CIRJE-F-503

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

Naoto Kunitomo
University of Tokyo

Kentaro Akashi
Graduate School of Economics, University of Tokyo

July 2007

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

Naoto Kunitomo †
and
Kentaro Akashi ‡

January 30, 2007 (1st Version)
July 12, 2007 (2nd Version)

Abstract

We propose the conditional limited information maximum likelihood (CLIML) approach for estimating dynamic panel structural equation models. When there are dynamic effects and endogenous variables with individual effects at the same time, the CLIML estimation method for the doubly-filtered data does give not only a consistent estimation, but also it attains the asymptotic efficiency when the number of orthogonal condition is large. Our formulation includes Alvarez and Arellano (2003), Blundell and Bond (2000) and other linear dynamic panel models as special cases.

Key Words

Dynamic Panel Models, Simultaneous Equations, Individual Effects, Conditional Limited Information Maximum Likelihood (CLIML), Many Orthogonal Conditions, Asymptotic Optimality

---

*KA07-7-12. The earlier version of this paper was presented at the Japan Statistical Society (JSS) meeting (at Sendai) in September 2006.

†Graduate School of Economics, University of Tokyo
‡Graduate School of Economics, University of Tokyo

1. Introduction

Recently there has been a growing interest on dynamic panel econometric models in econometrics. The main reason may be due to the fact that there have been a number of panel data available and their analyses have been growing in many applied fields of economics. Then the econometric methods of panel data are indispensable tools in econometrics by now. (See Hsiao (2003) for instance.) However, there are still non-trivial statistical problems of estimating dynamic panel econometric models to be investigated. In particular, when there are lagged endogenous variables with individual effects and the simultaneity effects in the structural equation of interest exist at the same time, it has been known that the standard statistical methods including the GMM (generalized method of moments) in the econometric literatures or the estimating equation (EE) method in the statistics literatures do not necessarily work well due to the *individual effects* and *the incidental parameters problem*.

In this paper we propose a new econometric method called *the conditional limited information maximum likelihood* (CLIML) approach to the estimation of dynamic panel structural equation models. It is actually an extension of the traditional limited information maximum likelihood (LIML) method, which was originally developed by Anderson and Rubin (1949, 1950). We intend to apply the LIML method to the estimation of dynamic panel structural models when there are dynamic effects and endogenous variables with individual effects at the same time. However, we need to modify the LIML method to handle the dynamic panel models with *incidental parameters* and *many orthogonal conditions*. The CLIML estimation method proposed in this paper gives a consistent estimator and it attains the asymptotic efficiency for a large number of dynamic panel structural equation models when the number of instrumental variables is large in the sense of Anderson and Kunitomo (2006). We also discuss the finite sample properties of the CLIML estimator and show that the finite sample bias of the CLIML estimator is small, which makes the CLIML estimation quite different from the standard GMM estimation. Since the existence of the exact moments of the CLIML estimator is not guaranteed, we need to conduct Monte Carlo experiments carefully. Once we notice the problem precisely, however, it is possible to deal with this moment problem and it is easy to modify the CLIML method without this problem if needed. We have obtained some promising results of the finite sample properties of the CLIML estimator based on Monte Carlo experiments.

In Section 2 we state the formulation of models and alternative estimation methods of unknown parameters in the dynamic panel structural equations with many instruments. Then in Section 3 we give the results of the asymptotic properties of the CLIML estimation method and its asymptotic optimality. In Section 4 we shall discuss some modified methods of the CLIML method and in Section 5 we discuss the finite sample properties of the CLIML estimator. Some concluding remarks will be given in Section 6 and some details of the proofs of our theorems will be in Section 7. Also we shall give some figures in Appendix.

2. Conditional Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equations

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

$$y_{it}^{(1)} = \sum_{j=2}^{1+G_2} \beta_{2j} y_{it}^{(j)} + \sum_{j=1}^{1+G_2} \beta_{3j} y_{it-1}^{(j)} + \sum_{j=1}^{k_1} \beta_{4j} z_{it}^{*(j)} + \eta_i + u_{it} \quad (i = 1, \cdots, N; t = 1, \cdots, T), \tag{2.1}$$

where $y_{it}^{(j)}$ $(j = 1, \cdots, 1 + G_2)$ are the endogenous variables in the system, $z_{it}^{*(j)}$ $(j = 1, \cdots, k_1)$ are the included exogenous variables, $\beta_{l,j}$ $(l = 2, 3, j = 1, \cdots, 1 + G_2; l = 4, j = 1, \cdots, k_1)$ are the unknown coefficients of the right-hand side variables, η_i $(i = 1, \cdots, N)$ are individual effects (fixed or random) and u_{it} are mutually independent (over individuals) disturbance terms with $\mathcal{E}(u_{it}) = 0$ and $\mathcal{E}(u_{it}^2) = \sigma^2$. In (2.1) we allow some coefficients in β_{3j} can be zeros and we denote the original sample size $n = NT$.

We rewrite the dynamic panel structural equation given by

$$
y_{it}^{(1)} = \boldsymbol{\beta}_2^{'} \boldsymbol{y}_{it}^{(2)} + \boldsymbol{\gamma}_1^{'} \boldsymbol{z}_{it}^{(1)} + \eta_i + u_{it} \quad (i = 1, \cdots, N; t = 1, \cdots, T), \tag{2.2}
$$

where $y_{it}^{(1)}$ and $\boldsymbol{y}_{it}^{(2)}$ $(G_2 \times 1)$ are $1 + G_2$ endogenous variables, $\boldsymbol{z}_{it}^{(1)}$ is the K_1 vector of the included predetermined variables in (2.1), $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters. We use the notation such that the vector $\boldsymbol{z}_{it}^{(1)}$ consists of the variables $y_{it-1}^{(j)}$ and $z_{it}^{*(j)}$ $(j = 1, \cdots, 1 + G_2)$ and possibly other lagged endogenous variables $y_{it-l}^{(j)}$ $(l = 1, \cdots, p)$ in this representation.

We assume that the reduced form equation is

$$
\mathbf{y}_{it} = \boldsymbol{\Pi}^{'} \mathbf{z}_{it} + \boldsymbol{\pi}_0^{'} \eta_i + \mathbf{v}_{it} \ , \tag{2.3}
$$

where $\mathbf{y}_{it} = (y_{it}^{(1)}, \boldsymbol{y}_{it}^{(2)'})'$ is the $(1+G_2)$ vector of endogenous variables, $\mathbf{z}_{it}$ is the $K \times 1$ $(n \geq 3)$ vector of predetermined variables at t including the K_1 exogenous variables and the lagged endogenous variables, $\boldsymbol{\Pi}$ and $\boldsymbol{\pi}_0$ are a $K \times (1+G_2)$ matrix and a $1 \times (1+G_2)$ vector of coefficients. We assume that the instruments $\mathbf{z}_{it}$ are $\mathcal{F}_{t-1}-$adapted, and $\mathcal{F}_{t-1}$ is the $\sigma-$field generated by $\mathbf{v}_{is}$ $(s \leq t-1)$ and $\mathbf{z}_{is}$ $(s \leq t)$ and $\mathcal{F}_0$ is the initial $\sigma-$field. (The individual effects η_i $(1 = 1, \cdots, N)$ and the initial conditions $\mathbf{y}_{i0}$ are adapted to $\mathcal{F}_0$.) The predetermined variables in $\mathbf{z}_{it}$ are correlated with $\mathbf{y}_{it}$ in the general case if the individual effects η_i $(i = 1, \cdots, N)$ are random; this aspect makes the panel model consisting of (2.2) and (2.3) different from the standard simultaneous equation models. The disturbance terms $\mathbf{v}_{it}$ are mutually independent over individuals with $\mathcal{E}(\mathbf{v}_{it}|\mathcal{F}_{t-1}) = \mathbf{0}$ $(a.s.)$ and $\mathcal{E}(\mathbf{v}_{it}\mathbf{v}_{it}^{'}|\mathcal{F}_{t-1}) = \boldsymbol{\Omega}$ $(a.s.)$.

The relation between the coefficients in (2.2) and (2.3) gives the condition $(1, -\boldsymbol{\beta}_2^{'})\boldsymbol{\Pi}^{'} = (\boldsymbol{\gamma}_1^{'}, \mathbf{0}^{'})$ and $\boldsymbol{\pi}_{21} = \boldsymbol{\Pi}_{22}\boldsymbol{\beta}_2$, where $\boldsymbol{\Pi}_1 = (\boldsymbol{\pi}_{11}, \boldsymbol{\Pi}_{12})$ is a $K_1 \times (1 + G_2)$ matrix, $\boldsymbol{\Pi}_2 = (\boldsymbol{\pi}_{21}, \boldsymbol{\Pi}_{22})$ is a $K_2 \times (1 + G_2)$ matrix and the $(K_1 + K_2) \times (1 + G_2)$ matrix of coefficients is partitioned as

$$
\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} \ .
$$

We give several examples of panel structural equations known in the econometric literatures.

Example 1 ; Alvarez and Arellano (2003) have considered the estimation problem of a simple dynamic model

$$y_{it} = \gamma y_{it-1} + \eta_i + u_{it} \ \ (i = 1, \cdots, N; t = 1, \cdots, T), \tag{2.4}$$

where $|\gamma| < 1$. They have applied the forward-filter [1] to the structural equation of interest and proposed to use the orthogonal condition $\mathcal{E}(z_{is}u_{it}) = 0 \ (s \leq t)$ and $z_{is} = y_{is-1}$. Then the number of orthogonal restrictions they used is $r_n = (T-1)T/2$ in their study. Recently Hayakawa (2006) has suggested that when we use only the variables y_{it-1} and $r_n = T$, we can recover an efficient information of the unknown parameters in some sense. The model of Alvarez and Arellano (2003) can be interpreted as the simple estimating equation in the sense that there is no simultaneity occurred when $G_2 = 0$ in (2.2).

