On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments *

T. W. Anderson †
Naoto Kunitomo ‡
and
Yukitoshi Matsushita §

December 17, 2007

Abstract

We consider the estimation of the coefficients of a linear structural equation in a simultaneous equation system when there are many instrumental variables. We derive some asymptotic properties of the limited information maximum likelihood (LIML) estimator when the number of instruments is large; some of these results are new and we relate them to results in some recent studies. We have found that the variance of the LIML estimator and its modifications often attain the asymptotic lower bound when the number of instruments is large and the disturbance terms are not necessarily normally distributed, that is, for the micro-econometric models with many instruments.

Key Words

Structural Equation, Simultaneous Equations System, Many Instruments, Limited Information Maximum Likelihood, Asymptotic Optimality

JEL Code: C13, C30

*This is the first part of a revised version of Discussion Paper CIRJE-F-321 under the title "A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models" (Graduate School of Economics, University of Tokyo, February 2005) which was presented at the Econometric Society World Congress 2005 at London (August 2005). We thank Yoichi Arai for some comments to the earlier version.

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

1. Introduction

Over the past three decades there has been increasing interest and research on the estimation of one structural equation in a system of simultaneous equations when the number of instruments (the number of exogenous variables excluded from the structural equation), say K_2, is large relative to the sample size, say n. The relevance of such models is due to collection of large data sets and the development of computational equipment capable of analysis of such data sets. One empirical example of this kind often cited in the econometric literature is Angrist and Krueger (1991) ; there has been some discussion by Bound et al. (1995) since then. Asymptotic distributions of estimators and test criteria are developed on the basis that both $K_2 \to \infty$ and $n \to \infty$. These asymptotic distributions are used as approximations to the distributions of the estimators and criteria when K_2 and n are large.

Bekker (1994) has written "To my knowledge a first mention of such a parameter sequence was made, with respect to the linear functional relationship model, in Anderson (1976 p.34). This work was extended to simultaneous equations by Kunitomo (1980) and Morimune (1983), who gave asymptotic expansions for the case of a single explanatory endogenous variable." Following Bekker there have been many studies of the behavior of estimators of the coefficients of a single equation when K_2 and n are large.

The main purpose of the present paper is to show that one estimator, the Limited Information Maximum Likelihood (LIML) estimator, has some optimum properties when K_2 and n are large. As background we state and derive some asymptotic distributions of the LIML and Two-Stage Least Squares (TSLS) estimators as $K_2 \to \infty$ and $n \to \infty$. Some of these results are improvements on Anderson (1976), Kunitomo (1980), Morimune (1983) and Bekker (1994), several of which are in the literature, and some results are new. They are presented in a uniform notation.

In addition to the LIML and TSLS estimators there are other instrumental variables (IV) methods. See Anderson, Kunitomo, and Sawa (1982) on the studies of their finite sample properties, for instance. Several semiparametric estimation methods have been developed including the generalized method of moments (GMM) estimation and the maximum empirical likelihood (MEL) method. (See Hayashi (2000) for instance.) However, it has been recently recognized that the classical methods have some advantages in microeconometric situations with many instruments. We call the case of many instruments *the large-K_2 asymptotic theory*.

In this paper we shall give the results on the asymptotic properties of the LIML estimator when the number of instruments is large. The TSLS and the GMM estimators lose even consistency in some of these situations. Our results on the asymptotic properties and optimality of the LIML estimator and its variants give new interpretations of the numerical information of the finite sample properties and some guidance on the use of alternative estimation methods in simultaneous equations and micro-econometric models with many instruments. There is a growing literature on the problem of many instruments in econometric models. We shall try to relate our results to some recent studies, including Donald and Newey (2001), Hahn (2002), Stock and Yogo (2005), Hansen, Hausman, and Newey (2004, 2006), Chao and Swanson (2005), and Bekker and Ploeg (2005).

