(\mathbf{M}_s\mathbf{\Lambda}_N^{(b)4}\mathbf{M}_s)\ ]^{\frac{1}{2}} \\
&= [\ \mathrm{tr}(\mathbf{\Lambda}_N^{(a)4}\mathbf{M}_t)\mathrm{tr}(\mathbf{\Lambda}_N^{(b)4}\mathbf{M}_s)\ ]^{\frac{1}{2}} \\
&\leq \mathrm{tr}(\mathbf{M}_t)
\end{aligned}
\tag{6.8}
$$

because each elements of $h_i^{(a)}, h_i^{(b)}$ are bounded and $s \leq t$. **Q.E.D.**

**Derivations of Theorem 3.1 and Theorem 3.2** : We shall show the derivations of Theorem 3.1 and Theorem 3.2 in several steps.

(**Step 1**) : At this step, we shall show the convergence of $\mathbf{G}^{(f)}$ and $\mathbf{H}^{(f)}$ based on $\mathbf{M}_t^D = \mathbf{M}_t - \mathbf{D}_t$. The many parts of the derivations are some modifications of the corresponding ones in Akashi and Kunitomo (2010b). We shall show that

$$\frac{1}{NT}\mathbf{G}^{(f)} \xrightarrow{p} \mathbf{G}_0 = \mathbf{B}'\mathbf{\Phi}^*\mathbf{B}\ , \tag{6.9}$$

$$\frac{1}{NT}\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0 = \begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} + \mathbf{B}'\mathbf{\Phi}_2^*\mathbf{B}\ , \tag{6.10}$$

where $\mathbf{B} = (\boldsymbol{\theta}, \mathbf{I}_{G_2+K_1})$, $\mathbf{\Phi}^* = \mathbf{\Phi}_1^* - \mathbf{\Phi}_2^*$ and

$$\mathbf{\Phi}_1^* = \mathbf{D}'\mathbf{J}' \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\mathcal{E}[(1-m_{ii}^{t,a})\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}']\mathbf{J}\mathbf{D} \tag{6.11}$$

$$\mathbf{\Omega} = \mathbf{J}_{1+G_2}'(\ \mathbf{\Omega}_v^{**} \circ \mathrm{plim}_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}[\boldsymbol{h}_i\boldsymbol{h}_i']\ )\mathbf{J}_{1+G_2} > 0\ , \tag{6.12}$$

$$\mathbf{\Phi}_2^* = \mathbf{D}'\mathbf{J}_K'\Big(\lim_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{W}_{t-1}'\mathbf{D}_t\mathbf{W}_{t-1}]\Big)\mathbf{J}_K\mathbf{D} \geq 0\ . \tag{6.13}$$

Consider the decomposition

$$\mathbf{G}^{(f)} = \mathbf{G}^{(f,1)} + \mathbf{G}^{(f,2)} + \mathbf{G}^{(f,2)'} + \mathbf{G}^{(f,3)}\ , \tag{6.14}$$

where

$$\mathbf{G}^{(f,1)} = \mathbf{B}'\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{Z}_{t-1}^{(f)}\mathbf{D}\mathbf{B}\ , \tag{6.15}$$

$$\mathbf{G}^{(f,2)} = \mathbf{B}'\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D(\mathbf{V}_t^{(f)}, \mathbf{O})\ , \tag{6.16}$$

$$\mathbf{G}^{(f,3)} = \sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)'} \\ \mathbf{O}\end{pmatrix}\mathbf{M}_t^D(\mathbf{V}_t^{(f)}, \mathbf{O})\ , \tag{6.17}$$

and $\mathbf{V}_t^{(f)'} = (\mathbf{v}_{1t}^{(f)}, \cdots, \mathbf{v}_{Nt}^{(f)})$, $\mathbf{v}_{it}^{(f)}$ are the corresponding forward-filtered disturbances of $\mathbf{v}_{it}$.

First, we notice that $(1/NT)\mathbf{G}^{(f,2)} \xrightarrow{p} \mathbf{O}$ because $(1/\sqrt{NT})\mathbf{G}^{(f,2)} = O_p(1)$ by using the argument in Step 2 below.

Second, we evaluate $\mathbf{G}^{(f,3)}$. For $g, h = 1, \cdots, (1+G_2)$, by the construction of $\mathbf{M}_t^D$, we have

$$\mathcal{E}[\mathbf{e}_g'\mathbf{J}_{1+G_2}'\mathbf{G}^{(f,3)}\mathbf{J}_{1+G_2}\mathbf{e}_h] = \sum_{t=1}^{T-1}\mathcal{E}[\mathrm{tr}(\mathbf{M}_t^D\mathcal{E}_t[\mathbf{v}_t^{(f,g)}\mathbf{v}_t^{(f,h)'}])] = 0 \, . \tag{6.18}$$

By using $\mathbf{V}_t^{(f)} = (\mathbf{V}_t - \bar{\mathbf{V}})/c_t$, we have

$$\begin{aligned} & \frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{e}_g'\mathbf{J}_{1+G_2}'\mathbf{G}^{(f,3)}\mathbf{J}_{1+G_2}\mathbf{e}_h \qquad (6.19)\\ = & \frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}\mathbf{v}_t^{(g)'}\mathbf{M}_t^D\mathbf{v}_t^{(h)'} - \frac{2}{NT}\sum_{t=1}^{T-1}c_t^{-2}\mathbf{v}_t^{(g)'}\mathbf{M}_t^D\bar{\mathbf{v}}_{tT}^{(h)'} + \frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}\bar{\mathbf{v}}_{tT}^{(g)'}\mathbf{M}_t^D\bar{\mathbf{v}}_{tT}^{(h)'} \, . \end{aligned}$$

By using *Lemma 1* and the fact $(c_t^{-2})^2 \leq 2$, the variance of the first term in (6.19) is given by

$$\mathcal{V}ar[\frac{1}{NT}\sum_{t=1}^{T-1}c_t^{-2}\mathbf{v}_t^{(g)'}\mathbf{M}_t^D\mathbf{v}_t^{(h)'}] \leq \frac{2}{(NT)^2}\sum_{t=1}^{T-1}[\mathbf{e}_g'\mathbf{\Omega}^{**}\mathbf{e}_g\mathbf{e}_h'\mathbf{\Omega}^{**}\mathbf{e}_h + (\mathbf{e}_g'\mathbf{\Omega}^{**}\mathbf{e}_h)^2]\mathrm{tr}(\mathbf{M}_t) \, .$$

The second and third terms of (6.19) can be evaluated analogously as $\Upsilon_{21}^{(k)}$ and $\Upsilon_{22}^{(k)}$ in Step 3 below. In fact, for the case (a), the order of variances of the second and third terms are $O(\log T/N^2T)$, $O((\log T)^2/N^2T)$, respectively. For the case of $(b)$, they are $O(\log T/(NT)^2)$, $O((\log T)^2/(NT)^2)$. Thus, we have $(1/NT)\mathbf{G}^{(f,3)} \xrightarrow{p} \mathbf{O}$.

Next, we use the representation

$$\begin{aligned} \mathbf{Z}_{t-1}^{(f)'} &= \mathbf{J}_K'\Big(c_t[\mathbf{I}_{K^*} - \frac{1}{T-t}(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})]\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}'\Big) \\ &= \mathbf{\Psi}_t'\mathbf{W}_{t-1}' - c_t\tilde{\mathbf{V}}_{tT}' \text{ (, } say\text{)}. \end{aligned} \tag{6.20}$$

Then we shall investigate the asymptotic behavior of $\mathbf{G}^{(f,1)}$ and we further decompose $(1/NT)\mathbf{G}^{(f,1)}$ as

$$\begin{aligned} \frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{Z}_{t-1}^{(f)} &= \frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{\Psi}_t'\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1}\mathbf{\Psi}_t - \frac{1}{NT}\sum_{t=1}^{T-1}c_t\mathbf{\Psi}_t'\mathbf{W}_{t-1}'\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT} \\ &- \frac{1}{NT}\sum_{t=1}^{T-1}c_t\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^D\mathbf{W}_{t-1}\mathbf{\Psi}_t + \frac{1}{NT}\sum_{t=1}^{T-1}c_t^2\tilde{\mathbf{V}}_{tT}'\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT}. \end{aligned} \tag{6.21}$$

The expected value of the fourth term of (6.21) is $\mathbf{O}$. By using the similar arguments for $\Upsilon_{22}^{(k)}$, the order of variance of the fourth term is $O((\log T)^2/N^2T)$ for the case $(a)$, and it is $O((\log T)^2/(NT)^2)$ for the case $(b)$.