Example 2 : Blundell and Bond (2000) have considered the estimation problem of a dynamic panel structural equation with two endogenous variables given by

$$y_{it} = \beta x_{it} + \gamma y_{it-1} + \eta_i + u_{it} \ \ (i = 1, \cdots, N; t = 1, \cdots, T) \tag{2.5}$$

and

$$x_{it} = \alpha x_{it-1} + \delta\eta_i + \epsilon_{it} \ \ , \tag{2.6}$$

where the disturbance terms u_{it} and ϵ_{it} are correlated, and we have the restrictions $|\gamma| < 1$ and $|\alpha| < 1$. They applied the standard GMM estimation by utilizing the orthogonal conditions $\mathcal{E}(\mathbf{z}_{it}v_{it}) = 0$ and the instrumental variables are $\mathbf{z}_{it}^{'} = (y_{it-1}, x_{it-1})$. The number of orthogonal restrictions in their study is $r_n = 2T$. We may interpret this model as the structural equation when $G_2 = 1$ if we take y_{it} and x_{it} as the endogenous variables, and $\mathbf{z}_{it} = (y_{it-1}, x_{it-1})$ as the vector of predetermined variables.

Example 3 : We maintain the structural equation of interest as (2.5). It may be reasonable to change the second equation of Blundell and Bond (2000) slightly as

$$x_{it} = \sum_{j=1}^{p} \alpha_j^{*} y_{it-j} + \sum_{j=1}^{q} \alpha_j x_{it-j} + \delta\eta_i + \epsilon_{it} \ , \tag{2.7}$$

[1] We shall discuss this procedure shortly.

where p and q are some positive integers. The number of orthogonal restrictions may be $r_n = (p+q)T$ or possibly many more as in Alvarez and Arellano (2003). We may interpret this model as the structural equation when $G_2 = 1$ if we take y_{it} and x_{it} as the endogenous variables, and $\mathbf{z}_{it} = (y_{it-1}, \cdots, y_{it-p}, x_{it-1}, \cdots, x_{it-q})$ are a vector of the predetermined variables.

There are several important aspects of the problem of estimating equations with instrumental variables in the dynamic panel models. First the standard statistical estimation methods do not necessarily have desirable properties because of the presence of incidental parameters η_i $(i = 1, \cdots, N)$. In many econometric applications the number of observations over individuals are large and we have the situation that there are many incidental parameters. In order to deal with this problem, there have been several statistical procedures for the estimating equations with individual effects developed. (See Hsiao (2003) for the details.) Second, some of the known estimation procedures have substantial bias when the panel models become dynamic in the sense that we have lagged explained variables as explanatory variables. In the dynamic panel models, the number of orthogonal conditions becomes large when we have a reasonable length of time series. But then it has been known that the bias of the standard GMM estimation procedure becomes serious. Third, when we have endogenous variables in the structural equations of interest, it has been also known that the standard estimation methods have serious drawbacks.

Instead of refining the traditional estimation methods, we shall develop a new estimation procedure which may overcome these problems at the same time by applying the conditional limited information maximum likelihood (CLIML) estimation approach. The asymptotic properties of the LIML estimation method of structural equation including its asymptotic optimality has been recently investigated by Anderson and Kunitomo (2006), and Anderson, Kunitomo and Matsushita (2006) when there are *many instruments*. We shall extend their analysis to the CLIML estimation method when the number of instruments increases as the sample size, which may be the common situation in the estimation problem of dynamic panel structural equation. Before we apply the CLIML estimation method, however, first we shall propose to use the doubly filtered procedure, which is a data transformation in both forward and backward directions of time and remove their individual effects before estimation.

Let $\mathbf{y}_i^{(1)} = (y_{it}^{(1)})$, $\mathbf{Y}_i^{(2)} = (\mathbf{y}_{it}^{(2)'})$ and $\mathbf{Z}_i^{(1)} = (\mathbf{z}_{it}^{(1)'})$ be $T \times 1$, $T \times G_2$ and $T \times K_1$ matrices. We define the forward deviation operator $\mathbf{A}_f$ ($(T-1)\times T$ upper triangular matrix [2]) used by Alvarez and Arellano (2003) such that $\mathbf{A}_f \mathbf{A}_f^{'} = \mathbf{I}_{T-1}$, $\mathbf{1} = (1, \cdots, 1)^{'}$ and

$$\mathbf{A}_f^{'} \mathbf{A}_f = \mathbf{Q}_T = \mathbf{I}_T - \mathbf{1}_T \mathbf{1}_T^{'} \ . \tag{2.8}$$

We apply the forward deviation operator to the random variables of $\mathbf{y}_i^{(1)} = (y_{it}^{(1)})$, $\mathbf{Y}_i^{(2)} = (\mathbf{y}_{it}^{(2)'})$, $\mathbf{Z}_i^{(1)} = (\mathbf{z}_{it}^{(1)'})$, $\mathbf{Z}_i = (\mathbf{z}_{it}^{'})$ and we denote the resulting variables as $\mathbf{y}_i^{(1,f)} = (y_{it}^{(1,f)})$, $\mathbf{Y}_i^{(2,f)} = (\mathbf{y}_{it}^{(2,f)'})$, $\mathbf{Z}_i^{(1,f)} = (\mathbf{z}_{it}^{(1,f)'})$ and $\mathbf{Z}_i^{(n,f)} = (\mathbf{z}_{it}^{(f)'})$ $(n \geq 3)$. By using the forward-filtered variables, (2.2) becomes

$$y_{it}^{(1,f)} = \boldsymbol{\beta}_2^{'} \boldsymbol{y}_{it}^{(2,f)} + \boldsymbol{\gamma}_1^{'} \boldsymbol{z}_{it}^{(1,f)} + u_{it}^{(f)} \ \ (i = 1, \cdots, N; t = 1, \cdots, T-1), \tag{2.9}$$

where $\mathbf{u}_i^{(f)} = (u_{it}^{(f)})$ is the transformed $((T-1) \times 1)$ vector by $\mathbf{u}_i^{(f)} = \mathbf{A}_f \mathbf{u}_i$ from the $T \times 1$ disturbance vector $\mathbf{u}_i = (u_{it})$.

On the other hand, we apply the backward deviation operator $\mathbf{A}_b$ ($(T-1) \times T$ lower triangular matrix [3]), which satisfies $\mathbf{A}_b \mathbf{A}_b^{'} = \mathbf{I}_{T-1}$, $\mathbf{A}_b^{'} \mathbf{A}_b = \mathbf{I}_T - \mathbf{1}_T \mathbf{1}_T^{'}$, and it removes the individual effects in the backward way instead of the forward way by using $\mathbf{A}_f$. By applying this backward deviation operator to the $(T-1) \times K$ matrix of instrumental variables of $\mathbf{Z}_i = (\mathbf{z}_{it}^{'})$, and we denote the transformed instrumental variables as $\mathbf{Z}_i^{(b)} = (\mathbf{z}_{it}^{(b)'})$ $(i = 1, \cdots, N; t = 2, \cdots, T)$.

Then the orthogonal conditions can be given by

$$\mathcal{E}\left[u_{it}^{(f)} \mathbf{z}_{is}^{(b)}\right] = \mathbf{0} \ \ (2 \leq s \leq t \leq T-1). \tag{2.10}$$

[2] The transformation from x_{it} $(i = 1, \cdots, N)$ to x_{it}^{*} are defined by $x_{it}^{*} = c_t[x_{it} - (1/(T - t))(x_{it+1} + \cdots + x_{iT})]$, $c_t^2 = (T-t)/(T-t+1)$ $(t = 1, \cdots, T-1, T \geq 2)$.

[3] The transformation can be calculated by reversing the time indices of the corresponding variables from the future to the past.

We notice that it is possible to use only the forward deviation operator and not to use the backward deviation operator [4] and then we have the orthogonal conditions of (2.10) with $\mathbf{z}_{is}^{(b)}$. In the above representation the number of orthogonal restrictions can be dependent on the sample size $n = NT$ when T is large and it is dependent on N. If we use the forward deviation operator and use all orthogonal conditions available, $r_n = J(T-1)(T-2)/2$ provided $T \geq 3$ and J $(\leq K)$ is the number of instruments used [5] at each period t. Also it is possible to use only a subset of orthogonal conditions such that

$$r_n = J \sum_{t=2}^{T-1} (t - s(t)), \tag{2.11}$$

where $t - s(t)$ $(t > s(t))$ is the number of past observations used at period t. If we take $J = K$ and $s(t) = t - 1$, we only use the orthogonal condition with $\mathbf{z}_{it}$ at period t and it corresponds to the traditional LIML estimation. In the general case we may call our method as the CLIML estimation because we can use a subset of orthogonal conditions to the forward-filtered data or the doubly-filtered data after transformations.

Let

$$\mathbf{Z}_t^{(n,b)} = \begin{pmatrix} \mathbf{z}_{1s(t)+1}^{(b)'} & \cdots & \mathbf{z}_{1t}^{(b)'} \\ \vdots & \vdots & \vdots \\ \mathbf{z}_{Ns(t)+1}^{(b)'} & \cdots & \mathbf{z}_{Nt}^{(b)'} \end{pmatrix}$$

be an $N \times J(t - s(t))$ matrix of the (backward) filtered instruments with $1 \leq s(t) \leq t$ $(t = 2, \cdots, T-1)$. Also let $\mathbf{y}_t^{(f)} = (y_{it}^{(1,f)}, \mathbf{y}_{it}^{(2,f)'})'$ be $(1 + G_2)$ vectors,

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

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

[4] This procedure may be reasonable when T is greater than 2, but it is not large.

[5] When there are many instruments of lagged endogenous variables as Example 3, we only use a part of variables in (2.3) to form $\mathbf{Z}_t^{(n,b)}$ and use (2.10). Alternatively, we should use a part of orthogonal conditions of (2.10) in order to avoild degeneracy.