In Section 2 we state the formulation of a simple linear structural model and the alternative estimation methods of unknown parameters in simultaneous equation models with possibly many instruments. Then in Section 3 we develop the large-K_2 asymptotics or the many instruments asymptotics and give some results on the asymptotic normality of the LIML estimator when n and K_2 are large. These results agree with the finite sample properties of alternative estimation methods and one application on t-ratios will be discussed. (We give a small number of figures and tables in Appendix. But the detail of the finite sample properties of the alternative estimators are discussed in Anderson, Kunitomo and Matsushita (2005), for instance.) In Section 4 we shall present new results on the asymptotic optimality of the LIML estimator and show that it often attains the lower bound of the asymptotic variance in a class of consistent estimators when the number of instruments is large. Also we shall discuss a more general formulation of the models and relate our results to some recent ones including Hansen et al. (2006) [1] in particular. Then brief concluding remarks will be given in Section 5. The proof of our theorems will be given in Section 6.

2. Alternative Estimation Methods in Structural Equation Models with Possibly Many Instruments

In Section 2 and Section 3 we consider the estimation problem of a structural equation in the classical linear simultaneous equations framework [2] . Let a single linear structural equation in an econometric model be

$$y_{1i} = \boldsymbol{\beta}_2^{'} \boldsymbol{y}_{2i} + \boldsymbol{\gamma}_1^{'} \boldsymbol{z}_{1i} + u_i \ \ (i = 1, \cdots, n), \tag{2.1}$$

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and a vector of G_2 endogenous variables, $\boldsymbol{z}_{1i}$ is a vector of K_1 (included) exogenous 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, and $u_1, \cdots, u_n$ are independent disturbance terms with $\mathcal{E}(u_i) = 0$ and $\mathcal{E}(u_i^2) = \sigma^2$ $(i = 1, \cdots, n)$. We assume that (2.1) is one equation in a system of $1 + G_2$ equations in $1 + G_2$ endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$. The reduced form of the model is

$$\mathbf{Y} = \mathbf{Z} \boldsymbol{\Pi}_n + \mathbf{V} \ , \tag{2.2}$$

where $\mathbf{Y} = (\mathbf{y}_i^{'})$ is the $n \times (1 + G_2)$ matrix of endogenous variables, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_{2n}) = (\mathbf{z}_i^{(n)'})$ is the $n \times K_n$ matrix of $K_1 + K_{2n}$ instrumental vectors $\mathbf{z}_i^{(n)} = (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{(n)'})^{'}$, $\mathbf{V} = (\mathbf{v}_i^{'})$ is the $n \times (1 + G_2)$ matrix of disturbances,

$$\boldsymbol{\Pi}_n = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21}^{(n)} & \boldsymbol{\Pi}_{22}^{(n)} \end{pmatrix}$$

---

[1] This is a revision of Hansen et al. (2004) after the second version of the present paper was written.

[2] We intentionally include the standard classic situation and state our results mainly because they are clear. Nonetheless a generalization of the formulation and the corresponding results will be discussed in Section 4.2.

is the $(K_1+K_{2n})\times(1+G_2)$ matrix of coefficients, and

$$\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \mathbf{\Omega} = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_2^{'} \\ \boldsymbol{\omega}_2 & \mathbf{\Omega}_{22} \end{bmatrix} .$$

The vector of K_n $(= K_1+K_{2n}, n > 2)$ instrumental variables $\mathbf{z}_i^{(n)}$ satisfies the orthogonality condition $\mathcal{E}[u_i \boldsymbol{z}_i^{(n)}] = \mathbf{0}$ $(i=1,\cdots,n)$. The relation between (2.1) and (2.2) is

$$\begin{pmatrix} \boldsymbol{\pi}_{11} & \mathbf{\Pi}_{12} \\ \boldsymbol{\pi}_{21}^{(n)} & \mathbf{\Pi}_{22}^{(n)} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \mathbf{0} \end{pmatrix} , \tag{2.3}$$

$u_i = (1, -\boldsymbol{\beta}_2^{'})\mathbf{v}_i = \boldsymbol{\beta}^{'}\mathbf{v}_i$, and

$$\sigma^2 = (1,-\boldsymbol{\beta}_2^{'}) \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_2^{'} \\ \boldsymbol{\omega}_2 & \mathbf{\Omega}_{22} \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_1 \end{bmatrix} = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} ,$$