The second and third terms of (6.21) have zero means, and for $j,k=1,\cdots,K$ we use the Cauchy-Schwarz inequality

$$
\begin{aligned}
& Var[\frac{1}{NT}\sum_{t=1}^{T-1} c_t \mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT}\mathbf{e}_k] \qquad (6.22)\\
\leq\ & \frac{1}{(NT)^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sqrt{c_t^2\mathcal{E}[(\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT}\mathbf{e}_k)^2]}\sqrt{c_s^2\mathcal{E}[(\mathbf{e}_k^{\prime}\tilde{\mathbf{V}}_{sT}\mathbf{M}_s^D\mathbf{W}_{s-1}\mathbf{\Psi}_s\mathbf{e}_j)^2]}\ .
\end{aligned}
$$

Also we have

$$
\begin{aligned}
& c_t^2\mathcal{E}[(\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT}\mathbf{e}_k)^2]\\
=\ & c_t^2\mathcal{E}\Big[\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}\mathbf{M}_t^D\Big[\frac{1}{(T-t)^2}\sum_{h=1}^{T-t}\mathbf{e}_{kJ}^{'}\mathbf{\Phi}_h^{\prime}\mathcal{E}_t[\mathbf{v}_{i0}^{*}\mathbf{v}_{i0}^{*'}]\mathbf{\Phi}_h\mathbf{e}_{kJ}\Big]\mathbf{M}_t^{D'}\mathbf{W}_{t-1}\mathbf{\Psi}_t\mathbf{e}_j\Big]\\
=\ & \sum_{l,\ell=1}^{K^*}\Big(\Big[\frac{\mathbf{e}_l^{\prime}\mathbf{\Omega}^{**}\mathbf{e}_\ell}{(T-t)^2}\sum_{h=1}^{T-t}\mathbf{e}_{kJ}^{'}\mathbf{\Phi}_h^{\prime}\mathbf{e}_l\mathbf{e}_\ell^{\prime}\mathbf{\Phi}_h\mathbf{e}_{kJ}\Big]\\
& \times c_t^2\mathcal{E}\Big[(\mathbf{e}_l^{\prime}\boldsymbol{\pi}_i\boldsymbol{\pi}_i^{\prime}\mathbf{e}_\ell)\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}(\mathbf{M}_t-\mathbf{D}_t)^2\mathbf{W}_{t-1}\mathbf{\Psi}_t\mathbf{e}_j\Big]\Big)\\
\leq\ & \sum_{l,\ell=1}^{K^*}\Big[\frac{|\mathbf{e}_l^{\prime}\mathbf{\Omega}^{**}\mathbf{e}_\ell|}{(T-t)^2}\sum_{h=1}^{T-t}|\mathbf{e}_{kJ}^{'}\mathbf{\Phi}_h^{\prime}\mathbf{e}_l||\mathbf{e}_\ell^{\prime}\mathbf{\Phi}_h\mathbf{e}_{kJ}|\Big]\mathcal{E}\Big[\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathbf{W}_{t-1}^{'}\mathbf{W}_{t-1}\mathbf{\Psi}_t\mathbf{e}_j\Big]\\
=\ & O\Big(\frac{K^{*2}N}{T-t}\Big)\mathbf{e}_j^{'}\mathbf{\Psi}_t^{\prime}\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}^{\prime}]\mathbf{\Psi}_t\mathbf{e}_j\ , \qquad (6.23)
\end{aligned}
$$

where $\mathbf{e}_{kJ}$ stands for $\mathbf{e}_{kJ}=\mathbf{J}_K\mathbf{e}_k$.
The inequality in (6.23) follows from the relations that $c_t^2<1,\ |\mathbf{e}_l^{\prime}\mathbf{h}_i\mathbf{h}_i^{\prime}\mathbf{e}_\ell|\leq 1$ and $\lambda_{\max}\{(\mathbf{M}_t-\mathbf{D}_t)^2\}\leq 1$, which is the result of $\mathbf{M}_t^D=(\mathbf{M}_t^D)^{\prime},\ \lambda_{\max}\{(\mathbf{M}_t-\mathbf{D}_t)^2\}=\max_i|\lambda_i\{\mathbf{M}_t-\mathbf{D}_t\}|^2$ and

$$
-1\leq\lambda_{\min}\{\mathbf{M}_t\}+\lambda_{\min}\{-\mathbf{D}_t\}\leq\lambda_i\{\mathbf{M}_t-\mathbf{D}_t\}\leq\lambda_{\max}\{\mathbf{M}_t\}+\lambda_{\max}\{-\mathbf{D}_t\}\leq 1\ .
$$

Hence

$$
\begin{aligned}
Var[\frac{1}{NT}\sum_{t=1}^{T-1}c_t\mathbf{e}_j^{'}\mathbf{\Psi}_t\mathbf{W}_{t-1}^{'}\mathbf{M}_t^D\tilde{\mathbf{V}}_{tT}\mathbf{e}_k] &\leq \frac{1}{(NT)^2}\sum_{t=1}^{T-1}\sqrt{O(\frac{N}{T-t})}\sum_{s=1}^{T-1}\sqrt{O(\frac{N}{T-s})}\\
&= O\Big(\frac{(\sqrt{T})^2}{NT^2}\Big)\ . \qquad (6.24)
\end{aligned}
$$

For the first term of (6.21), we further decompose it as

$$
\begin{aligned}
& \frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{\Psi}_t'\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1}\mathbf{\Psi}_t \\
= & \frac{1}{NT}\sum_{t=1}^{T-1}\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1} - \frac{1}{NT}\sum_{t=1}^{T-1}(\frac{1}{T-t+1})\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1} \qquad (6.25) \\
& -\frac{1}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{T-t}\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1}(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})' - \frac{1}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{T-t}(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1} \\
& +\frac{1}{NT}\sum_{t=1}^{T-1}(\frac{c_t}{T-t})^2(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1}(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})' \\
\xrightarrow{p} & \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathcal{E}[\mathbf{w}_{i(t-1)}\mathbf{w}_{i(t-1)}'] - \lim_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{W}_{t-1}'\mathbf{D}_t\mathbf{W}_{t-1}]\ ,
\end{aligned}
$$

where we have used the fact $c_t^2 = 1 - 1/(T-t+1)$.
For the second term of (6.25), for $j,k = 1,\cdots,K$, we have

$$
\begin{aligned}
& \mathcal{E}\Big[\ \Big|\frac{1}{NT}\sum_{t=1}^{T-1}(\frac{1}{T-t+1})\mathbf{e}_{jJ}'\mathbf{W}_{t-1}'\mathbf{M}_t^D\mathbf{W}_{t-1}\mathbf{e}_{kJ}\Big|\ \Big] \qquad (6.26) \\
\leq & \frac{1}{NT}\sum_{t=1}^{T-1}(\frac{1}{T-t+1})\mathcal{E}[|\mathbf{w}_{t-1}^{(j)\prime}(\mathbf{M}_t-\mathbf{D}_t)^2\mathbf{w}_{t-1}^{(j)}\mathbf{w}_{t-1}^{(k)\prime}\mathbf{w}_{t-1}^{(k)}|^{\frac{1}{2}}] \qquad (6.27) \\
= & \frac{O(N\log T)}{NT}\ ,
\end{aligned}
$$

where the inequality is due to the Cauchy-Schwartz inequality, and we have used it as $\mathcal{E}[|\mathbf{w}_{t-1}^{(j)\prime}\mathbf{w}_{t-1}^{(j)}\mathbf{w}_{t-1}^{(k)\prime}\mathbf{w}_{t-1}^{(k)}|^{\frac{1}{2}}] \leq \mathcal{E}[(\mathbf{w}_{t-1}^{(j)\prime}\mathbf{w}_{t-1}^{(j)})(\mathbf{w}_{t-1}^{(k)\prime}\mathbf{w}_{t-1}^{(k)})]^{\frac{1}{2}} = O(N)$.
Therefore the second term of (6.25) converges in probability to 0. By using the similar arguments and the boundedness of $|\mathbf{e}_j(\sum_{j=1}^{T-t}\mathbf{\Pi}^{*j})\mathbf{e}_{kJ}|$, the third and fifth terms of (6.25), which are of the orders of $O(\log T/T)$ and $O(1/T)$ respectively, converge to $\mathbf{O}$. The first term of (6.25) converges as $N,T\to\infty$ by using *Lemma 3* and *Lemma 4* of Akashi and Kunitomo (2010b) and the assumption (A-III). Hence, we have shown that $(1/NT)\mathbf{G}^{(f)} \xrightarrow{p} \mathbf{G}_0$.