$K_1) \times (1 + G_2 + K_1)$ matrices by

$$\mathbf{G} = \sum_{t=2}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_t^{(1,f)'} \end{pmatrix} \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)'} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)'} \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_t^{(1,f)} \right) \tag{2.12}$$

and

$$\mathbf{H} = \sum_{t=2}^{T-1} \begin{pmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_t^{(1,f)'} \end{pmatrix} \left[\mathbf{I}_N - \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)'} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)'} \right] \left(\mathbf{Y}_t^{(f)}, \mathbf{Z}_t^{(1,f)} \right) , \tag{2.13}$$

where we assume that $0 < J(T-2) < N$ and the $[J(t-s(t))] \times [J(t-s(t))]$ matrices $\mathbf{Z}_t^{(n,b)'} \mathbf{Z}_t^{(n,b)}$ $(t = 2, \cdots, T-1)$ are non-singular (a.s.).
Then the CLIML estimator $\hat{\boldsymbol{\theta}}_{LI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.LI}^{'}, -\hat{\boldsymbol{\gamma}}_{1.LI}^{'})')$ of $\boldsymbol{\theta} = (1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})'$ is defined by

$$\left[\frac{1}{n^*} \mathbf{G} - \lambda_n \frac{1}{q_n} \mathbf{H} \right] \hat{\boldsymbol{\theta}}_{LI} = \mathbf{0} , \tag{2.14}$$

where $n^* = N(T-2)$, $q_n = n^* - r_n$ $(T \geq 3, q_n \geq 2)$ and λ_n is the smallest root of

$$\left| \frac{1}{n^*} \mathbf{G} - l \frac{1}{q_n} \mathbf{H} \right| = 0 . \tag{2.15}$$

The solution to (2.14) gives the minimum of the variance ratio

$$\mathrm{VR}_n = \frac{\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right] \mathbf{G} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_2 \end{bmatrix}}{\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right] \mathbf{H} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_2 \end{bmatrix}} . \tag{2.16}$$

Similarly, we define the conditional GMM (or CTSLS) estimator (or the special case of the GMM estimator) $\hat{\boldsymbol{\theta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'}, -\hat{\boldsymbol{\gamma}}_{1.TS}^{'})')$ of $\boldsymbol{\theta} = (1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})'$ by

$$[\mathbf{0}, \mathbf{I}_{G_2+K_1}] \sum_{t=2}^{T-1} \begin{bmatrix} \mathbf{Y}_t^{(f)'} \\ \mathbf{Z}_t^{(1,f)'} \end{bmatrix} \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)'} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)'} \left[\mathbf{Y}_t^{(f)}, \mathbf{Z}_t^{(1,f)} \right] \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \\ -\hat{\boldsymbol{\gamma}}_{2.TS} \end{bmatrix} = \mathbf{0}. \tag{2.17}$$

It minimizes the numerator of the variance ratio (2.16). The LIML and TSLS estimation methods were originally developed by Anderson and Rubin (1949, 1950), and we modify them slightly to develop the conditional LIML and the conditional TSLS methods for the dynamic panel simultaneous equations models with individual effects.

When we use only the forward-filter, the summation in $\mathbf{G}$ and $\mathbf{H}$ should run from $t = 1$ to $t = T - 1$ instead of from $t = 2$ to $t = T - 1$. Then the sample size is $n^* = N(T - 1)$.

3. Asymptotic Properties of the CLIML Method

We investigate the limiting distribution of the CLIML estimator under a set of alternative assumptions when the number of orthogonal conditions r_n can be dependent on n and $n \to \infty$ (as $N, T \to \infty$). We consider the situation when

$$(\mathbf{I}) \qquad \frac{r_n}{n} \longrightarrow c \ \ (0 \leq c < 1) \ .$$

Condition (I) controls that the number of orthogonal conditions is proportional to the number of observations including the case when $c = 0$. Because we need to estimate the covariance matrix of $\mathbf{v}_{it}$, we need the restriction $c < 1$. Let a $K \times (G_2 + K_1)$ matrix

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

As the first condition on the non-centrality, we assume

$$(\mathbf{II}) \qquad \frac{1}{n} \mathbf{D}' \left[\sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)'} \mathbf{M}_t^{(b)} \mathbf{Z}_t^{(n,f)} \right] \mathbf{D} \xrightarrow{p} \mathbf{\Phi}_1^* ,$$

where

$$\mathbf{M}_t^{(b)} = \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)'} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)'} \tag{3.1}$$

and $\mathbf{\Phi}_1^*$ is a $(G_2 + K_1) \times (G_2 + K_1)$ positive definite matrix.
As the second condition on the non-centrality, we assume

$$(\mathbf{III}) \quad \frac{1}{q_n} \mathbf{D}' \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)'} \left[\mathbf{I}_N - \mathbf{M}_t^{(b)} \right] \mathbf{Z}_t^{(n,f)} \mathbf{D} \xrightarrow{p} \mathbf{\Phi}_2^* ,$$

and $\mathbf{\Phi}_2^*$ is a $(G_2+K_1)\times(G_2+K_1)$ non-negative definite matrix.
Condition (II) and Condition (III) control that the non-centrality is proportional to the sample size and they may be quite natural conditions. If we further have the condition $\mathbf{\Phi}_2^*=\mathbf{O}$, then the analysis would be greatly simplified and it could be reasonable if we deal with the stationary processes [6] with respect to t. Since r_n grows with n, it may correspond to the case of *many instruments* in micro-econometric literatures. These conditions on r_n and the non-centrality are the maximal rates of growing the number of incidental parameters in a sense.

It is also convenient to state our results in terms of $\mathcal{E}(\mathbf{v}_{it}\mathbf{v}_{it}^{'})=\mathbf{\Omega}$ (a.s.),

$$\mathbf{w}_{it}^{(2)}=(\mathbf{0},\mathbf{I}_{G_2})\left[\mathbf{v}_{it}-\mathbf{Cov}(\mathbf{v}_{it},u_{it})\frac{u_{it}}{\sigma^2}\right]\ (i=1,\cdots,N), \tag{3.2}$$

and $\sigma^2=\boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}$. Because $u_{it}=\boldsymbol{\beta}^{'}\mathbf{v}_{it}$ and $(1,0,\cdots,0)=\boldsymbol{\beta}^{'}+(0,\boldsymbol{\beta}_2^{'})$ in (2.3) and (2.3), we have a decomposition

$$\begin{aligned}\mathbf{v}_{it} &= \frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}u_{it}+\left[\mathbf{I}_{G_2+1}-\frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\right]\mathbf{v}_{it} \\ &= \frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}u_{it}+\begin{bmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{bmatrix}\mathbf{w}_{it}^{(2)}\ .\end{aligned} \tag{3.3}$$

Then the random variables u_{it} and $\mathbf{w}_{it}^{(2)}(i=1,\cdots,N)$ are uncorrelated and

$$\mathcal{E}(\mathbf{w}_{it}^{(2)}\mathbf{w}_{it}^{(2)'})=\frac{1}{\sigma^2}\left[\mathbf{\Omega}\sigma^2-\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{\Omega}\right]_{22}\ , \tag{3.4}$$

where $[\,\cdot\,]_{22}$ is the $G_2\times G_2$ right-lower corner sub-matrix. We are ready to state that the CLIML estimator is consistent and asymptotically normal under a set of reasonable conditions.

Theorem 3.1 : Let $\mathbf{z}_{it}$ be a set of $K\times 1$ vector, which is $\mathcal{F}_{t-1}$ adapted. Let also $\mathbf{v}_{it}$ be a vector of $(1+G_2)\times 1$ martingale difference sequences such that $\mathcal{E}(\mathbf{v}_{it}|\mathcal{F}_{t-1})=\mathbf{0}$ and $\mathcal{E}(\mathbf{v}_{it}\mathbf{v}_{it}^{'}|\mathcal{F}_{t-1})=\mathbf{\Omega}$ (a.s.). Suppose $r_n\to\infty$ and $q_n=n-r_n\to\infty$ as $N,T\longrightarrow\infty$. In addition to Conditions (I), (II) and (III), we assume

$$\textbf{(IV)}\qquad \frac{1}{n}\max_{2\le t\le T-1}\|\mathbf{D}^{'}\mathbf{Z}_t^{(n,*)}\|^2\xrightarrow{p}0\ ,$$

[6] If $\mathbf{Z}_t^{(n,f)}-\mathbf{Z}_t^{(n,b)}=o_p(1)$, then $\mathbf{\Phi}_2^*=\mathbf{O}$. This condition holds automatically when there are no individuals effects in the panel structural equation of interest.

where $\mathbf{Z}_t^{(n,*)} = \mathbf{Z}_t^{(n,f)}\mathbf{M}_t^{(b)}$.

(i) For $c = 0$,

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^*) \ , \tag{3.5}$$

where $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}$ and

$$\boldsymbol{\Psi}^* = \sigma^2 \boldsymbol{\Phi}_1^{*-1} \ . \tag{3.6}$$

(ii) For $0 < c < 1$, let $\boldsymbol{\Phi}^* = \boldsymbol{\Phi}_1^* - c\boldsymbol{\Phi}_2^*$ and we assume that $\boldsymbol{\Phi}^*$ is a non-singular matrix. Furthermore, suppose that $\mathcal{E}[\|\mathbf{v}_{it}\|^6]$ are bounded and there exist matrices $\boldsymbol{\Xi}_{3.2}$ and $\boldsymbol{\Xi}_{4.2}$ such that

$$\textbf{(V)} \quad \boldsymbol{\Xi}_{3.2} = \left[\frac{1}{1-c}\right]^2 \operatorname*{plim}_{n\to\infty} \mathbf{D}'\frac{1}{n}\sum_{t=2}^{T-1}\sum_{j=1}^{N}\left[\mathbf{Z}_t^{(n,f)'}\left(\mathbf{M}_t^{(b)} - c\mathbf{I}_N\right)\right]_{\cdot j}\left[a_{jj}^{(t)} - c\right]\mathcal{E}(u_{jt}^{(f)2}\mathbf{w}_{jt}^{(f)'}),$$

$$\textbf{(VI)} \quad \boldsymbol{\Xi}_{4.2} = \left[\frac{1}{1-c}\right]^2 \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{t=2}^{T-1}\sum_{j=1}^{N}\left[a_{jj}^{(t)} - c\right]^2 \boldsymbol{\Gamma}_{44.2}^{(t)} \ ,$$

where $a_{jj}^{(t)} = \left[\mathbf{M}_t^{(b)}\right]_{jj}$, $\mathbf{w}_{it}^{(f)} = [\mathbf{0}, \mathbf{I}_{G_2}]\left[\mathbf{I}_{1+G_2} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\right]\mathbf{v}_{it}^{(f)}$, $\boldsymbol{\Gamma}_{44.2}^{(t)} = \mathcal{E}(u_{it}^{(f)2}\mathbf{w}_{it}^{(f)}\mathbf{w}_{it}^{(f)'}) - \mathcal{E}(u_{it}^{(f)2})\mathcal{E}(\mathbf{w}_{it}^{(f)}\mathbf{w}_{it}^{(f)'})$, $\mathbf{v}_{it}^{(f)}$ are the forward-filtered disturbances and $[\ \cdot\ ]_{\cdot j}$ means the j-th column of the corresponding matrix.