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

Let $\mathbf{\Pi}_{2n} = (\boldsymbol{\pi}_{21}^{(n)}, \mathbf{\Pi}_{22}^{(n)})$ be a $K_{2n}\times(1+G_2)$ matrix of coefficients. Define the $(1+G_2)\times(1+G_2)$ matrices by

$$\mathbf{G} = \mathbf{Y}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y} = \mathbf{P}_2^{'}\mathbf{A}_{22.1}\mathbf{P}_2 , \tag{2.4}$$

and

$$\mathbf{H} = \mathbf{Y}^{'}\left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\right)\mathbf{Y} , \tag{2.5}$$

where $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}^{'}\mathbf{Z}_{2.1}$, $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{P_2} = \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}$,

$$\mathbf{Z}_1 = \begin{pmatrix} \mathbf{z}_{11}^{'} \\ \vdots \\ \mathbf{z}_{1n}^{'} \end{pmatrix} , \mathbf{Z}_{2n} = \begin{pmatrix} \mathbf{z}_{21}^{(n)'} \\ \vdots \\ \mathbf{z}_{2n}^{(n)'} \end{pmatrix} , \tag{2.6}$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{Z}_1^{'} \\ \mathbf{Z}_{2n}^{'} \end{pmatrix} (\mathbf{Z}_1, \mathbf{Z}_{2n}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{2.7}$$

is a nonsingular matrix (a.s.). Then the LIML estimator $\hat{\boldsymbol{\beta}}_{LI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.LI}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ is the solution of

$$(\frac{1}{n}\mathbf{G} - \frac{1}{q_n}\lambda_n\mathbf{H})\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0} , \tag{2.8}$$

where $q_n = n - K_n$ $(n > 2)$ and λ_n $(n > 2)$ is the smallest root of

$$|\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}| = 0 \ . \tag{2.9}$$

The solution to (2.8) minimizes the variance ratio

$$L_{1n} = \frac{[\sum_{i=1}^{n} \mathbf{z}_i^{(n)'}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})][\sum_{i=1}^{n} \mathbf{z}_i^{(n)}\mathbf{z}_i^{(n)'}]^{-1}[\sum_{i=1}^{n} \mathbf{z}_i^{(n)}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})]}{\sum_{i=1}^{n}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})^2} . \tag{2.10}$$

The TSLS estimator $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ is given by

$$\mathbf{Y}_2^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{pmatrix} = \mathbf{0} \ . \tag{2.11}$$

It minimizes the numerator of the variance ratio (2.10). The LIML and the TSLS estimators of $\boldsymbol{\gamma}_1$ are

$$\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Y}\hat{\boldsymbol{\beta}} \ , \tag{2.12}$$

where $\hat{\boldsymbol{\beta}}$ is $\hat{\boldsymbol{\beta}}_{LI}$ or $\hat{\boldsymbol{\beta}}_{TS}$, respectively. The LIML and TSLS estimators and their properties were originally developed by Anderson and Rubin (1949, 1950). See also Anderson (2005).

3 Asymptotic Properties of the LIML Estimator with Many Instruments

3.1 Asymptotic Normality of the LIML Estimator

We state the limiting distribution of the LIML estimator under a set of alternative assumptions when K_{2n} and $\boldsymbol{\Pi}_{2n}$ can depend on n and $n \to \infty$. We first consider the case when

$$\textbf{(I)} \qquad \frac{K_{2n}}{n} \longrightarrow c \ \ (0 \le c < 1),$$

$$\textbf{(II)} \qquad \frac{1}{n}\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}^{(n)} \stackrel{p}{\longrightarrow} \boldsymbol{\Phi}_{22.1} \ ,$$

where $\boldsymbol{\Phi}_{22.1}$ is a nonsingular constant matrix.

Condition (I) implies that the number of coefficient parameters is proportional to the number of observations. Because we want to estimate the covariance matrix of $\mathbf{v}_i^{(n)}$ $(i = 1, \cdots, n)$, we want $c < 1$. Then (I) implies $q_n \longrightarrow \infty$ as $n \longrightarrow \infty$. Condition (II) controls the noncentrality (or concentration) parameter to be proportional to the sample size. Since K_{2n} grows, it may be called the case of *many instruments*. These conditions define the maximal rates of growth of the number of incidental parameters.

We shall give our first result in *Theorem 1* and *Theorem 2*. Although the present formulation and *Theorem1* are similar to the corresponding results reported in Hansen et al. (2004) and Hasselt (2006), we shall give the proofs in Section 6 because the method of our proofs are relatively simple such that the underlying assumptions are clear and the method of proof can be extended easily to the more general cases as we shall discussed in Section 4.2 [3] .