Next, we turn to show that $(1/NT)\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{H}_0$ by evaluating each terms of the decomposition

$$
\begin{aligned}
& \frac{1}{NT}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{Y}_t^{(f)'}\\ \mathbf{Z}_{t-1}^{(1,f)'}\end{pmatrix}(\mathbf{Y}_t^{(f)},\mathbf{Z}_{t-1}^{(1,f)}) \\
= & \frac{1}{NT}\mathbf{B}'\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{Z}_{t-1}^{(f)}\mathbf{DB}+\frac{1}{NT}\mathbf{B}'\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}(\mathbf{V}_t^{(f)},\mathbf{O}) \\
& +\frac{1}{NT}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)'}\\ \mathbf{O}\end{pmatrix}\mathbf{Z}_{t-1}^{(f)}\mathbf{DB}+\frac{1}{NT}\sum_{t=1}^{T-1}\begin{pmatrix}\mathbf{V}_t^{(f)'}\\ \mathbf{O}\end{pmatrix}(\mathbf{V}_t^{(f)},\mathbf{O})\ . \qquad (6.28)
\end{aligned}
$$

For the fourth term, by using the property of the forward filter

$$
\begin{aligned}
\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{e}_g'\mathbf{V}_t^{(f)'}\mathbf{V}_t^{(f)}\mathbf{e}_h] &= \frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}[\mathbf{v}_i^{(g)'}\mathbf{Q}_T\mathbf{v}_i^{(h)}] \\
&\xrightarrow{p} \boldsymbol{\Omega}-\frac{1}{NT}\lim_{N\to\infty}\sum_{i=1}^{N}\mathcal{E}[h_i^{(g)}h_i^{(h)}\mathbf{e}_g'\boldsymbol{\Omega}^{**}\mathbf{e}_h]\ , \qquad (6.29)
\end{aligned}
$$

where $\mathbf{Q}_T=\mathbf{I}_T-\boldsymbol{\iota}_T\boldsymbol{\iota}_T'/T$ and the second equality follows from the fact that $\mathcal{E}[(h_i^{(g)}v_{is}^{**(g)})(h_t^{(h)}v_{it}^{**(h)})]=0$ if $s\neq t$. Using the mutual independence of $\mathbf{v}_{it}$ over $i=1,\cdots,N$, the variance of the first term of (6.28) is given by

$$
\begin{aligned}
\mathcal{V}ar[\frac{1}{NT}\sum_{i=1}^{N}\mathbf{v}_i^{(g)'}\mathbf{v}_i^{(h)}] &= \frac{1}{(NT)^2}\sum_{i=1}^{N}\mathcal{V}ar[\sum_{t=1}^{T}(h_i^{(g)}v_{it}^{**(g)})(h_i^{(h)}v_{it}^{**(h)})] \qquad (6.30)\\
&= \frac{N}{(NT)^2}\times O\Big(T(T-1)\mathcal{V}ar[h_i^{(g)}h_i^{(h)}](\mathbf{e}_g'\boldsymbol{\Omega}^{**}\mathbf{e}_h)^2\Big)\ ,
\end{aligned}
$$

where $\mathcal{C}ov[h_i^{(g)}v_{is}^{**(g)}h_i^{(h)}v_{is}^{**(h)},h_i^{(g)}v_{it}^{**(g)}h_i^{(h)}v_{it}^{**(h)}]=\mathcal{V}ar[h_i^{(g)}h_i^{(h)}](\mathbf{e}_g'\boldsymbol{\Omega}^{**}\mathbf{e}_h)^2$ for $s\neq t$. Then the variance of the second term of (6.29) is given by

$$
\mathcal{V}ar[\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{v}_i^{(g)'}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{v}_i^{(h)}] = \frac{N}{(NT^2)^2}\mathcal{V}ar[\mathbf{v}_i^{(g)'}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{v}_i^{(h)}]=\frac{O(NT^4)}{(NT^2)^2},\qquad (6.31)
$$

since $\mathbf{v}_i^{(g)'}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{v}_i^{(h)}=O_p(T^2)$. Hence we find that the variance order of the last term of (6.28) is $O(1/N)$.

Turning to consider the second term of (6.28), we have

$$
\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{Z}_{t-1}^{(f)'}\mathbf{V}_t^{(f)}]=\mathbf{O}-\frac{N}{NT^2}\times O\Big((T-1)\mathbf{J}_K'(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^*)^{-1}\mathcal{E}[\mathbf{v}_{i0}^*\mathbf{v}_{i0}^{*'}\mathbf{J}_{1+G_2}]\Big)\ .
$$

By the similar arguments as used for evaluating (6.31), the variance order of the second term of (6.28) is $O(1/N)$.

Finally, we consider the first term of (6.28),

$$\frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{e}_{jJ}'\mathbf{Z}_{t-1}^{(f)'}\mathbf{Z}_{t-1}^{(f)}\mathbf{e}_{kJ}] \tag{6.32}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}[w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)}] - \frac{N}{NT^2}\mathbf{e}_{jJ}'[\sum_{l=0}^{T-1}\sum_{h=0}^{l}\mathbf{\Gamma}_{ih} + \sum_{l=1}^{T-1}\sum_{h=1}^{l}\mathbf{\Gamma}_{ih}']\mathbf{e}_{kJ}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}[w_{i(t-1)}^{(j)}w_{i(t-1)}^{(k)}] - O(\frac{NT}{NT^2}) , \tag{6.33}$$

where $\mathbf{\Gamma}_{ih} = \mathcal{E}[\mathbf{w}_{it}\mathbf{w}_{i(t+h)}']$.
By using the similar arguments as for evaluating (6.31), the variance order of the first term of (6.28) is also $O(1/N)$.
Therefore we have established that as $N, T \to \infty$

$$\frac{1}{NT}\mathbf{H}^{(f)} \xrightarrow{p} \mathbf{B}'\mathbf{\Phi}_1^*\mathbf{B} + \begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} - \mathbf{G}_0 = \mathbf{H}_0 . \tag{6.34}$$

(**Step 2**) : We consider the consistency and the limiting distribution of the AOL-LIML estimator $\hat{\boldsymbol{\theta}}_{MLI}$. For this purpose, we need to evaluate the sampling error form for $\hat{\boldsymbol{\theta}}_{MLI}$. We notice that the continuity of the minimum eigenvalue function

$$\left|\mathbf{B}'\mathbf{\Phi}^*\mathbf{B} - \text{plim}_{n\to\infty}\lambda_n\Big(\begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} + \mathbf{B}'\mathbf{\Phi}_2^*\mathbf{B}\Big)\right| = 0 ,$$

and we have 0 as a solution due to the singularity of $\mathbf{G}_0$. If $\text{plim}_{n\to\infty}\lambda_n < 0$, we observe $|\mathbf{G}_0 - \text{plim}_{n\to\infty}\lambda_n\mathbf{H}_0| > 0$. This is because for any $(1 + G_2 + K_1)$ non-zero vector $\mathbf{d}' = (\mathbf{d}_1', \mathbf{d}_2')$ we have $\mathbf{d}_1'\mathbf{\Omega}\mathbf{d}_1 > 0$, or for $\mathbf{d}_1 = \mathbf{0}$ and $\mathbf{d}_2 \neq \mathbf{0}$,

$$(\mathbf{0}', \mathbf{d}_2')\mathbf{B}'\mathbf{\Phi}^*\mathbf{B}\begin{pmatrix} \mathbf{0} \\ \mathbf{d}_2 \end{pmatrix} = (\mathbf{0}', \mathbf{d}_2')\mathbf{\Phi}^*\begin{pmatrix} \mathbf{0} \\ \mathbf{d}_2 \end{pmatrix} > 0 . \tag{6.35}$$

Thus we have $\text{plim}_{n\to\infty}\lambda_n = 0$. Then the non-singularity of (6.35) and $\mathbf{\Phi}^*(\hat{\boldsymbol{\theta}}_{MLI} - \boldsymbol{\theta}) = \mathbf{0} + o_p(1)$, we obtain $\hat{\boldsymbol{\theta}}_{MLI} \xrightarrow{p} \boldsymbol{\theta}$.

Next, we consider the limiting distribution form of the modified LIML estimator. Define $\mathbf{G}_1^{(f)} = \sqrt{n}[(1/n)\mathbf{G}^{(f)} - \mathbf{G}_0]$, $\mathbf{H}_1^{(f)} = \sqrt{n}[(1/n)\mathbf{H}^{(f)} - \mathbf{H}_0]$, $\mathbf{b}_1 = \sqrt{n}[\hat{\boldsymbol{\theta}}_{MLI} - \boldsymbol{\theta}]$ and $\lambda_{1n}^{(f)} = \sqrt{n}[\lambda_n - 0]$. By substituting these random variables into (2.19), it is asymptotically equivalent to

$$\mathbf{G}_0\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + \frac{1}{\sqrt{n}}[\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0]\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + \frac{1}{\sqrt{n}}\mathbf{G}_0\mathbf{b}_1 = o_p(\frac{1}{\sqrt{n}}) . \tag{6.36}$$

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

$$(\mathbf{B}'\mathbf{\Phi}^*\mathbf{B})\mathbf{b}_1 = [\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0] \begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + o_p(1) \ . \tag{6.37}$$

Multiplication of (6.37) from the left by $(1,-\boldsymbol{\theta})$ yields

$$\lambda_{1n}^{(f)} = \frac{(1,-\boldsymbol{\theta}')\mathbf{G}_1^{(f)}(1,-\boldsymbol{\theta}')'}{(1,-\boldsymbol{\theta}')\mathbf{H}_0(1,-\boldsymbol{\theta}')'} + o_p(1) \ . \tag{6.38}$$

Also the multiplication of (6.37) from the left by $(\mathbf{0}, \mathbf{I}_{G_2+K_1})$ and substitution of $\lambda_{1n}^{(f)}$ for (6.37) yields

$$\mathbf{\Phi}^*\sqrt{n}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{2MLI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1MLI} - \boldsymbol{\gamma}_1 \end{bmatrix} = [\mathbf{0}, \mathbf{I}_{G_2+K_1}] \left[\mathbf{G}_1^{(f)} - \lambda_{1n}^{(f)}\mathbf{H}_0\right] \begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} + o_p(1) \tag{6.39}$$