Then

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^*) \ , \tag{3.7}$$

where

$$\begin{aligned} \boldsymbol{\Psi}^* \ &= \ \boldsymbol{\Phi}^{*-1}\left\{\sigma^2\boldsymbol{\Phi}^* + c_*\sigma^2\boldsymbol{\Phi}_2^* + c_*\begin{bmatrix}\mathbf{I}_{G_2}\\ \mathbf{O}\end{bmatrix}\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22}[\mathbf{I}_{G_2}, \mathbf{O}]\right. \\ &\qquad \left. + \boldsymbol{\Xi}_{3.2} + \boldsymbol{\Xi}_{3.2}^{'} + \boldsymbol{\Xi}_{4.2}\right\}\boldsymbol{\Phi}^{*-1} \end{aligned} \tag{3.8}$$

and $c_* = c/(1-c)$.

In the general case, the asymptotic covariance (3.8) of the CLIML estimator depend on the third and fourth order moments of disturbance terms $\mathbf{v}_{it} = (v_{it}^{(j)})$.

When the random vectors are followed by the class of elliptically contoured distribution EC($\mathbf{\Omega}$) (see Section 2.7 of Anderson (2003)), for instance, we could simplify (3.8) considerably because the third order moments are zeros and there is a simple expression on the fourth order moments. When the disturbances are normally distributed in particular, $\Xi_{3.2} = \mathbf{O}$ and $\Xi_{4.2} = \mathbf{O}$ are automatically zeros.

Instead of making an assumption on the distributions of disturbance terms except the existence of their moments, alternatively we utilize the condition

$$\textbf{(VII)} \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=2}^{T-1} \sum_{i=1}^{N} \left[a_{ii}^{(t)} - c \right]^2 = 0$$

and $a_{ii}^{(t)} = \left[\mathbf{M}_t^{(b)} \right]_{ii}$.

This condition is often satisfied in practical situations as shown by Anderson and Kunitomo (2006). Then we can simplify the covariance-matrix in *Theorem 3.1* as the next theorem.

Theorem 3.2 : For the case of (ii) of *Theorem 3.1*, in addition to Conditions (I)-(IV) and (VII), we assume that $\mathcal{E}[\|\mathbf{v}_{it}\|^4]$ $(i = 1, \cdots, n)$ are bounded instead of the 6-th order moments. Then

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^*) \tag{3.9}$$

where

$$\mathbf{\Psi}^* = \sigma^2 \mathbf{\Phi}^{*-1} + c_* \mathbf{\Phi}^{*-1} \mathbf{\Phi}_2^* \mathbf{\Phi}^{*-1} + c_* \mathbf{\Phi}^{*-1} \left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{\Omega} \right]_{22} \mathbf{\Phi}^{*-1} \tag{3.10}$$

and $c_* = c/(1-c)$.

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 $\mathbf{G}$ and $\mathbf{H}$. We shall consider a class of estimators which are some functions of these two matrices and then we have a new result on the asymptotic optimality of the CLIML estimator under a set of assumptions. The proof is given in Section

7.

Theorem 3.3 : In the panel structural equations model of (2.2) and (2.3), assume that $\mathbf{\Phi}^* = \mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*$ is a positive definite matrix. Define the class of consistent estimators for $\boldsymbol{\theta} = (1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})'$ by

$$\text{(3.11)} \qquad \begin{pmatrix} \hat{\boldsymbol{\beta}}_2 \\ \hat{\gamma}_1 \end{pmatrix} = \phi(\frac{1}{n}\mathbf{G}, \frac{1}{q_n}\mathbf{H}) ,$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices $\mathbf{G}/n$ and $\mathbf{H}/q_n$ in (2.12) and (2.13) as $r_n \to \infty$ (N and $T \to \infty$) and $0 \le c < 1$. Furthermore, we assume that there exists a positive constant d ($d > 0$) such that

$$\text{(3.12)} \qquad \frac{\partial \mathbf{\Phi}^*}{\partial \rho_{ij}} = d\frac{\partial \mathbf{\Phi}_1^*}{\partial \rho_{ij}}$$

for any $\mathbf{\Phi}_1^* = (\rho_{ij})$. Then either under the conditions of (i) of *Theorem 3.1*, *Corollary 3.1* or *Theorem 3.2*, as $n \to \infty$

$$\text{(3.13)} \qquad \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}) ,$$

where

$$\text{(3.14)} \qquad \mathbf{\Psi} \ge \mathbf{\Psi}^* ,$$

and $\mathbf{\Psi}^*$ is given by (3.6) or (3.10), respectively.

The results reported in this section can be regarded as extensions of *Theorem 1-Theorem 3* of Anderson and Kunitomo (2006) for the standard linear structural equations model to the general panel structural equations model. In their case $\mathbf{\Phi}_2^* = \mathbf{O}$ and (3.12) is automatically satisfied. In the more general cases in the LINEAR panel econometric models, however, the condition given by (3.12) may be strong and it could be weaken in some situations.

4. Improving CLIML methods

In the estimation problem of structural equation methods, there have been some discussions on improving the limited information maximum likelihood (LIML) method. Anderson, Kunitomo and Morimune (1986) have investigated various types of estimation methods. One important class of estimation methods is a class of modified LIML method including the one proposed by Fuller (1977). We consider the class of modified CLIML estimators given by $\hat{\boldsymbol{\theta}}_{MLI}$ $(= (1, -\hat{\boldsymbol{\beta}}'_{2.MLI}, -\hat{\boldsymbol{\gamma}}'_{1.MLI})')$ of $\boldsymbol{\theta} = (1, -\boldsymbol{\beta}'_2, -\boldsymbol{\gamma}'_1)'$, which is the solution

$$\left[\frac{1}{n}\mathbf{G} - \lambda_n^* \frac{1}{q_n}\mathbf{H}\right] \hat{\boldsymbol{\theta}}_{LI} = \mathbf{0} \ , \tag{4.1}$$

where $\lambda_n^* = \lambda_n - a/n$, a is a constant $(0 \leq a \leq 4)$ and λ_n is the smallest root of (2.14).

By using a set of Monte Carlo experiments, we have investigated the finite sample properties of some modified estimators when the LIML estimator ($a = 0$), $a = 4$ (due to T.W. Anderson) and $a = 1$ (due to W. Fuller) in particular. Since the Fuller modified LIML eatimator has finite moments in the simplest case, we have expected that it improves the asymptotic bias of the LIML estimator. When we take $a = 1$, we have found that the bias of the CLIML estimator is reduced in the sense of the Monte Carlo expectation. However, we also find that the CLIML estimator is almost median-unbiased around the true parameter values. When we take $a = 4$, the mean squared errors (MSE) can be further reduced. These observations are similar to those in the eariler studies on the finite sample properties of alternative estimators in the simple structural equations without individual effects (see Anderson et al. (1982, 1986)).

5. Finite Sample Properties of Alternative Estimators

There have been many studies on the finite sample properties of alternative estimators for the structural equation models. One common method often used has been to conduct Monte Carlo experiments. However, there should be non-trivial problem existed and it has been known in econometrics that the LIML estimator does not possess any moments of positive integer order under a set of reasonable assumptions. Therefore, instead of moments we need to investigate the exact cumulative distributions of the CLIML estimator and its modifications directly in a systematic way. The problem of non-existence of moments had been already discussed in the econometric literature, but it does not imply that the LIML estimator should not be used and we should be careful for the loss function. One common example in the statistics literatures is the estimation of the reciprocal of non-zero Gaussian mean. (See Anderson et al. (2005) for the details.)

The evaluation method of the cdfs of estimators we have used in this study is based on the simulation method. In order to describe our evaluation method, we use the classical notation of Anderson et al. (2006) for the ease of comparison except the sample size being n and we concentrate on the comparison of the estimators of the coefficient parameter of the endogenous variable when $G_2 = 1$ for the ease of interpretation. To specify the exact distributions of estimators we use the *key parameters* used by Anderson et. al. (2005) in the study of the finite sample properties of the CLIML estimator. We have investigated the exact finite sample distributions of the estimator normalized such that the limiting distribution is the standard normal in Theorem 3.2 as

$$\sqrt{n}\boldsymbol{\Psi}^{*-1/2}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{pmatrix} . \tag{5.1}$$

For the experiments we use Example 2 and Example 3 as typical cases. We have the structural equation of interest in (2.5) and

$$x_{it} = \sum_{j=1}^{q} \alpha_j x_{it-j} + \delta\eta_i + \epsilon_{it} \ , \tag{5.2}$$

where we take $q = 2, 5$ and we take $\rho = 0.3$ as the correlation coefficient between the disturbances u_{it} and ϵ_{it}. We have investigated a number of cases and we give only eight figures in Appendix. They show the distribution functions of the CLIML estimator and the GMM estimator (or the CTSLS estimator) for β and γ in (2.5). In figures *all* means the case when we use all available orthogonal conditions and instruments while *min* means the case when we use a set of instruments only with lag one. We have normalized the estimators by the limiting covariance matrices such that the limiting distribution of the CLIML estimator is N(0,1) as Theorem 3.2.

Because the true process of x_{it} is AR(5), we can expected that the CLIML estimator with all instruments gives a reasonable performance. One important finding is that the GMM (or TSLS) estimator with many instruments has significant bias. The speed of approaching to the limiting distribution N(0,1) of the CLIML estimator is much faster than other estimators. These findings agree with the results reported by Anderson et al. (2006). Additionally, we give some figures of the distribution function of the Within-Group (WG) estimator in some cases. As we had expected, its biases of the WG estimator are very large in comparison with other estimation methods.

Although we have presented a limited number of figures from a large number of our simulations, we have found several interesting observations. The finite sample bias of the CLIML estimator is much more smaller than the corresponding GMM estimator. The variance of the GMM estimator may be smaller than the CLIML estimator, but the effects of finite sample bias dominates the MSE. As n grows, the finite sample distribution of the CIML estimator approaches to the standard normal distribution and its speed is much faster than the GMM estimator.