To state our results conveniently we transform $\mathbf{v}_i$ to

$$
\begin{aligned}
\mathbf{w}_{2i} &= (\mathbf{0}, \mathbf{I}_{G_2})\left[\mathbf{I}_{1+G_2} - \frac{1}{\sigma^2}\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\right]\mathbf{v}_i \\
&= (\mathbf{0}, \mathbf{I}_{G_2})\left[\mathbf{v}_i - \frac{1}{\sigma^2}Cov(\mathbf{v}, u)u_i\right]
\end{aligned} \tag{3.1}
$$

and $u_i = \boldsymbol{\beta}^{'}\mathbf{v}_i$. Then $\mathcal{E}(\mathbf{w}_{2i}u_i) = \mathbf{0}$ and

$$
\mathcal{E}(\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}) = \frac{1}{\sigma^2}\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22} , \tag{3.2}
$$

where $[\,\cdot\,]_{22}$ is the $G_2 \times G_2$ lower right-hand corner of the matrix.

Theorem 1 : Let $\mathbf{z}_i^{(n)}, i = 1, 2, \cdots, n$, be a set of $K_n \times 1$ vectors ($K_n = K_1 + K_{2n}, n > 2$). Let $\mathbf{v}_i, i = 1, 2, \cdots, n$, be a set of $(1 + G_2) \times 1$ independent random vectors independent of $\mathbf{z}_1^{(n)}, \cdots, \mathbf{z}_n^{(n)}$ such that $\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \boldsymbol{\Omega}$ (a.s.), and the third- and fourth-order moments of $\mathbf{v}_i$ do not depend on i. Suppose that (I) and (II) hold. In addition assume

$$
\textbf{(III)} \qquad \frac{1}{n}\max_{1 \le i \le n} \|\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^{*}\|^2 \xrightarrow{p} 0 ,
$$

[3] A consequence of our method is that the proofs are self-contained.

where $\mathbf{z}_{in}^*$ is the i-th row vector of $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_{2n}$.

(i) For $c = 0$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^*) \ , \tag{3.3}$$

where $\boldsymbol{\Psi}^* = \sigma^2 \boldsymbol{\Phi}_{22.1}^{-1}$ and $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}$.

(ii) For $0 < c < 1$, define $\mathcal{E}(u^2\mathbf{w}_2\mathbf{w}_2^{'}) - \sigma^2\mathcal{E}(\mathbf{w}_2\mathbf{w}_2^{'}) = \boldsymbol{\Gamma}_{44.2}$. Suppose that $\mathcal{E}[\|\mathbf{v}_i\|^6]$ are bounded $(i = 1, \cdots, n)$ and there exist limits

$$\textbf{(IV)} \qquad \boldsymbol{\Xi}_{3.2} = \left[\frac{1}{1-c}\right] \operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{\Pi}_{22}^{(n)'} \sum_{i=1}^{n} \mathbf{z}_{in}^* \left[p_{ii}^{(n)} - c\right] \mathcal{E}(u^2\mathbf{w}_2^{'}) \ ,$$

$$\textbf{(V)} \qquad \eta = \left[\frac{1}{1-c}\right]^2 \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[p_{ii}^{(n)} - c\right]^2 \ ,$$

where $p_{ii}^{(n)} = (\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'})_{ii}$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{**}) \ , \tag{3.4}$$

where

$$\begin{aligned} \boldsymbol{\Psi}^{**} &= \sigma^2 \boldsymbol{\Phi}_{22.1}^{-1} + \boldsymbol{\Phi}_{22.1}^{-1} \left\{ c_* \left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22} \right. \\ &\qquad \left. + \left[(\boldsymbol{\Xi}_{3.2} + \boldsymbol{\Xi}_{3.2}^{'}) + \eta\boldsymbol{\Gamma}_{44.2}\right] \right\} \boldsymbol{\Phi}_{22.1}^{-1} \end{aligned} \tag{3.5}$$

and $c_* = c/(1-c)$. If $G_2 = 1$, then $[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}]_{22} = \omega_{11}\omega_{22} - \omega_{12}^2 = |\boldsymbol{\Omega}|$.