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

Using the relations of (6.38), we have

$$\mathbf{G}_1^{(f)}\begin{bmatrix} 1 \\ -\boldsymbol{\theta} \end{bmatrix} = \frac{1}{\sqrt{n}}\mathbf{B}'\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{pmatrix}\mathbf{M}_t^D\mathbf{u}_t^{(f)} \ , \tag{6.40}$$

where $\mathbf{M}_t^D = \mathbf{M}_t - \mathbf{D}_t$. Therefore

$$\mathbf{\Phi}^*\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{2MLI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1MLI} - \boldsymbol{\gamma}_1 \end{pmatrix} = \frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{U}_t^{(\perp,f)'} \\ \mathbf{O} \end{pmatrix}\mathbf{M}_t^D\mathbf{u}_t^{(f)} + o_p(1) \ , \tag{6.41}$$

where

$$\mathbf{U}_t^{(\perp,f)'} = [\mathbf{0}, \mathbf{I}_{G_2}]\left[\mathbf{I}_{1+G_2} - \frac{\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}}\right]\mathbf{V}_t^{(f)'} = (\mathbf{u}_{1t}^{(\perp,f)}, ..., \mathbf{u}_{Nt}^{(\perp,f)}) \ . \tag{6.42}$$

(**Step 3**) : At this step we shall evaluate the effects of the forward-filtering on the limiting distribution of the modified LIML estimator. The sampling error $\mathbf{\Phi}^*\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ can be written as

$$\frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{Z}_{t-1}^{(f)'}\mathbf{M}_t^D\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{U}_t^{(\perp,f)'} \\ \mathbf{O} \end{pmatrix}\mathbf{M}_t^D\mathbf{u}_t^{(f)} + o_p(1) \tag{6.43}$$

$$= \frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'[\mathbf{M}_t - \mathbf{D}_t]\mathbf{u}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{U}_t^{\perp'} \\ \mathbf{O} \end{pmatrix}[\mathbf{M}_t - \mathbf{D}_t]\mathbf{u}_t + \mathbf{b}^{(f)} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\mathbf{D}'\sum_{t=1}^{T-1}\mathbf{J}'\mathbf{W}_{t-1}'[\mathbf{I}_N - \mathbf{D}_t]\mathbf{u}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\begin{pmatrix} \mathbf{U}_t^{\perp'} \\ \mathbf{O} \end{pmatrix}[\mathbf{M}_t - \mathbf{D}_t]\mathbf{u}_t + o_p(1)$$

$$= \mathbf{a}_{1n} + \mathbf{a}_{2n} + o_p(1), \ (, say) \ .$$

We shall show the first and second equalities and that the bias $\mathbf{b}^{(f)}$ due to the forward filtering is $\mathbf{0}$. In order to show the first equality,] we use the fact $\mathbf{u}_t^{(f)} = (\mathbf{u}_t - \mathbf{u}_{tT})/c_t$ and consider the decompositions, for $k = 1, ..., K$ ; $g = 1, ..., G_2$,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \mathbf{Z}_{t-1}^{(f)'} \mathbf{M}_t^D \mathbf{u}_t^{(f)} \\ = & \ ( \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \mathbf{W}_{t-1}' \mathbf{M}_t^D \mathbf{u}_t - \Upsilon_{11n}^{(k,.)} - \Upsilon_{12n}^{(k,.)} ) - ( \Upsilon_{21n}^{(k,.)} - \Upsilon_{22n}^{(k,.)} ) , \quad (6.44) \\ & \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{(\perp,f)'} \mathbf{M}_t^D \mathbf{u}_t^{(f)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_g' \mathbf{U}_t^{\perp'} \mathbf{M}_t^D \mathbf{u}_t + \Upsilon_{4n}^{(g,.)} , \quad (6.45) \end{aligned}$$

where

$$\Upsilon_{11n}^{(k,.)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \mathbf{W}_{t-1}' \mathbf{M}_t^D \bar{\mathbf{u}}_{tT} , \quad (6.46)$$

$$\Upsilon_{12n}^{(k,.)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} (\frac{c_t}{T-t}) \mathbf{e}_{kJ}' \tilde{\mathbf{W}}_{t-1}' \mathbf{M}_t^D \mathbf{u}_t^{(f)} , \quad (6.47)$$

$$\Upsilon_{21n}^{(k,.)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \tilde{\mathbf{V}}_{tT}' \mathbf{M}_t^D \mathbf{u}_t , \quad (6.48)$$

$$\Upsilon_{22n}^{(k,.)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \mathbf{e}_{kJ}' \tilde{\mathbf{V}}_{tT}' \mathbf{M}_t^D \bar{\mathbf{u}}_{tT} , \quad (6.49)$$

$$\begin{aligned} \Upsilon_{4n}^{(g,.)} = & \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \Big[ (\frac{1}{T-t}) \mathbf{e}_g' \mathbf{U}_t^{\perp'} \mathbf{M}_t^D \mathbf{u}_t - c_t^{-2} \mathbf{e}_g' \mathbf{U}_t^{\perp'} \mathbf{M}_t^D \bar{\mathbf{u}}_{tT} \\ & - c_t^{-2} \mathbf{e}_g' \bar{\mathbf{U}}_{tT}^{\perp'} \mathbf{M}_t^D \mathbf{u}_t + c_t^{-2} \mathbf{e}_g' \bar{\mathbf{U}}_{tT}^{\perp'} \mathbf{M}_t^D \bar{\mathbf{u}}_{tT} \Big] , \quad (6.50) \end{aligned}$$

and

$$\begin{aligned} \bar{\mathbf{u}}_{tT} &= (\mathbf{u}_t + \cdots + \mathbf{u}_T)/(T-t+1) , \ \ \mathbf{u}_t = (u_{1t}, \cdots, u_{Nt})' , & (6.51) \\ \tilde{\mathbf{W}}_{t-1}' &= (\sum_{h=1}^{T-t} \mathbf{\Pi}^{*h}) \mathbf{W}_{t-1}' , \ \ \tilde{\mathbf{V}}_{tT}' = \frac{1}{T-t} \sum_{h=1}^{T-t} \mathbf{\Phi}_h \mathbf{V}_{T-h}^{*'} , & (6.52) \\ \mathbf{V}_h^{*'} &= (\mathbf{v}_{1h}^*, ..., \mathbf{v}_{Nh}^*) = (\mathbf{v}_h^{*(1)}, \cdots, \mathbf{v}_h^{*(K^*)})' , & (6.53) \\ \mathbf{\Phi}_h &= (\mathbf{I}_{K^*} - \mathbf{\Pi}^*)^{-1} (\mathbf{I}_{K^*} - \mathbf{\Pi}^{*h}) , & (6.54) \\ \mathbf{U}_t^{\perp'} &= [\mathbf{0}, \mathbf{I}_{G_2}] \left[ \mathbf{I}_{1+G_2} - \frac{\mathbf{\Omega} \boldsymbol{\beta} \boldsymbol{\beta}'}{\boldsymbol{\beta}' \mathbf{\Omega} \boldsymbol{\beta}} \right] \mathbf{V}_t' = (\mathbf{u}_{1t}^{\perp}, ..., \mathbf{u}_{Nt}^{\perp}) , & (6.55) \\ \bar{\mathbf{U}}_t^{\perp} &= (\mathbf{U}_t^{\perp} + \cdots + \mathbf{U}_T^{\perp})/(T-t+1) . & (6.56) \end{aligned}$$

We shall investigate the convergences from (6.46) to (6.50). For $\Upsilon_{11n}^{(k)}$, we have

$$
\begin{aligned}
&\mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}\mathbf{M}_t^D\mathcal{E}_t[\bar{\mathbf{u}}_{tT}\bar{\mathbf{u}}_{sT}^{\prime}]\mathbf{M}_s^{D\prime}\mathbf{w}_{s-1}^{(k)}] \qquad (6.57)\\
&= \frac{1}{(T-s+1)}\mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}(\mathbf{M}_t\mathbf{\Lambda}_N^{(\sigma)}\mathbf{M}_s - \mathbf{M}_t\mathbf{\Lambda}_N^{(\sigma)}\mathbf{D}_s - \mathbf{D}_t\mathbf{\Lambda}_N^{(\sigma)}\mathbf{M}_s + \mathbf{D}_t\mathbf{\Lambda}_N^{(\sigma)}\mathbf{D}_s)\mathbf{w}_{s-1}^{(k)}]\ .
\end{aligned}
$$

where $\mathbf{\Lambda}_N^{(\sigma)} = \mathrm{diag}\{\sigma_i^2\}$. Consider the first term of (6.57),

$$
\begin{aligned}
&\frac{1}{(T-s+1)}\mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}\mathbf{M}_t\mathbf{\Lambda}_N^{(\sigma)}\mathbf{M}_s\mathbf{w}_{s-1}^{(k)}] \qquad (6.58)\\
= &\frac{1}{(T-s+1)}\Big[\mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}\mathbf{\Lambda}_N^{(\sigma)}(\mathbf{I}_N-\mathbf{M}_s)\boldsymbol{\epsilon}_{s-1}^{(k)}] - \mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}\mathbf{\Lambda}_N^{(\sigma)}\mathbf{w}_{s-1}^{(k)}]\\
&-\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k)\prime}(\mathbf{I}_N-\mathbf{M}_t)\mathbf{\Lambda}_N^{(\sigma)}(\mathbf{I}_N-\mathbf{M}_s)\boldsymbol{\epsilon}_{s-1}^{(k)}] + \mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k)\prime}(\mathbf{I}_N-\mathbf{M}_t)\mathbf{\Lambda}_N^{(\sigma)}\mathbf{w}_{s-1}^{(k)}]\Big]\ ,
\end{aligned}
$$