It is possible to use the CLIML method with the backward-filtered data to the cases when $T \geq 3$. When T is very small, however, it may be reasonable to use the forward-filtered transformation only because ot the resulting small sample considerations.

6. Conclusions

In this paper we have proposed to use the conditional limited information maximum likelihood (CLIML) approach for the estimation of dynamic panel simultaneous equation models. When there are dynamic effects and lagged endogenous variables with individual effects at the same time, the CLIML estimation method for the doubly-filtered (or the forward-filtered) data does give not only a consistent estimator, but also it attains the asymptotic efficiency when the number of orthogonal conditions is large or many instruments in some sense.

We also develop some modified CLIML methods which improve the finite sample properties of the standard CLIML estimation. We have given some persuasive numerical results on the finite sample properties of the distribution functions of the CLIML estimator based on a set of extensive Monte Carlo experiments. Because our approach can be applied to the general panel simultaneous equations with dynamic effects and individual effects, the CLIML method would be important for solving practical problems with panel data.

There are several problems remained to be investigated on the estimation of structural dynamic panel econometric models. Our approach can be extended to the multivariate case and also the case with time specific effects. It is also important to develop the test procedures in the dynamic panel structural equations model and the choice procedure of instruments in estimation. They are currently under investigation and the results will be reported in another occasion.

7 Mathematical Details

In section we give the proofs of *Theorems* in Section 3. The method of proofs are similar to those used in Anderson and Kunitomo (2006).

Proof of Theorem 3.1 :

[**Step 1**] Substitution of (2.3) and (2.9) into (2.12) yields

$$\mathbf{G} = \mathbf{G}^{(1)} + \mathbf{G}^{(2)} + \mathbf{G}^{(2)\prime} + \mathbf{G}^{(3)} , \tag{7.1}$$

where

$$\begin{aligned}
\mathbf{G}^{(1)} &= \mathbf{D}^{*\prime} \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)\prime} \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)\prime} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)\prime} \mathbf{Z}_t^{(n,f)} \mathbf{D}^* , \\
\mathbf{G}^{(2)} &= \mathbf{D}^{*\prime} \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)\prime} \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)\prime} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)\prime} (\mathbf{V}_t^{(f)}, \mathbf{O}) , \\
\mathbf{G}^{(3)} &= \sum_{t=2}^{T-1} \begin{pmatrix} \mathbf{V}_t^{(f)\prime} \\ \mathbf{O} \end{pmatrix} \mathbf{Z}_t^{(n,b)} (\mathbf{Z}_t^{(n,b)\prime} \mathbf{Z}_t^{(n,b)})^{-1} \mathbf{Z}_t^{(n,b)\prime} (\mathbf{V}_t^{(f)}, \mathbf{O}) ,
\end{aligned}$$

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

$$\mathbf{D}^* = \left[\mathbf{\Pi}, \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix} \right] .$$

Then by Conditions (I)-(III) as $n \longrightarrow \infty$ $(N, T \to \infty)$,

$$\frac{1}{n}\mathbf{G} \xrightarrow{p} \mathbf{G}_0 = \begin{bmatrix} (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1') \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \mathbf{\Phi}_1^* \left[\begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}, \mathbf{I}_{G_2+K_1} \right] + c \begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \tag{7.2}$$

and

$$\frac{1}{q_n}\mathbf{H} \xrightarrow{p} \mathbf{H}_0 = \begin{bmatrix} (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1') \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \mathbf{\Phi}_2^* \left[\begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}, \mathbf{I}_{G_2+K_1} \right] + \begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} . \tag{7.3}$$

Then we have

$$\left| [\mathbf{\Phi}_1 - (\mathrm{plim}_{n\to\infty}\lambda_n)\mathbf{\Phi}_2] - [c - (\mathrm{plim}_{n\to\infty}\lambda_n)] \begin{bmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \right| = 0 , \tag{7.4}$$

where

$$\mathbf{\Phi}_1 = \begin{bmatrix} (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1') \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \mathbf{\Phi}_1^* \left[\begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}, \mathbf{I}_{G_2+K_1} \right], \mathbf{\Phi}_2 = \begin{bmatrix} (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1') \\ \mathbf{I}_{G_2+K_1} \end{bmatrix} \mathbf{\Phi}_2^* \left[\begin{pmatrix} \boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{pmatrix}, \mathbf{I}_{G_2+K_1} \right].$$

By setting $\mathbf{\Phi}^* = \mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*$ (which is nonsingular) and $\mathbf{\Phi} = \mathbf{\Phi}_1 - c\mathbf{\Phi}_2$, we find that $\lambda_n \xrightarrow{p} c$, $\hat{\beta}_{LI} \xrightarrow{p} \beta$ and $\hat{\gamma}_{1.LI} \xrightarrow{p} \gamma_1$ as $n \to \infty$ because $\mathbf{\Phi}$ $(= \mathbf{\Phi}_1 - c\mathbf{\Phi}_2)$ satisfies

$$\mathbf{\Phi} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{pmatrix} = \mathbf{0} . \tag{7.5}$$

Define $\mathbf{G}_1$ and $\mathbf{H}_1$ by $\mathbf{G}_1 = \sqrt{n}[(1/n)\mathbf{G} - \mathbf{G}_0]$, $\mathbf{H}_1 = \sqrt{q_n}[(1/q_n)\mathbf{H} - \mathbf{H}_0]$, $\lambda_{1n} = \sqrt{n}[\lambda_n - c]$ and

$$\mathbf{b}_1 = \sqrt{n} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta} \\ -(\hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1) \end{bmatrix} .$$

By substituting these random variables into (2.14), it is asymptotically equivalent to

$$[\mathbf{G}_0 - c\,\mathbf{H}_0]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ \boldsymbol{\gamma}_1 \end{bmatrix} \;+\; \frac{1}{\sqrt{n}}[\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} + \frac{1}{\sqrt{n}}[\mathbf{G}_0 - c\,\mathbf{H}_0]\,\mathbf{b}_1$$

$$-\frac{1}{\sqrt{q_n}}[c\mathbf{H}_1]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} = o_p(\frac{1}{\sqrt{n}}) \ .$$

Then by using (7.4), we have

$$\text{(7.6)}\quad [\boldsymbol{\Phi}_1 - c\boldsymbol{\Phi}_2]\begin{bmatrix} 0 \\ \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{bmatrix} = [\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} + o_p(1) \ .$$

Multiplication of (7.6) from the left by $(\boldsymbol{\beta}^{'}, -\boldsymbol{\gamma}_1^{'}) = (1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'})$ yields

$$\text{(7.7)}\qquad \lambda_{1n} = \frac{\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right]\left[\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1\right]\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right]^{'}}{\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right]\mathbf{H}_0\left[1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}\right]^{'}} + o_p(1) \ .$$

Also the multiplication of (7.6) from the left by $(\mathbf{0}, \mathbf{I}_{G_2+K_1})$ and substitution for λ_{1n} from (7.6) yields

$$\begin{aligned}
&\boldsymbol{\Phi}^*\sqrt{n}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{bmatrix} \\
= \;& [\mathbf{0}, \mathbf{I}_{G_2+K_1}]\,[\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} + o_p(1) \\
= \;& [\mathbf{0}, \mathbf{I}_{G_2+K_1}]\left[\mathbf{I}_{1+G_2+K_1} - \frac{1}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\begin{pmatrix} \boldsymbol{\Omega}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}(1, -\boldsymbol{\beta}_2^{'}, -\boldsymbol{\gamma}_1^{'}))\right][\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1]\begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} \\
&+ o_p(1) \ .
\end{aligned}$$

By using the similar arguments of Anderson and Kunitomo (2006) to the present situation, we have

$$
\begin{aligned}
(7.8)\quad & [\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1] \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \\ -\boldsymbol{\gamma}_1 \end{bmatrix} \\
= \quad & \frac{1}{\sqrt{n}}\mathbf{D}^{*'} \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)'}\mathbf{Z}_t^{(n,b)}(\mathbf{Z}_t^{(n,b)'}\mathbf{Z}_t^{(n,b)})^{-1}\mathbf{Z}_t^{(n,b)'}\mathbf{u}_t^{(f)} \\
& + \frac{1}{\sqrt{n}}\left[\sum_{t=2}^{T-1}\begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{pmatrix} \mathbf{Z}_t^{(n,b)}(\mathbf{Z}_t^{(n,b)'}\mathbf{Z}_t^{(n,b)})^{-1}\mathbf{Z}_t^{(n,b)'}\mathbf{u}_t^{(f)} - r_n(\begin{matrix} \boldsymbol{\Omega\beta} \\ \mathbf{O} \end{matrix})\right] \\
& - \sqrt{cc_*}\frac{1}{\sqrt{q_n}}\mathbf{D}' \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)'}[\mathbf{I}_N - \mathbf{Z}_t^{(n,b)}(\mathbf{Z}_t^{(n,b)'}\mathbf{Z}_t^{(n,b)})^{-1}\mathbf{Z}_t^{(n,b)'}]\mathbf{u}_t^{(f)} \\
& - \sqrt{cc_*}\frac{1}{\sqrt{q_n}}\left[\sum_{t=2}^{T-1}(\begin{matrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{matrix})[\mathbf{I}_N - \mathbf{Z}_t^{(n,b)}(\mathbf{Z}_t^{(n,b)'}\mathbf{Z}_t^{(n,b)})^{-1}\mathbf{Z}_t^{(n,b)'}]\mathbf{u}_t^{(f)} - q_n(\begin{matrix} \boldsymbol{\Omega\beta} \\ \mathbf{O} \end{matrix})\right] ,
\end{aligned}
$$

where $\mathbf{u}_t^{(f)'} = (u_{1t}^{(f)}, \cdots, u_{Nt}^{(f)})$ and $r_n + q_n = n$.

[**Step 2**] When $c = 0$, both third and fourth terms of (7.8) are zeros, and the second term of (7.8) converges to zeros. Then by using the central limit theorem for martingale differences, we have the result.