Corollary 1 : When $\mathbf{v}_i$ $(= (v_{ji}))$ $(i = 1, \cdots, n; j = 1, \cdots, G_2 + 1)$ has an elliptically contoured (EC) distribution [4] in *Theorem 1*, the fourth order moments $\mathcal{E}(v_{ji}v_{ki}v_{li}v_{mi}) = (1 + \kappa/3)(\omega_{jk}\omega_{lm} + \omega_{jl}\omega_{km} + \omega_{jm}\omega_{kl})$ and $\mathcal{E}[(\beta^{'}\mathbf{v}_i)^2\mathbf{v}_i\mathbf{v}_i^{'}] = (1 + \kappa/3)(\sigma^2\boldsymbol{\Omega} + 2\boldsymbol{\Omega}\beta\beta^{'}\boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = (\omega_{jk})$, $\mathcal{E}(v_{ji}v_{ki}) = \omega_{jk}$ and κ is the kurtosis

[4] The precise definition of elliptically contoured (EC) distribution has been given by Section 2.7 of Anderson (2003). The standardized fourth order cumulant of any linear combination of the random vector $\mathbf{X}$ followed by the EC distribution, say, $\gamma\mathbf{X}$, is a constant

$$\kappa = \frac{\mathcal{E}[\gamma^{'}(\mathbf{X} - \mu)]^4}{(\mathcal{E}[\gamma^{'}(\mathbf{X} - \mu)]^2)^2} - 3 \ ,$$

independent of γ and is known as the *kurtosis* of $\gamma^{'}\mathbf{X}$. The multivariate normal distribution is a member of the EC class; the kurtosis of any normal distribution is 0.

of EC($\mathbf{\Omega}$). Then $\mathbf{\Gamma}_{44.2} = (\kappa/3)\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]_{22}$ and (3.5) is given by

$$\text{(3.6)} \qquad \mathbf{\Psi}^{**} = \sigma^2\mathbf{\Phi}_{22.1}^{-1} + (c_* + \frac{1}{3}\eta\kappa)\mathbf{\Phi}_{22.1}^{-1}\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]_{22}\mathbf{\Phi}_{22.1}^{-1} \, .$$

Instead of making an assumption on the distribution of disturbance terms except the existence of their moments, alternatively we assume

$$(\mathbf{VI}) \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)} - c\right]^2 = 0 \, ,$$

where $p_{ii}^{(n)} = (\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}')_{ii}$. The typical example for Condition (VI) is the case when we have dummy variables which have 1 or -1 in their all components so that $(1/n)\mathbf{A}_{22.1} = \mathbf{I}_{K_{2n}}$ and $p_{ii}^{(n)} = K_{2n}/n \; (i = 1, \cdots, n)$.

Condition (VI) is the same as $\eta = 0$ in Condition (V), which in turn implies $\mathbf{\Xi}_{3.2} = \mathbf{O}$ in Condition (IV) by the Cauchy-Schwarz inequality using Conditions (II) and (VI). These consequences of Condition (VI) imply the following theorem :

Theorem 2 : For $0 \leq c < 1$ assume Conditions (I), (II), (III), (VI) and assume that $\mathcal{E}[\|\mathbf{v}_i\|^6] \; (i = 1, \cdots, n)$ is bounded. Then

$$\text{(3.7)} \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^{**}) \, ,$$

where

$$\text{(3.8)} \qquad \mathbf{\Psi}^{**} = \sigma^2\mathbf{\Phi}_{22.1}^{-1} + c_*\mathbf{\Phi}_{22.1}^{-1}\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]_{22}\mathbf{\Phi}_{22.1}^{-1}$$

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

Corollary 2 : Suppose $\mathbf{z}_1^{(n)}, \cdots, \mathbf{z}_n^{(n)}$ are independently distributed with $\mathcal{E}(\mathbf{z}_i^{(n)}\mathbf{z}_i^{(n)'}) = \mathbf{M} = (m_{ij})$, $(1/n)a_{ij} \xrightarrow{p} m_{ij}$ $(a_{ij} = (\mathbf{A})_{ij})$ and $Var[\mathbf{z}_i^{(n)'}\mathbf{M}^{-1}\mathbf{z}_i^{(n)}] = o(K_n^2)$. Then (3.7) and (3.8) hold without Condition (VI) in *Theorem 2.*

The asymptotic properties of the LIML estimator hold when K_{2n} increases as $n \to \infty$ and $K_{2n}/n \to 0$. In this case the limiting distribution of the LIML estimator can be different from that of the TSLS estimator. (The proof of Theorem 3 will be given in Section 6.)