where we used the decomposition $\mathbf{w}_{h-1}^{(k)\prime}\mathbf{M}_h = \mathbf{w}_{h-1}^{(k)\prime} - \boldsymbol{\epsilon}_h^{(k)\prime}[\mathbf{I}_N - \mathbf{M}_h]$ for $h = t, s$, and $\boldsymbol{\epsilon}_h^{(k)}$ was defined in Lemma 3 and Lemma 4 of Akashi and Kunitomo (2010b) for the cases of $(a)$,$(b)$. For the second term of (6.57), we use

$$
\begin{aligned}
\mathcal{E}[\mathbf{w}_{t-1}^{(k)\prime}\mathbf{\Lambda}_N^{(\sigma)}\mathbf{w}_{s-1}^{(k)}] &= \sum_{i=1}^{N}\mathcal{E}[\sigma_i^2\mathcal{E}_s[w_{it-1}^{(k)}]w_{is-1}^{(k)}] \qquad (6.59)\\
&\leq \sum_{i=1}^{N}\mathcal{E}[\sigma_i^2|\sum_{j=1}^{K^*}(\mathbf{e}_{kJ}^{\prime}\mathbf{\Pi}^{*t-s}\mathbf{e}_j)||w_{is-1}^{(j)}w_{is-1}^{(k)}|]\\
&\leq \bar{\sigma}^2\Big(\sum_{j=1}^{K^*}|\mathbf{e}_{kJ}^{\prime}\mathbf{\Pi}^{*t-s}\mathbf{e}_j|\Big)\Big(\sum_{i=1}^{N}\max_j\mathcal{E}[|w_{i0}^{(j)}w_{i0}^{(k)}|]\Big)\ ,
\end{aligned}
$$

since $\mathcal{E}_s[\sigma_i^2 w_{it-1}] = \sigma_i^2\mathcal{E}_s[w_{it-1}]$ and $0 < \sigma_i^2 \leq \bar{\sigma}^2$ w.p.1. Thus, by using *Lemma 2* in Akashi and Kunitomo (2010b), the corresponding order is $O(\log T/T)$. For the third of (6.57),

$$
\begin{aligned}
&\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k)\prime}(\mathbf{I}_N-\mathbf{M}_t)\mathbf{\Lambda}_N^{(\sigma)}(\mathbf{I}_N-\mathbf{M}_s)\boldsymbol{\epsilon}_{s-1}^{(k)}] \qquad (6.60)\\
\leq\ &(\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k)\prime}(\mathbf{I}_N-\mathbf{M}_t)\mathbf{\Lambda}_N^{(\sigma)2}(\mathbf{I}_N-\mathbf{M}_t)\boldsymbol{\epsilon}_{t-1}^{(k)}\boldsymbol{\epsilon}_{s-1}^{(k)\prime}(\mathbf{I}_N-\mathbf{M}_s)^2\boldsymbol{\epsilon}_{s-1}^{(k)}])^{\frac{1}{2}}\\
\leq\ &\bar{\sigma}^2(\mathcal{E}[\boldsymbol{\epsilon}_{t-1}^{(k)\prime}\boldsymbol{\epsilon}_{t-1}^{(k)}\boldsymbol{\epsilon}_{s-1}^{(k)\prime}\boldsymbol{\epsilon}_{s-1}^{(k)}])^{\frac{1}{2}}\ ,
\end{aligned}
$$

since $\lambda_{max}\{\mathbf{\Lambda}_N^{(\sigma)2}\} \leq (\bar{\sigma}^2)^2$ and $(\mathbf{I}_N - \mathbf{M}_t)$ is idempotent. Thus, by using the same arguments used in Akashi and Kunitomo (2010b), we obtain the corresponding order of (6.57) as $O(\log T/\sqrt{T})$. Similarly, we apply the same arguments to the second, third and forth terms of (6.57) by using the fact $0 < [\mathbf{D}_t]_{ii} = m_{ii}^{(t)} \leq 1$. Therefore, $\mathcal{V}ar[\Upsilon_{11n}^{(k)}] = O(\log T/\sqrt{T})$.

For $\Upsilon_{12n}^{(k,a)}$ and $\Upsilon_{12n}^{(k,b)}$, by using the facts that $\mathcal{E}_t[u_{it}^{(f)2}] = \sigma_i^2$ and $\mathcal{E}_t[u_{is}^{(f)} u_{it}^{(f)}] = 0\ (s < t)$,

$$
\begin{aligned}
\mathcal{V}ar[\Upsilon_{12n}^{(k,.)}] &= \frac{1}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}[\tilde{\mathbf{w}}_{t-1}^{(k)\prime}\mathbf{M}_t^D\mathbf{\Lambda}_N^{(\sigma)}\mathbf{M}_t^{D\prime}\tilde{\mathbf{w}}_{t-1}^{(k)}] \\
&\leq \frac{\bar{\sigma}^2}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}[\tilde{\mathbf{w}}_{t-1}^{(k)\prime}(\mathbf{M}_t-\mathbf{D}_t)^2\tilde{\mathbf{w}}_{t-1}^{(k)}] \\
&\leq \frac{\bar{\sigma}^2}{NT}\sum_{t=1}^{T-1}\frac{c_t^2}{(T-t)^2}\mathcal{E}[\tilde{\mathbf{w}}_{t-1}^{(k)\prime}\tilde{\mathbf{w}}_{t-1}^{(k)}] = O(\frac{1}{T})\ .
\end{aligned}
$$

For $\Upsilon_{21n}^{(k,a)}$, by using *Lemma 1* and the same arguments in Akashi and Kunitomo (2010b), we can obtain $\mathcal{V}ar[\Upsilon_{21n}^{(k,a)}] = O(\log T/N)$.

For $\Upsilon_{22n}^{(k,a)}$, it is sufficient to show that

$$
\begin{aligned}
\mathcal{V}ar[\tilde{\mathbf{v}}_{tT}^{*(k,j)\prime}\mathbf{M}_t^D\bar{\mathbf{u}}_{tT}] &= \mathcal{V}ar[\ (\boldsymbol{h}_N^{(j)}\circ\tilde{\mathbf{v}}_{tT}^{**(k,j)})'\mathbf{M}_t^D(\sum_{g=1}^{1+G_2}\beta_g\boldsymbol{h}_N^{(g)}\circ\bar{\mathbf{v}}_{tT}^{(g)})\ ] \qquad (6.61)\\
&\leq (1+G_2)^2\max_g\{\mathcal{V}ar[\ (\boldsymbol{h}_N^{(j)}\circ\tilde{\mathbf{v}}_{tT}^{**(k,j)})'\mathbf{M}_t^D(\boldsymbol{h}_N^{(g)}\circ\beta_g\bar{\mathbf{v}}_{tT}^{(g)})\ ]\} \\
&= O(\mathrm{tr}(\mathbf{M}_t))[m^{(1)}(\tilde{\mathbf{v}}_{tT}^{**(k,j)},\beta_g\bar{\mathbf{v}}_{tT}^{(g)}) + m^{(3)}(\tilde{\mathbf{v}}_{tT}^{**(k,j)},\beta_g\bar{\mathbf{v}}_{tT}^{(g)})]\ ,
\end{aligned}
$$

where the first inequality is due to the Cauchy-Schwartz inequality, and the last equality follows from that the same arguments used for the case of $l = r = p = q$ in *Lemma 1*, and $m^{(2)} \leq m^{(1)}(\tilde{\mathbf{v}}_{tT}^{**(k,j)},\beta_g\bar{\mathbf{v}}_{tT}^{(g)}) = \mathcal{E}[(\mathbf{e}_i'\tilde{\mathbf{v}}_{tT}^{**(k,j)})^2(\beta_g\mathbf{e}_i'\bar{\mathbf{v}}_{tT}^{(g)})^2]$. Therefore, the rest of the proof is to show that $\mathcal{V}ar[\Upsilon_{22n}^{(k,a)}] = O((\log T)^2/N)$, which is quite similar to the ones in Akashi and Kunitomo (2010b).

For the case of $(b)$, by using the fact $\mathrm{tr}(\mathbf{M}_t) = O(1)$, we have

$$
\mathcal{V}ar[\Upsilon_{21n}^{(k,b)}] = O(\frac{\log T}{NT}),\quad \mathcal{V}ar[\Upsilon_{22n}^{(k,b)}] = O(\frac{(\log T)^2}{NT})\ . \qquad (6.62)
$$

For the first term of $\Upsilon_{4n}^{(g,.)}$, by using *Lemma 1*, the order of its variance is $(1/NT)\sum_{t=1}^{T-1}(1/T-t)^2\mathrm{tr}(\mathbf{M}_t) = o(1)$. The orders of the variances of second and third terms for $(\Upsilon_{4n}^{(g,a)},\Upsilon_{4n}^{(g,b)})$ are the same as those of $(\Upsilon_{21n}^{(k,a)},\Upsilon_{21n}^{(k,b)})$, respectively. By the same token, for both $(a)$ and $(b)$ cases, the variance order of fourth term of $\Upsilon_{4n}^{(g,.)}$ are the same as the ones for $(\Upsilon_{22n}^{(k,a)},\Upsilon_{22n}^{(k,b)})$.