[**Step 3**] We consider the case when $0 < c < 1$. We use the relations $\sqrt{cc_*}/\sqrt{q_n} - c_*/\sqrt{n} = o(1)$ and set

$$
(7.9) \qquad \mathbf{N}_t^{(b)} = \mathbf{M}_t^{(b)} - c_*(\mathbf{I}_N - \mathbf{M}_t^{(b)}) \sim \frac{1}{1-c}\left[\mathbf{M}_t^{(b)} - c\mathbf{I}_N\right] .
$$

Then

$$
\begin{aligned}
\boldsymbol{\Phi}^*\sqrt{n}\begin{bmatrix} \hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\gamma}}_{1.LI} - \boldsymbol{\gamma}_1 \end{bmatrix} &= \frac{1}{\sqrt{n}}\mathbf{D}' \sum_{t=2}^{T-1} \mathbf{Z}_t^{(n,f)'}\mathbf{N}_t^{(b)}\mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}}\sum_{t=2}^{T-1}\mathbf{W}_t^{(f)'}\mathbf{N}_t^{(b)}\mathbf{u}_t^{(f)} \\
&= \mathbf{A}_{1n} + \mathbf{A}_{2n}\ (, say) ,
\end{aligned}
$$

where

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

Then we can evaluate the asymptotic variance-covariance terms of the CLIML estimator. We first notice that

$$\left[\mathbf{N}_t^{(b)}\right]^2 = \mathbf{M}_t^{(b)} + c_*^2(\mathbf{I}_N - \mathbf{M}_t^{(b)})$$

and

$$\mathcal{E}[\mathbf{A}_{1n}\mathbf{A}_{1n}^{'}] = \frac{1}{n}\mathbf{D}^{'}\sum_{t=2}^{T-1}\mathbf{Z}_t^{(n,f)'}\left[\mathbf{M}_t^{(b)} + c_*^2(\mathbf{I}_N - \mathbf{M}_t^{(b)})\right]\mathbf{Z}_t^{(n,f)}\mathbf{D}\ .$$

By using the $i-th$ unit vector $\mathbf{e}_i^{'} = (0,\cdots,0,1,0,\cdots,0)$,

$$\begin{aligned}
\mathcal{E}[\mathbf{A}_{1n}\mathbf{A}_{2n}^{'}] &= \frac{1}{n}\mathbf{D}^{'}\sum_{t=2}^{T-1}\mathbf{Z}_t^{(n,f)'}\mathbf{N}_t^{(b)}\mathcal{E}\left[\mathbf{u}_t^{(f)}\mathbf{u}_t^{(f)'}\mathbf{N}_t^{(b)}\mathbf{W}_t^{(f)}\right] \\
&= \frac{1}{n}\mathbf{D}^{'}\sum_{t=2}^{T-1}\mathbf{Z}_t^{(n,f)'}\mathbf{N}_t^{(b)}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{e}_i\mathbf{e}_i^{'}\mathcal{E}[u_{it}^{(f)2}\mathbf{N}_t^{(b)}\mathbf{e}_j\mathbf{w}_{jt}^{(f)'}] \\
&= \left[\frac{1}{1-c}\right]^2\frac{1}{n}\mathbf{D}^{'}\sum_{t=2}^{T-1}[\sum_{i=1}^{N}\mathbf{Z}_t^{(n,f)'}(\mathbf{M}_t^{(b)} - c\mathbf{I}_N)]_{\cdot i}[a_{ii}^{(t)} - c]\mathcal{E}[u_{it}^{(f)2}\mathbf{w}_{it}^{(f)'}]\ ,
\end{aligned}$$

and

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

Then the first term converges

$$\frac{1}{n}\sum_{t=2}^{T-1}\mathrm{tr}(\mathbf{N}_t^{(b)})\sigma^2\mathcal{E}\left[\mathbf{w}_{it}^{(f)}\mathbf{w}_{it}^{(f)'}\right] \longrightarrow c_*\mathcal{E}\left[\mathbf{w}_{it}^{(2)}\mathbf{w}_{it}^{(2)'}\right] \tag{7.10}$$

as $n\to\infty$ because we have

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

as $n\to\infty$. For any vector $\mathbf{b}$, the second term becomes

$$\begin{aligned}
&\mathbf{b}^{'}\frac{1}{n}\sum_{t=1}^{T-1}\mathcal{E}\left[\mathbf{W}_t^{(f)'}\mathbf{N}_t^{(b)}(\mathbf{u}_t^{(f)}\mathbf{u}_t^{(f)'} - \sigma^2\mathbf{I}_N)\mathbf{N}_t^{(b)}\mathbf{W}_t^{(f)}\right]\mathbf{b} \qquad (7.11)\\
&= \frac{1}{n}\sum_{t=2}^{T-1}\sum_{j=1}^{N}(\mathbf{e}_j^{'}\mathbf{N}_t^{(b)}\mathbf{e}_j)^2\mathcal{E}\left[(u_{it}^{(f)2} - \sigma^2)(\mathbf{w}_{jt}^{(f)'}\mathbf{b})^2\right] \\
&= \left[\frac{1}{1-c}\right]^2\frac{1}{n}\sum_{t=2}^{T-1}\sum_{j=1}^{N}(\mathbf{e}_j^{'}\mathbf{M}_t^{(b)}\mathbf{e}_j)^2\mathcal{E}\left[(u_{it}^{(f)2} - \sigma^2)(\mathbf{w}_{jt}^{(f)'}\mathbf{b})^2\right] \longrightarrow \mathbf{b}^{'}\mathbf{\Xi}_{4.2}\mathbf{b}
\end{aligned}$$

as $n \to \infty$ by using the similar calculations as $\mathcal{E}(\mathbf{A}_{1n}\mathbf{A}_{2n}^{'})$.
Finally, by using the Lyapounov-type martingale central limit theorem, we have the asymptotic normality of (7.10) with the asymptotic covariance matrix $\mathbf{\Psi}^*$. This proves the second part of *Theorem 3.1*.
Q.E.D.

Proof of Theorem 3.2 :

Let

$$\mathbf{\Xi}_{4.2}^{n} = \left[\frac{1}{1-c}\right]^2 \frac{1}{n} \sum_{t=2}^{T-1} \sum_{j=1}^{N} \left[a_{jj}^{(t)} - c\right]^2 \mathbf{\Gamma}_{44.2}^{(t)} \tag{7.12}$$

and $\mathbf{\Xi}_{3.2}^{(n)}$ be defined by Conditions (V) and (VI) of *Theorem 3.1* without their limits. Since $\mathcal{E}[u_{it}^{(f)2}\mathbf{w}_{jt}^{(f)}]$ and $\mathcal{E}[u_{it}^{(f)2}\mathbf{w}_{jt}^{(f)}\mathbf{w}_{jt}^{(f)'}]$ are bounded, Condition (VII) implies that $\mathbf{\Xi}_{4.2}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$, which in turn leads to $\mathbf{\Xi}_{3.2}^{(n)} \xrightarrow{p} 0$ as $n \to \infty$.
Q.E.D.

Proof of Theorem 3.3 :

Without the loss of generality, we consider the case when $K_1 = 0$ and $\gamma_1 = \mathbf{0}$. We set the vector of true parameters $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'}) = (1, \beta_2, \cdots, \beta_{1+G_2})$. Then an estimator of the vector $\boldsymbol{\beta}_2$ is composed of

$$\hat{\beta}_i = \phi_i(\frac{1}{n}\mathbf{G}, \frac{1}{q_n}\mathbf{H}) \;\; (i = 2, \cdots, 1 + G_2)\ . \tag{7.13}$$

For the estimator to be consistent, we need the conditions

$$\beta_i = \phi_i \left[\begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{\Phi}_1^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + c\,\mathbf{\Omega}, \begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{\Phi}_2^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + \mathbf{\Omega} \right] \;\; (i = 2, \cdots, 1{+}G_2) \tag{7.14}$$

as identities with respect to parameters $\boldsymbol{\beta}_2$, $\mathbf{\Phi}_1^*$, $\mathbf{\Phi}_2^*$ and $\mathbf{\Omega}$.

First, we consider the role of $\mathbf{\Omega}$ in (7.11). By differentiating (7.11) with respect to ω_{ij} $(i, j = 1, \cdots, 1 + G_2)$, we have the condition

$$c\frac{\partial \phi_k}{\partial g_{ij}} = -\frac{\partial \phi_k}{\partial h_{ij}} \; (k = 2, \cdots, 1 + G_2; i, j = 1, \cdots, 1 + G_1) \tag{7.15}$$

evaluated at the probability limit of (7.11). Let two $(1+G_2)\times(1+G_2)$ matrices

$$\text{(7.16)}\qquad \mathbf{T}_1^{(k)} = \left(\frac{\partial \phi_k}{\partial g_{ij}}\right) = (\tau_{ij}^{(k)}) \quad (k=2,\cdots,1+G_2; i,j=1,\cdots,1+G_2)$$

and $\mathbf{T}_2^{(k)} = \left(\frac{\partial \phi_k}{\partial h_{ij}}\right)$ evaluated at the probability limits of (7.11). Then the condition in (7.13) implies

$$\text{(7.17)}\qquad c\,\mathbf{T}_1^{(k)} + \mathbf{T}_2^{(k)} = \mathbf{O}\ .$$

We write two $(1+G_2)\times(1+G_2)$ matrices $\mathbf{\Phi}_1$ $(=(\theta_{ij}))$ and $\mathbf{\Phi}_2$ as

$$\mathbf{\Phi}_1 = \begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{\Phi}_1^*(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) = \begin{bmatrix} \boldsymbol{\beta}_2^{'}\mathbf{\Phi}_1^*\boldsymbol{\beta}_2 & \boldsymbol{\beta}_2^{'}\mathbf{\Phi}_1^* \\ \mathbf{\Phi}_1^*\boldsymbol{\beta}_2 & \mathbf{\Phi}_1^* \end{bmatrix}$$

and

$$\mathbf{\Phi}_2 = \begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{\Phi}_2^*(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) = \begin{bmatrix} \boldsymbol{\beta}_2^{'}\mathbf{\Phi}_2^*\boldsymbol{\beta}_2 & \boldsymbol{\beta}_2^{'}\mathbf{\Phi}_2^* \\ \mathbf{\Phi}_2^*\boldsymbol{\beta}_2 & \mathbf{\Phi}_2^* \end{bmatrix}\ ,$$

where we denote $\mathbf{\Phi}_1^* = (\rho_{ml})$ $(m,l=2,\cdots,1+G_2)$, $(\mathbf{\Phi}_1^*\boldsymbol{\beta}_2)_l = \sum_{j=2}^{1+G_2}\beta_j\rho_{lj}$ $(l=2,\cdots,1+G_2)$, $(\boldsymbol{\beta}_2^{'}\mathbf{\Phi}_1^*)_m = \sum_{i=2}^{1+G_2}\beta_i\rho_{im}$ $(m=2,\cdots,1+G_2)$, and $\boldsymbol{\beta}_2^{'}\mathbf{\Phi}_1^*\boldsymbol{\beta}_2 = \sum_{i,j=2}^{1+G_2}\rho_{ij}\beta_i\beta_j$.