Theorem 3 : Let $\{\mathbf{v}_i, \mathbf{z}_i^{(n)}; i = 1, \cdots, n\}$ be a set of independent random vectors. Assume that (2.1) and (2.2) hold with $\mathcal{E}(\mathbf{v}_i|\mathbf{z}_i) = \mathbf{0}$ $(a.s.)$ and $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}|\mathbf{z}_i^{(n)}) = \mathbf{\Omega}_i^{(n)}$ $(a.s.)$ is a function of $\mathbf{z}_i^{(n)}$, say, $\mathbf{\Omega}_i[n, \mathbf{z}_i^{(n)}]$. The further assumptions on $(\mathbf{v}_i, \mathbf{z}_i^{(n)})$ $(\mathbf{v}_i = (v_{ji}))$ are that $\mathcal{E}(v_{ji}^4|\mathbf{z}_i^{(n)})$ are bounded, there exists a constant matrix $\mathbf{\Omega}$ such that $\sqrt{n}\|\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}\|$ is bounded and $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} > 0$. Suppose

$$
\begin{aligned}
&(\mathbf{I}^{'}) && \frac{K_{2n}}{n^{\eta}} \longrightarrow c \ (0 \leq \eta < 1, \ 0 < c < \infty), \\
&(\mathbf{II}) && \frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}^{(n)} \xrightarrow{p} \mathbf{\Phi}_{22.1} \ , \\
&(\mathbf{III}) && \frac{1}{n}\max_{1 \leq i \leq n} \|\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^{*}\|^2 \xrightarrow{p} 0 \ ,
\end{aligned}
$$

where $\mathbf{\Phi}_{22.1}$ is a nonsingular constant matrix and $\mathbf{z}_{in}^{*'}$ is the i-th row vector of $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_{2n}$.

(i) Then for the LIML estimator when $0 \leq \eta < 1$,

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1}^{-1}) \ , \tag{3.9}
$$

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

(ii) For the TSLS estimator when $1/2 < \eta < 1$,

$$
n^{1-\eta}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{p} \mathbf{\Phi}_{22.1}^{-1} \ c(\boldsymbol{\omega}_{21}, \mathbf{\Omega}_{22})\boldsymbol{\beta} \ , \tag{3.10}
$$

when $\eta = 1/2$,

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{d} N\left[c\mathbf{\Phi}_{22.1}^{-1}(\boldsymbol{\omega}_{21}, \mathbf{\Omega}_{22})\boldsymbol{\beta}, \sigma^2\mathbf{\Phi}_{22.1}^{-1}\right] \ , \tag{3.11}
$$

where $(\omega_{21}, \mathbf{\Omega}_{22})$ is the $G_2 \times (1 + G_2)$ lower submatrix of $\mathbf{\Omega}$. When $0 \leq \eta < 1/2$,

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1}^{-1}) \ . \tag{3.12}
$$

It is possible to interpret the standard large sample theory as a special case of *Theorem 3.* The asymptotic property of the LIML and TSLS estimators for $\boldsymbol{\gamma}_1$ can be derived from *Theorem 1.* Donald and Newey (2001) (in their *Lemma A.6*) has investigated the asymptotic properties of the LIML estimator when $K_{2n}/n \longrightarrow 0$. Also Stock and Yogo (2005), and Hansen et al. (2004) have discussed the asymptotic properties of the GMM estimators in some cases of the large-K_2 theory when $0 < \eta < 1/2$. In this case, however, the asymptotic lower bound of the covariance matrix is the same as in the case of the large sample asymptotic theory as we shall see in Section 4.