Thus, the variances of (6.46) to (6.50) converge to zeros as $N, T \to \infty$. Hence, $\mathbf{b}^{(f)}$ can be evaluated as the sum of expectations of these terms by the mean squared convergence. These expectations are zeros by the construction of $\mathbf{M}_t^D = \mathbf{M}_t - \mathbf{D}_t$ and then $\mathbf{b}^{(f)} = \mathbf{0}$ for both $(a)$ and $(b)$ cases.

The second equality of (6.43) holds for both $(a)$ and $(b)$ cases, since $\mathbf{M}_t = \mathbf{I}_N - (\mathbf{I}_N - \mathbf{M}_t)$ and the results due to *Lemma 3* and *Lemma 4* of Akashi and Kunitomo (2010), which is

$$\mathcal{V}ar[\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\mathbf{W}'_{t-1}(\mathbf{I}_N - \mathbf{M}_t)\mathbf{u}_t] = O(\frac{\log T}{T}) \ . \qquad (6.63)$$

(**Step 4**) : At this step we shall evaluate the asymptotic variance-covariance terms of the modified LIML estimator. For this purpose, we utilize the representation of (6.43). First,

$$\begin{aligned}\mathcal{E}[\mathbf{a}_{1n}\mathbf{a}'_{1n}] &= \frac{1}{NT}\mathbf{D}'\mathbf{J}'\sum_{t=1}^{T-1}\sum_{i,j=1}^{N}\mathcal{E}[\mathbf{w}_{it-1}(1-m_{ii}^{(t)})u_{it}u_{jt}(1-m_{jj}^{(t)})\mathbf{w}'_{it-1}]\mathbf{JD} \\ &= \frac{1}{NT}\mathbf{D}'\mathbf{J}'\sum_{t=1}^{T-1}\sum_{i=1}^{N}\mathcal{E}[\mathcal{E}_t[u_{it}^2](1-m_{ii}^{(t)})^2\mathbf{w}_{it-1}\mathbf{w}'_{it-1}]\mathbf{JD} \ , \qquad (6.64)\end{aligned}$$

where the first equality is from $[\mathbf{I}_N - \mathbf{D}_t]_{ij} = 0$ for $i \neq j$, and the second equality follows from $\mathcal{E}_t[u_{it}u_{jt}] = 0$ for $i \neq j$.
Second,

$$\begin{aligned}\mathcal{E}[\mathbf{a}_{1n}\mathbf{a}'_{2n}] &= \Big(\frac{1}{NT}\mathbf{D}'\mathbf{J}'\sum_{t=1}^{T-1}\mathcal{E}\Big[\sum_{i=1}^{N}\mathbf{w}_{it}(1-m_{ii}^{(t)})u_{it}\sum_{j,k=1}^{N}u_{jt}(m_{jk}^{(t)} - \delta_{jk}m_{jk}^{(t)})\mathbf{u}_{kt}^{\perp\prime}\Big], \mathbf{O}\Big) \\ &= \Big(\mathbf{O}, \mathbf{O}\Big) \ , \qquad (6.65)\end{aligned}$$

where $\delta_{jk} = 1$ if $j = k$, 0 otherwise. The second equality follows from that $(m_{jk}^{(t)} - m_{jj}^{(t)})\mathcal{E}_t[u_{it}u_{jt}\mathbf{u}_{kt}^{\perp}] = \mathbf{0}$ and $\mathcal{E}_t[\mathbf{u}_{kt}^{\perp}] = \mathbf{0}$ for any $i, j, k$.
Third,

$$\begin{aligned}&\mathcal{E}[\mathbf{J}'_{G_2}\mathbf{a}_{2n}\mathbf{a}'_{2n}\mathbf{J}_{G_2}] \\ &= \frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}\Big[\sum_{i,j,i\neq j}^{N}\mathbf{u}_{it}^{\perp}m_{ij}^{(t)}u_{jt}\sum_{k,l,k\neq l}^{N}u_{kt}m_{kl}^{(t)}\mathbf{u}_{lt}^{\perp\prime}\Big] \qquad (6.66) \\ &= \frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}\Big[\sum_{i,j,i\neq j}^{N}\mathbf{u}_{it}^{\perp}m_{ij}^{(t)}u_{jt}(u_{it}m_{ij}^{(t)}\mathbf{u}_{jt}^{\perp\prime} + u_{jt}m_{ji}^{(t)}\mathbf{u}_{it}^{\perp\prime})\Big] \\ &= \frac{1}{NT}\sum_{t=1}^{T-1}\mathcal{E}\Big[\sum_{i,j=1}^{N}(\mathcal{E}_t[\mathbf{u}_{it}^{\perp}u_{it}]\mathcal{E}_t[u_{jt}\mathbf{u}_{jt}^{\perp\prime}] + \mathcal{E}_t[u_{jt}^2]\mathcal{E}_t[\mathbf{u}_{it}^{\perp}\mathbf{u}_{it}^{\perp\prime}])(m_{ij}^{(t)}[1-\delta_{ij}])^2\Big] \ ,\end{aligned}$$

where the last equality is from $m_{ji}^{(t)} = m_{ij}^{(t)}$.

Consider the statements in (3.14) and (3.15) or the cases $c_a = 0, c_b = 0$. As for $\mathbf{\Phi}^*$, by the Cauchy-Schwartz inequality

$$\mathcal{E}\Big[\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{i=1}^{N} m_{ii}^{(t)}|\mathbf{e}_{jJ}'\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\mathbf{e}_{kJ}|\Big] \tag{6.67}$$

$$\leq \mathcal{E}\Big[\frac{1}{T}\sum_{t=1}^{T-1}\Big(\frac{1}{N}\sum_{i=1}^{N} m_{ii}^{(t)2}\Big)^{\frac{1}{2}}\Big(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{e}_{jJ}'\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\mathbf{e}_{kJ})^2\Big)^{\frac{1}{2}}\Big]$$

$$\leq \frac{1}{T}\sum_{t=1}^{T-1}\Big(\frac{\mathrm{tr}(\mathbf{M}_t)}{N}\Big)^{\frac{1}{2}}\Big(\mathcal{E}[(\mathbf{e}_{jJ}'\mathbf{w}_{i0}\mathbf{w}_{i0}'\mathbf{e}_{kJ})^2]\Big)^{\frac{1}{2}},$$

where the last quantity is $O(T\sqrt{T}/T\sqrt{N}) = O(\sqrt{c_a})$ or $O(T/T\sqrt{N}) = O(\sqrt{c_b})$. Because of the mean convergence, these terms converge in probability to zeros, and then $\mathbf{\Phi}^* \to \mathbf{\Phi}_1^*$.
By using the similar arguments, $\mathbf{\Psi}_1^*$ converges to a constant matrix in probability. Moreover, for $g, h = 1, \cdots, G_2$, we can show

$$|\mathbf{e}_g'\mathbf{\Psi}_2^*\mathbf{e}_h| \leq \frac{1}{NT}\sum_{t=1}^{T-1}\mathrm{tr}(\mathbf{M}_t)\times O(1) \tag{6.68}$$

since $|\mathbf{e}_g'(\sigma_i^2\mathcal{E}_t[\mathbf{u}_{jt}^{\perp}\mathbf{u}_{jt}^{\perp\prime}]+\mathcal{E}_t[\mathbf{u}_{it}^{\perp}u_{it}]\mathcal{E}_t[u_{jt}\mathbf{u}_{jt}^{\perp\prime}])\mathbf{e}_h|$ are bounded w.p.1 and $\sum_{i,j=1}^{N} m_{ij}^{(t)2} = \mathrm{tr}(\mathbf{M}_t)$. Thus, we can conclude that $\mathbf{\Psi}_2^* \to \mathbf{O}$.

[ **Step 5** ]: Finally we consider the asymptotic normality of the modified LIML estimator and the modified GMM estimator. For this purpose, define the $(G_2 + K_1)\times 1$ martingale difference sequence by

$$\boldsymbol{\alpha}_t = \frac{1}{\sqrt{N}}\Big[\mathbf{D}'\mathbf{J}_K'\mathbf{W}_{t-1}[\mathbf{I}_N - \mathbf{D}_t]\mathbf{u}_t + (\mathbf{U}_t^{\perp}, \mathbf{O})'[\mathbf{M}_t - \mathbf{D}_t]\mathbf{u}_t\Big] = \boldsymbol{\alpha}_{1t} + \boldsymbol{\alpha}_{2t}\ (, say)$$

where $(1/\sqrt{T})\sum_{t=1}^{T-1}\boldsymbol{\alpha}_{1t} = \mathbf{a}_{1n}$ and $(1/\sqrt{T})\sum_{t=1}^{T-1}\boldsymbol{\alpha}_{2t} = \mathbf{a}_{2n}$ for (6.43). For any $(G_2+K_1)\times 1$ constant vector $\mathbf{d}$ and any function $f(.)$ such $N = f(T)$, $f(T) \to \infty$ as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{d}'\mathcal{E}_t[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t']\mathbf{d} \xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{d}'\mathcal{E}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t']\mathbf{d}\,. \tag{6.69}$$

For the case of $(a)$, (6.69) holds by the assumptions (A-I) to (A-IV) and

$$\frac{1}{NT}\sum_{t=1}^{T-1}\sum_{i=1}^{N}\sigma_i^2\mathbf{w}_{it-1}\mathbf{w}_{it-1}' \xrightarrow{p} \mathcal{E}[\sigma_i^2\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\,. \tag{6.70}$$

For the case of $(b)$, the condition (6.69) is equivalent to (6.70).