By differentiating each components of $\mathbf{\Phi}_1$ with respect to β_j $(j=1,\cdots,G_2)$, we have

$$\text{(7.18)}\qquad \frac{\partial \mathbf{\Phi}_1}{\partial \beta_j} = (\frac{\partial \theta_{lm}}{\partial \beta_j})\ ,$$

where $\frac{\partial \theta_{11}}{\partial \beta_j} = 2\sum_{i=2}^{1+G_2}\rho_{ji}\beta_i$ $(j=2,\cdots,1+G_2)$, $\frac{\partial \theta_{1m}}{\partial \beta_j} = \rho_{jm}$ $(m=2,\cdots,1+G_2)$, $\frac{\partial \theta_{l1}}{\partial \beta_j} = \rho_{lj}$ $(l=2,\cdots,1+G_2)$, and $\frac{\partial \theta_{lm}}{\partial \beta_j} = 0$ $(l,m=2,\cdots,1+G_2)$.

By using the same arguments to $\mathbf{\Phi}_2$, the condition (7.13) implies

$$\text{(7.19)}\qquad \mathbf{tr}\left[\mathbf{T}_1^{(k)}(\frac{\partial \mathbf{\Phi}_1}{\partial \beta_j} - c\,\frac{\partial \mathbf{\Phi}_2}{\partial \beta_j})\right] = \delta_j^k\ ,$$

where we define $\delta_k^k = 1$ and $\delta_j^k = 0$ $(k \neq j)$.

Define a $(1+G_2)\times(1+G_2)$ partitioned matrix

$$\text{(7.20)}\qquad \mathbf{T}_1^{(k)} = \begin{bmatrix} \tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)'} \\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix}\ .$$

Then (7.17) is represented as

$$2\tau_{11}^{(k)}[\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*]\boldsymbol{\beta} + 2[\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*]\boldsymbol{\tau}_2^{(k)} = \boldsymbol{\epsilon}_k \ , \tag{7.21}$$

where $\boldsymbol{\epsilon}_k^{'} = (0, \cdots, 0, 1, 0, \cdots, 0)$ with 1 in the k-th place and zeros in other elements. By the assumption that $\mathbf{\Phi}^* = \mathbf{\Phi}_1^* - c\ \mathbf{\Phi}_2^*$ is positive definite, we solve (7.19) as

$$\boldsymbol{\tau}_2^{(k)} = \frac{1}{2}\mathbf{\Phi}^{*-1}\boldsymbol{\epsilon}_k - \tau_{11}^{(k)}\boldsymbol{\beta}_2 \ . \tag{7.22}$$

Further by differentiating $\mathbf{\Phi}_1$ with respect to ρ_{ij}, we have

$$\frac{\partial \mathbf{\Phi}_1}{\partial \rho_{ii}} = (\frac{\partial \theta_{lm}}{\partial \rho_{ii}}) \ , \tag{7.23}$$

where $\frac{\partial \theta_{11}}{\partial \rho_{ii}} = \beta_i^2$, $\frac{\partial \theta_{1m}}{\partial \rho_{ii}} = \beta_i \ (m = i), 0 \ (m \neq i)$, $\frac{\partial \theta_{l1}}{\partial \rho_{ii}} = \beta_i \ (l = i), 0 \ (l \neq i)$ and $\frac{\partial \theta_{lm}}{\partial \rho_{ii}} = 1 \ (l = m = i), 0$ (otherwise).

For $i \neq j$

$$\frac{\partial \mathbf{\Phi}_1}{\partial \rho_{ij}} = (\frac{\partial \theta_{lm}}{\partial \rho_{ij}}) \ , \tag{7.24}$$

where $\frac{\partial \theta_{11}}{\partial \rho_{ij}} = 2\beta_i\beta_j$, $\frac{\partial \theta_{1m}}{\partial \rho_{ij}} = \beta_j \ (m = i), \beta_i \ (m = j), 0 \ (m \neq i, j)$, $\frac{\partial \theta_{l1}}{\partial \rho_{ij}} = \beta_j \ (l = i), \beta_i \ (l = j), 0 \ (l \neq i, j)$, and $\frac{\partial \theta_{lm}}{\partial \rho_{ij}} = 1 \ (l = i, m = j \ or \ l = j, m = i), 0$ (otherwise) for $(2 \leq l, m \leq 1 + G_2)$.

Then under the condition (3.13) we have the representation

$$\mathrm{tr}\left(\mathbf{T}_1^{(k)}\frac{\partial \mathbf{\Phi}_1 - c\mathbf{\Phi}_2}{\partial \rho_{ij}}\right) = \begin{cases} d[\beta_i^2\tau_{11}^{(k)} + 2\tau_{1i}^{(k)}\beta_i + \tau_{ii}^{(k)}] & (i = j) \\ \\ d[2\beta_i\beta_j\tau_{11}^{(k)} + 2\tau_{1j}^{(k)}\beta_i + 2\tau_{1i}^{(k)}\beta_j + 2\tau_{ij}^{(k)}] & (i \neq j) \end{cases} . \tag{7.25}$$

Under the assumption of *Theorem 3.3* we have the relation as

$$\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} + \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} + \mathbf{T}_{22}^{(k)} = \mathbf{O} \tag{7.26}$$

and then we have the representation

$$\begin{aligned} \mathbf{T}_{22}^{(k)} &= -\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} - \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} \\ &= \tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \frac{1}{2}\left[\mathbf{\Phi}^{*-1}\boldsymbol{\epsilon}_k\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\epsilon}_k^{'}\mathbf{\Phi}^{*-1}\right] \ . \end{aligned}$$

Let

$$\mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1 = \begin{bmatrix} s_{11} & \mathbf{s}_2^{'} \\ \mathbf{s}_2 & \mathbf{S}_{22} \end{bmatrix} \quad . \tag{7.27}$$

Since $\phi(\ \cdot\ )$ is differentiable and its first derivatives are bounded at the true parameters by assumption, the linearized estimator of β_k in the class of our concern can be represented as

$$\begin{aligned}
\sum_{g,h=1}^{1+G_2} \tau_{gh}^{(k)} s_{gh} &= \tau_{11}^{(k)} s_{11} + 2\boldsymbol{\tau}_2^{(k)'}\mathbf{s}_2 + \mathbf{tr}\left[\mathbf{T}_{22}^{(k)}\mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)} s_{11} + \left(\boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}^{*-1} - 2\tau_{11}^{(k)}\boldsymbol{\beta}_2^{'}\right)\mathbf{s}_2 + \mathbf{tr}\left[\left(\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \boldsymbol{\Phi}^{*-1}\boldsymbol{\epsilon}_k\boldsymbol{\beta}_2^{'}\right)\mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)}\left[s_{11} - 2\boldsymbol{\beta}_2^{'}\mathbf{s}_2 + \boldsymbol{\beta}_2^{'}\mathbf{S}_{22}\boldsymbol{\beta}_2\right] + \boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}^{*-1}(\mathbf{s}_2 - \mathbf{S}_{22}\boldsymbol{\beta}_2) \\
&= \tau_{11}^{(k)}\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta} + \boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}^{*-1}(\mathbf{s}_2, \mathbf{S}_{22})\boldsymbol{\beta} \quad .
\end{aligned}$$

Let

$$\boldsymbol{\tau}_{11} = \begin{bmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{bmatrix} \tag{7.28}$$

and we consider the asymptotic behavihor of the normalized estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ as

$$\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\right]\mathbf{S}\boldsymbol{\beta} \, . \tag{7.29}$$

Since the asymptotic variance-covariance matrix of $\mathbf{S}\boldsymbol{\beta}$ has been obtained by the proof of *Theorem 3.1*, we have

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

where $\boldsymbol{\Psi}^*$ has been given by *Theorem 3.1.*

This covariance matrix is the sum of a positive semi-definite matrix of rank 1 and a positive definite matrix. It has a minimum if

$$\boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta} \quad . \tag{7.30}$$

Hence we have completed the proof of *Theorem 3.3*.

Q.E.D.

APPENDIX : Some Figures

In Figures the cdf of alternative estimators for β and γ of (2.5) with (5.2) are shown in their standardized term, that is, (5.1). We give the distribution functions of the LIML, GMM, and the WG(within-group) estimators. For the comparative purpose we give the standard normal distribution as the bench mark for each case.

References

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

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

[3] 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.

[4] 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.

[5] Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol. 50, 1009-1027.

[6] Anderson, T.W., N. Kunitomo and K. Morimune (1986), "Comparing Single-equation Estimators in a Simultaneous Equation System," *Econometric Theory*, Vol. 2, 1-32.

[7] Anderson, T.W. and N. Kunitomo (2006), "On Asymptotic Properties of the LIML Estimator with Possibly Many Instruments," Unpublished Mnuscript.

[8] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations with Possibly Many Instruments," Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp(2005)).

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

[10] Fuller, W. (1977), "Some Properties of a Modification of the Limited Information Maximum Likelihood Estimaror," *Econometrica*, Vol. 45, 939-953.

[11] Hayakawa, K. (2006), "Efficient GMM Estimation of Dynamic Panel Data Models where Large Heterogeneity May be present," Unpublished Manuscript (Hitotsubashi University).

[12] Hsiao, Cheng (2003) *Analysis of Panel Data*, Cambridge University Press.