3.2 On the Asymptotic Variance and t-Ratios

There is a notable difference between the results in *Theorem 1* and *Theorem 2*, that is, the asymptotic variance depends on the 3rd and 4th order moments of the disturbance terms in the former. The finite sample properties of the LIML estimator have been investigated by Anderson, Kunitomo and Matsushita (2005) in a systematic way and as typical examples we present only eight figures (Figures 1A-8A) in Appendix when $\alpha = 0.5, 1.0$ and $G_2 = 1$. We have used the numerical estimation of the cumulative distribution function (cdf) of the LIML estimator based on the simulation and we have enough numerical accuracy in most cases. See Anderson et al. (2005) for the details of the numerical computation method. The key parameters in figures and tables are K_2 (or K_{2n}), $n - K$ (or $n - K_n$), $\alpha = [\omega_{22}/|\mathbf{\Omega}|^{1/2}](\beta_2 - \omega_{12}/\omega_{22})$ ($\mathbf{\Omega} = (\omega_{ij})$) and $\delta^2 = \mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}^{(n)}/\omega_{22}$. See Anderson et al. (1982) for the details of these notations.

The figures (Figures 1A-8A) show the estimated cdf of the LIML estimator in the standard form, that is,

$$\frac{\sqrt{n}}{\sigma}\mathbf{\Phi}_{22.1}^{1/2}\left(\hat{\beta}_{2.LI} - \beta_2\right) . \tag{3.13}$$

By using (3.3) the limiting distributions of the LIML estimators are $N(0,1)$ in the large sample asymptotics and they are denoted by "o". By using (3.8) the corresponding limiting distributions of the LIML estimators in the large K_2 asymptotics are $N(0,a)$ ($a = \mathbf{\Psi}^{*-1}\mathbf{\Psi}^{**}, a \geq 1$), which are denoted by *large-K-normal* in Figures 1A-8A, and they are traced by the dashed curves. In Figures 3A-4A and 7A-8A we also have the approximations based on the variance formula (3.5) with the third and fourth order moments of disturbance terms, which are denoted by *large-K-nonnormal* and traced by "x".

From these figures we have found that the effects of many instruments on the cdf of the LIML estimator are significant and the approximations based on the large sample asymptotics are often inferior. At the same time we also have found that the effects of non-normality of disturbance terms on the cdf of the LIML estimator are often very small. (The dashed curves and x are almost identical.)

One important application of the asymptotic variance is to construct a t-ratio for testing a hypothesis on the coefficients. We can use the asymptotic variance of the LIML estimator given by (3.5) or (3.8) replaced by its estimator. (We have used $\mathbf{P}_2$ for $\mathbf{\Pi}_{2n}$, $(1/q_n)\mathbf{H}$ for $\mathbf{\Omega}$ and the sample moments from residuals for σ^2 and $\mathcal{E}(u^2\mathbf{w}_2)$, for instance.) We have investigated this problem and as typical examples we give four tables (Tables 1B-4B) on the cdfs of t-ratios

$$t(\hat{\beta}_{2.LI}) = \frac{\sqrt{n}(\hat{\beta}_{2.LI} - \beta_2)}{s(\hat{\beta}_{2.LI})} , \tag{3.14}$$

which is constructed by the LIML estimation, where $s^2(\hat{\beta}_{2.LI})$ is the estimator of the variance. The formulas (3.3),(3.5), (3.6) and (3.8) are used. (Matsushita (2006) has investigated the finite sample properties of alternative t-ratios in detail and derived their asymptotic expansions of their distribution functions.) From these tables we have found that the effect of many instruments on the cdf of the null distributions of t-ratios are often significant. The approximations based on the large sample asymptotics are often inferior. At the same time we also have found that the effects of non-normality of disturbance terms on the null-distributions of the t-ratios are often small, that is, the differences between the effects of (3.5) in *Theorem 1* and (3.8) in *Theorem 2* are small for practical purposes.

Bekker (1994) derived the asymptotic variance formula (3.8) for the LIML estimator under the condition that the disturbance terms are normally distributed. It is identical to the asymptotic covariance matrix of the LIML estimator in the large-K_2 asymptotics reported by Kunitomo (1982). From our investigations it may be advisable to use (3.8) for statistical inferences on the structural coefficients even under the cases when the disturbances are not normally distributed for practical purposes.