By using the similar arguments as the modified LIML estimator, the limiting distribution of $\Phi^* \sqrt{n}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta})$ for the modified GMM estimator is given by

$$\frac{1}{\sqrt{n}} \mathbf{D}' \sum_{t=1}^{T-1} \mathbf{J}' \mathbf{W}_{t-1}' [\mathbf{I}_N - \mathbf{D}_t] \mathbf{u}_t + \frac{1}{\sqrt{n}} \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{J}_{G_2}' \mathbf{V}_t' \\ \mathbf{O} \end{pmatrix} [\mathbf{M}_t - \mathbf{D}_t] \mathbf{u}_t + o_p(1) \ . \tag{6.71}$$

Then for both the modified LIML and GMM estimators, it is enough to investigate the 4-th order moments as the Lyapounov condition for the asymptotic normality. For the case of $(b)$, $|\mathbf{d}'\alpha_{1t}|^4$ are uniformly bounded in $h$ by (A-I). For the case of $(a)$, it sufficient to show the boundedness of $|\mathbf{d}'\alpha_{1t}|^4$ and $|\mathbf{h}'\alpha_{1t}|^4$.
Define $t_i^{(t)}$ by $(\mathbf{d}'\mathbf{U}_t^{\perp\prime}\mathbf{e}_i)$ or $(\mathbf{d}'\mathbf{J}_{G_2}\mathbf{V}'{}_t\mathbf{e}_i)$, $m_{ij}^{(t,D)} = \mathbf{e}_i'[\mathbf{M}_t - \mathbf{D}_t]\mathbf{e}_j$ and re-write $u_j^{(t)} = u_{jt}$, then $\mathcal{E}_t[\ |\mathbf{d}'\mathbf{U}_t^{\perp\prime}\mathbf{M}_t^D\mathbf{u}_t|^4\ ]$ or $\mathcal{E}_t[\ |\mathbf{d}'\mathbf{J}_{G_2}\mathbf{V}_t'\mathbf{M}_t^D\mathbf{u}_t|^4\ ]$ are given by

$$\sum_{i,i',i'',i'''}^{N} \sum_{j,j',j'',j'''}^{N} m_{ij}^{(t,D)} m_{i'j'}^{(t,D)} m_{i''j''}^{(t,D)} m_{i'''j'''}^{(t,D)} \sigma(i,i',i'',i''',j,j',j'',j''') \tag{6.72}$$

where $\sigma(i,i',i'',i''',j,j',j'',j''') = \mathcal{E}_t[t_i^{(t)} t_{i'}^{(t)} t_{i''}^{(t)} t_{i'''}^{(t)} u_j^{(t)} u_{j'}^{(t)} u_{j''}^{(t)} u_{j'''}^{(t)}]$ does not depend on $t$. Then after some evaluations we can summarize the results as *Lemma 2* below, which is similar to *Lemma 5* of Akashi and Kunitomo (2010b). (The proof is omitted.) **(Q.E.D.)**

**Lemma 2:** Define the $(G_2 + K_1) \times 1$ martingale difference sequence by

$$\boldsymbol{\alpha}_{1t} = \frac{1}{\sqrt{N}} \Big[ \mathbf{D}'\mathbf{J}' \sum_{i=1}^{N} \mathbf{w}_{i(t-1)} \mathbf{u}_t \Big] \ , \boldsymbol{\alpha}_{2t} = \frac{1}{\sqrt{N}} \Big[ \begin{pmatrix} \mathbf{U}_t^{\perp\prime} \\ \mathbf{O} \end{pmatrix} \mathbf{N}_t \mathbf{u}_t \Big] \ . \tag{6.73}$$

Then for any $t, N$ and any constant vector $\mathbf{d}$, there exists a positive constant $\Delta$ such that $(i)$ $\mathcal{E}[\ |\mathbf{d}'\alpha_{1t}|^4\ ] \leq \Delta$ and $(ii)$ $\mathcal{E}[\ |\mathbf{d}'\alpha_{2t}|^4\ ] \leq \Delta$ .

# References


[1] Akashi, K. and N. Kunitomo (2010a), "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," Discussion paper CIRJE-F-707, Graduate school of Economics, University of Tokyo.

[2] Akashi, K. and N. Kunitomo (2010b), "The Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equations," Discussion paper CIRJE-F-708, Graduate school of Economics, University of Tokyo.

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

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

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

[7] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[8] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2008), "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.*

[9] Anderson, T.W., Kunitomo, N. and Matsushita, Y. (2010), "On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments," *Journal of Econometrics,*, Vol. 157, 191-204.

[10] Chao, J., N. Swanson, J. Hausman, W. Newey and T. Woutersen (2009), "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments," *Unpublished Manuscript.*

[11] 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.

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

[13] 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.

[14] Kunitomo, N. (2008), "An Optimal Modification of the LIML Estimation for Many Instruments and Persistent Heteroscedasiticity," Discussion paper CIRJE-F-576, Graduate School of Economics, University of Tokyo, Revised.

# APPENDIX : Some Figures

In Figures the distribution functions of the modified GMM and LIML estimators and the GMM and LIML proposed by Akashi and Kunitomo (2010a, b) are shown with the same normalization (4.4) in each case $(a)$ and $(b)$. The marginal limiting distributions for the modified LIML estimator for $(\beta_2, \gamma_{11})$ are $\mathcal{N}(0,1)$ as $N, T \rightarrow \infty$, which is denoted as "o". Note that the modified LIML estimator denoted as $(b)^*$ in Figures 5-8 are normalized by the different way as explained in Section 4. The parameters of our settings and the details of numerical computation method have been explained in Akashi and Kunitomo (2010a, b).