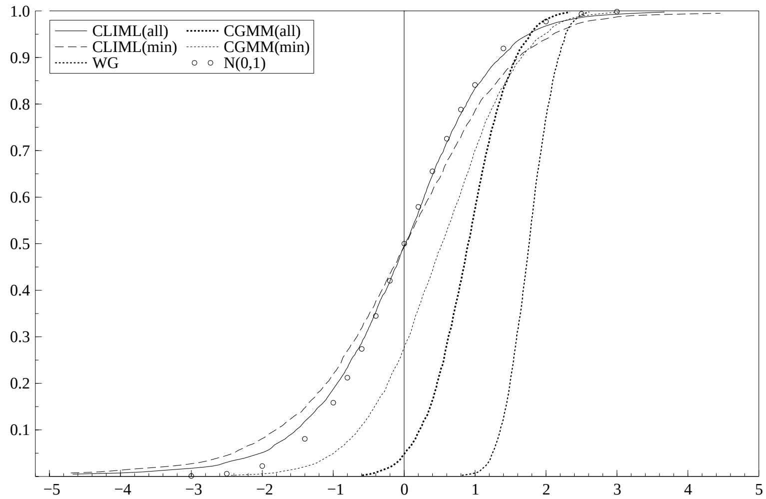

Figure 1: β : AR(5), N=100, T=10

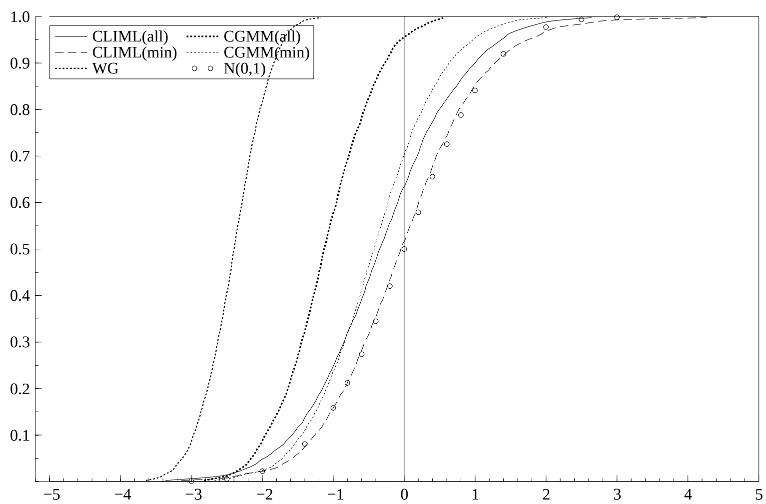

Figure 2: γ : AR(5), N=100, T=10

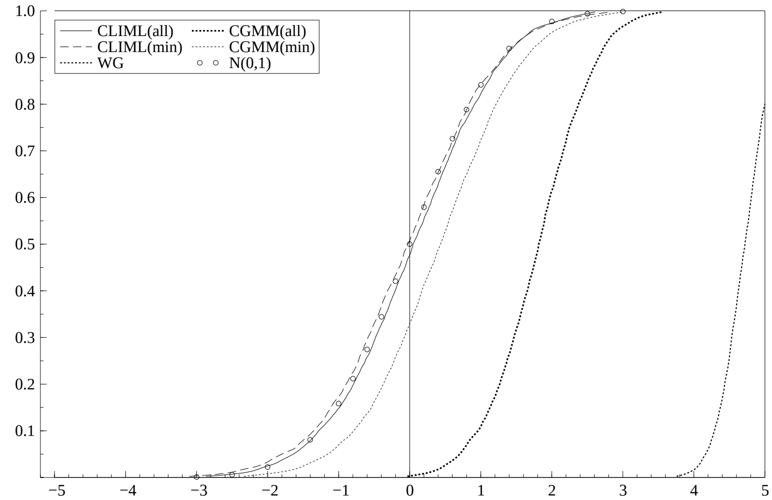

Figure 3: β : AR(5), N=200, T=17

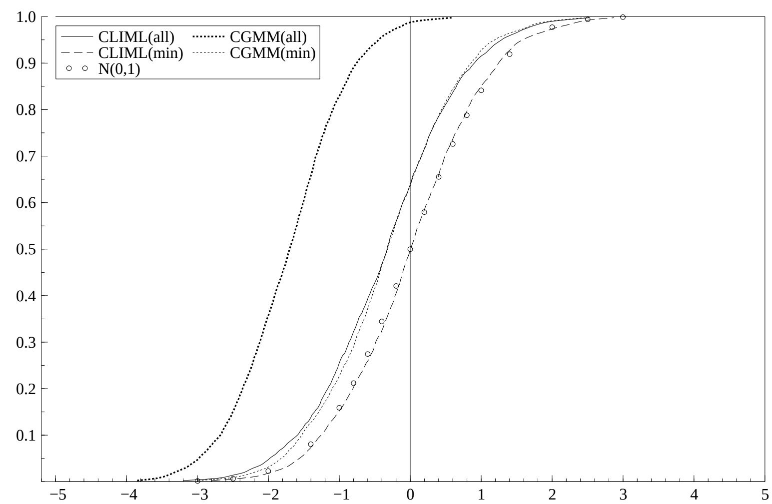

Figure 4: γ : AR(5), N=200, T=17

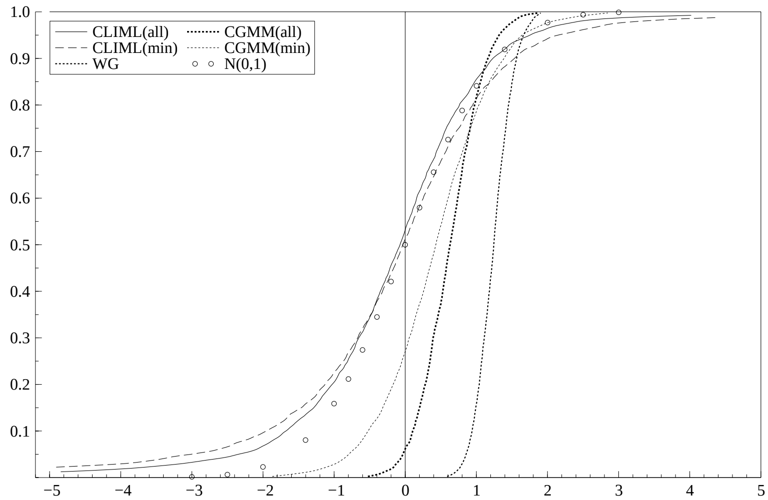

Figure 5: β : AR(2), N=100, T=10

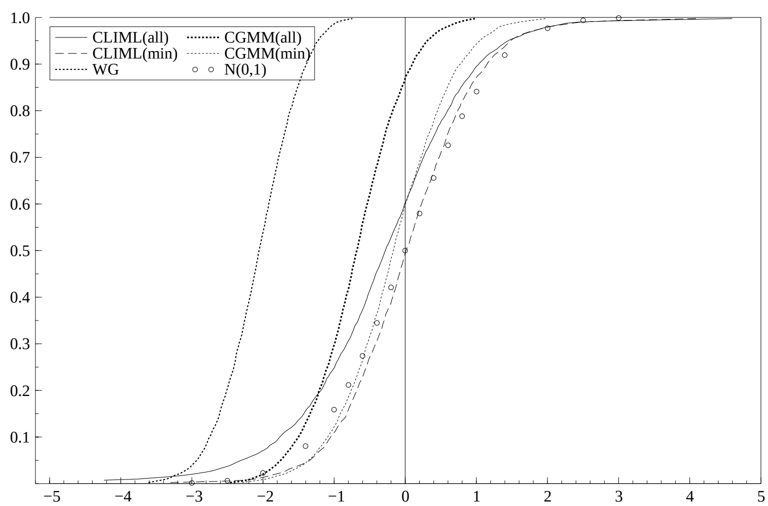

Figure 6: γ : AR(2), N=100, T=10

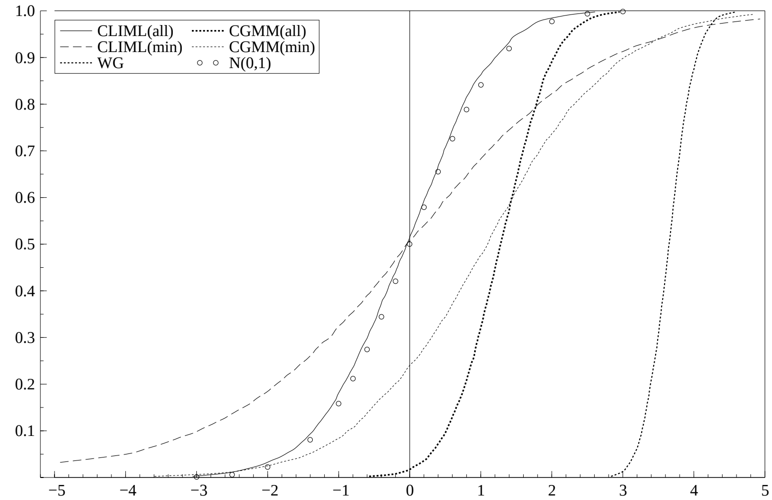

Figure 7: β : AR(2), N=200, T=17

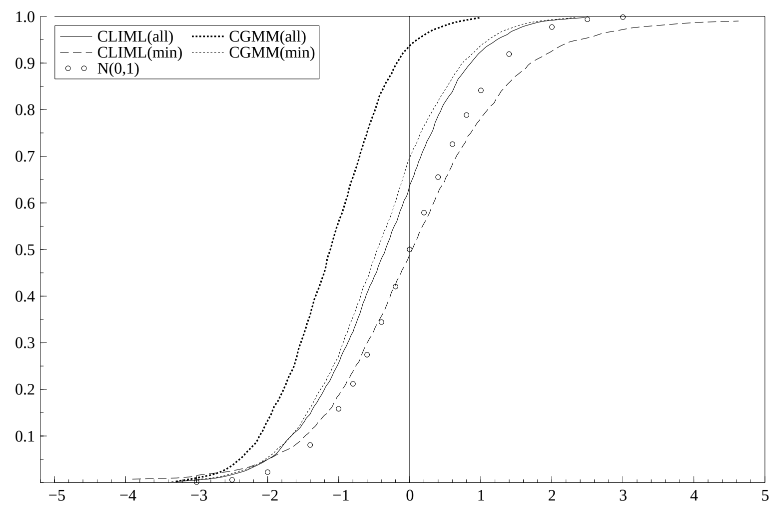

Figure 8: γ : AR(2), N=200, T=17