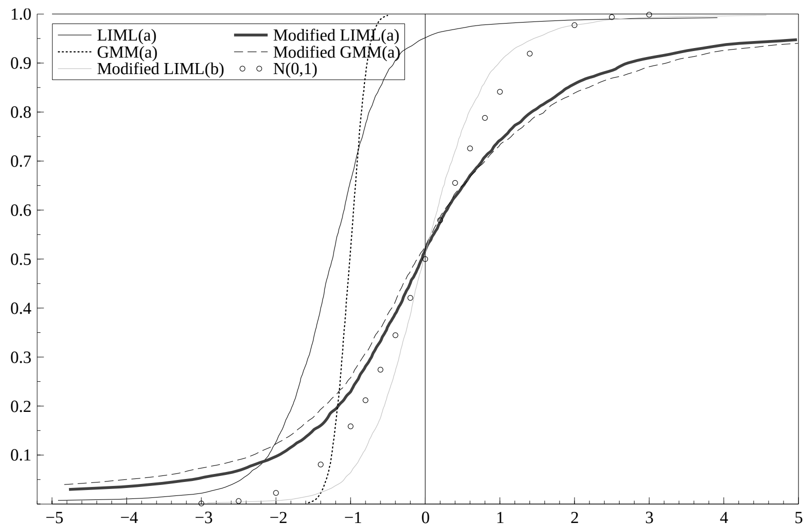


Figure 1: $\beta_2: \ N = 75, \ T = 25, \ c_a = \frac{3}{2}\frac{25}{75}$

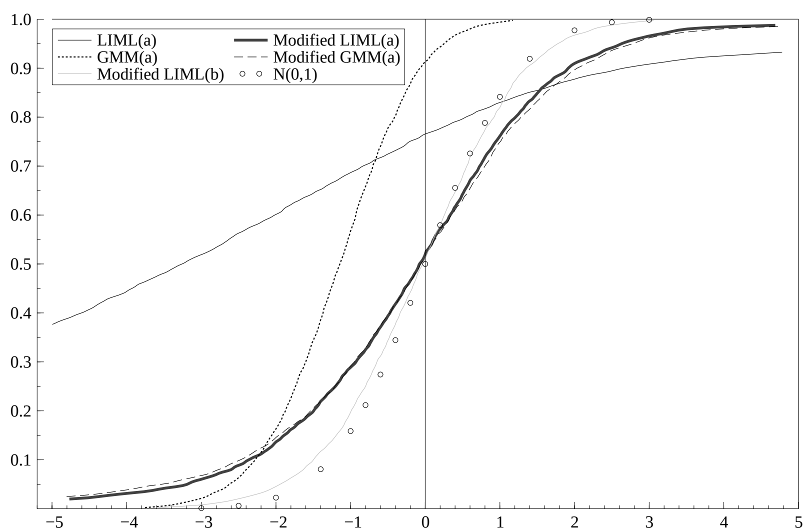


Figure 2: $\gamma_{11}: \ N = 75, \ T = 25, \ c_a = \frac{3}{2}\frac{25}{75}$

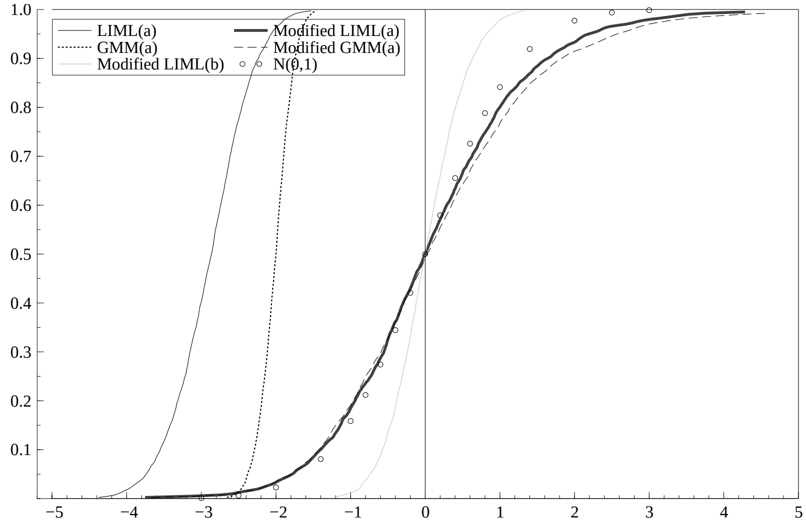


Figure 3: $\beta_2: \ N = 150, \ T = 50, \ c_a = \frac{3}{2}\frac{50}{150}$

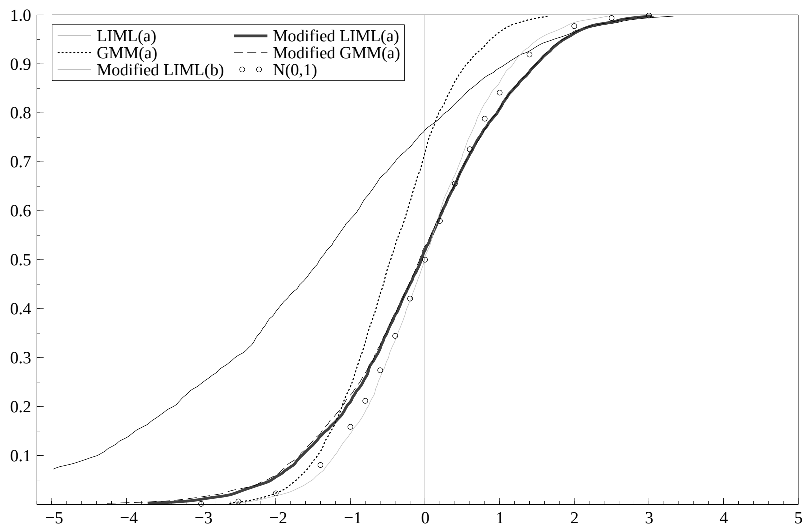


Figure 4: $\gamma_{11}: \ N = 150, \ T = 50, \ c_a = \frac{3}{2}\frac{50}{150}$

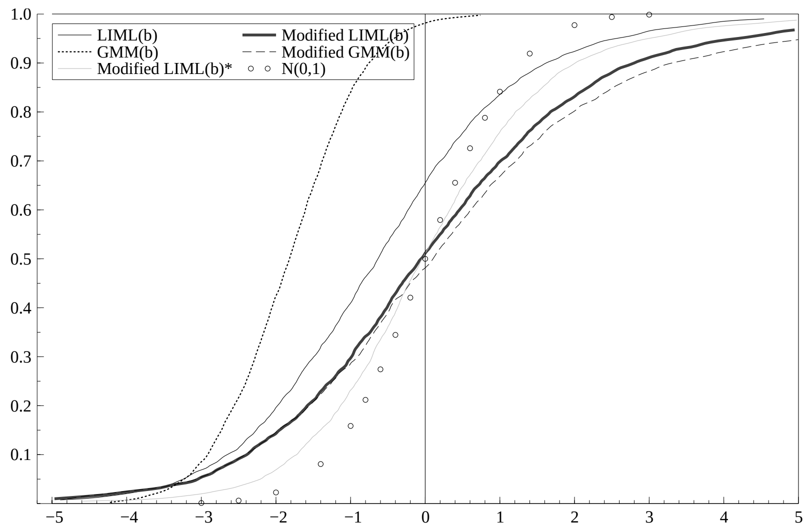


Figure 5: $\beta_2: \ N = 75, \ T = 25, \ c_b = \frac{4}{75}$

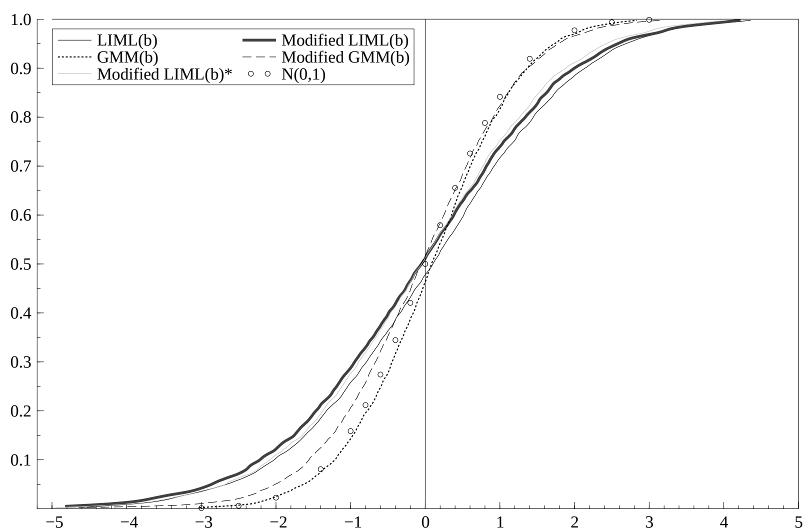


Figure 6: $\gamma_{11}: \ N = 75, \ T = 25, \ c_b = \frac{4}{75}$

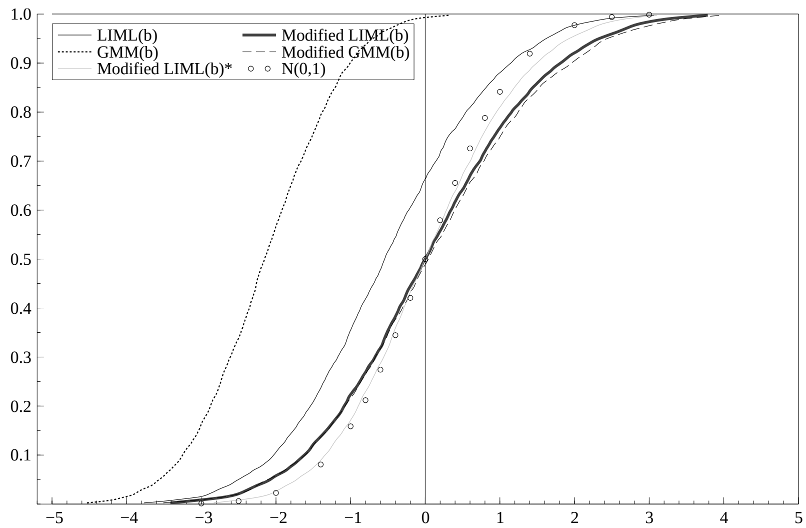


Figure 7: $\beta_2:\ N = 150,\ T = 50,\ c_b = \frac{4}{150}$

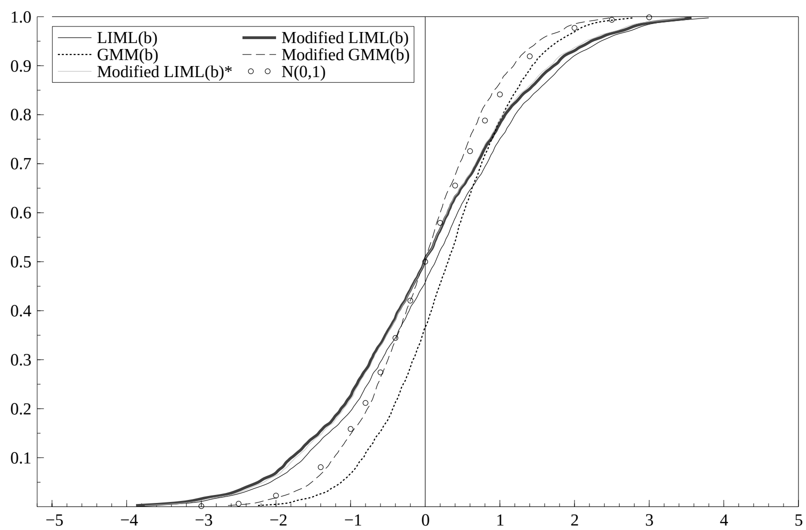


Figure 8: $\gamma_{11}:\ N = 150,\ T = 50,\ c_b = \frac{4}{150}$