4 Asymptotic Optimality of the LIML Estimator

4.1 Main Result

For the estimation of the vector of structural parameters $\boldsymbol{\beta}$, it seems natural to consider procedures based on two $(1+G_2)\times(1+G_2)$ matrices $\mathbf{G}$ and $\mathbf{H}$. We shall consider a class of estimators which are functions of these matrices. The typical examples of this class are the OLS estimator, the TSLS estimator, and the modified versions of the LIML estimator including the one proposed by Fuller (1977). Then we have a new result on the asymptotic optimality of the LIML estimator. We shall discuss the modified version of LIML estimator which attains the lower bound of the asymptotic covariance under alternative assumptions in Section 4.2. The proof of *Theorem 4* will be given in Section 6.

Theorem 4 : Assume that (2.1) and (2.2) hold. Define a class of consistent estimators for $\boldsymbol{\beta}_2$ by

(4.1) $$\hat{\boldsymbol{\beta}}_2 = \phi(\frac{1}{n}\mathbf{G}, \frac{1}{q_n}\mathbf{H}) ,$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of $\mathbf{G}$ and $\mathbf{H}$ as $K_{2n} \to \infty$ and $n \to \infty$ and $0 \le c < 1$. Then under the assumptions of the case (i) of *Theorem 1*, *Corollary 1* or *Theorem 2*,

(4.2) $$\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}) ,$$

where

(4.3) $$\boldsymbol{\Psi} \ge \boldsymbol{\Psi}^* \text{ (or } \boldsymbol{\Psi}^{**}) ,$$

and $\boldsymbol{\Psi}^*$ (or $\boldsymbol{\Psi}^{**}$) is given by (3.3), (3.6) or (3.8), respectively.

When the distribution of $\mathbf{V}$ is normal $N(\mathbf{0}, \boldsymbol{\Omega})$ and $\mathbf{Z}$ is exogenous, $\mathbf{P} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ and $\mathbf{H} = \mathbf{Y}'[\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}$ are a sufficient set of statistics for $\boldsymbol{\Pi}_n$ and $\boldsymbol{\Omega}$, the parameters of a model. When K_n is fixed, it is known that of all consistent estimators of $\boldsymbol{\beta}_2$ the LIML estimator suitably normalized has the minimum asymptotic variance and the optimality of $\hat{\boldsymbol{\beta}}_{2.LI}$ extends to the class of all consistent estimators including the MEL estimator (provided that it is consistent), not only the form of (4.1) in this case. When K_n is dependent on n, however, there is a further problem with many incidental parameters.

The above theorems are the generalized versions of the results given by Kunitomo (1982) and Theorem 3.1 of Kunitomo (1987). Furthermore, Kunitomo (1987) has investigated the higher order efficiency property of the LIML estimator when $G_2 = 1$, $0 \leq c < 1$ and the disturbances are normally distributed. Chao and Swanson (2005) recently have investigated the consistency issue of instrumental variables methods when K_{2n} is dependent on n and the disturbances are not necessarily normally distributed. In the large-K_2 asymptotic theory with $0 < c < 1$, the LIML estimator is asymptotically efficient and attains the lower bound of the variance-covariance matrix, which is strictly larger than the information matrix and the asymptotic Cramér-Rao lower bound under a set of assumptions, while both the TSLS and the GMM estimators are inconsistent. This is a non-regular situation because the number of incidental parameters increases as K_{2n} increases in the simultaneous equation models [5] .

4.2 A general formulation of the asymptotic optimality

We shall consider a model with the single structural equation (2.1) and a nonlinear replacement for the last G_2 columns of the reduced form (2.2). We treat (2.1) and

$$\mathbf{Y}_2 = \mathbf{\Pi}_2^{(n)}(\mathbf{Z}) + \mathbf{V}_2 \ , \tag{4.4}$$

where $\mathbf{\Pi}_2^{(n)}(\mathbf{Z}) = (\boldsymbol{\pi}_{2i}^{'}(\mathbf{z}_i^{(n)}))$ is an $n \times G_2$ matrix, the i-th row of which $\boldsymbol{\pi}_{2i}^{'}(\mathbf{z}_i^{(n)})$ depends on the $K_n \times 1$ vector $\mathbf{z}_i^{(n)}$ $(i = 1, \cdots, n)$, $\mathbf{V}_2$ is a $n \times G_2$ matrix, $\mathbf{v}_1 = \mathbf{u} + \mathbf{V}_2\boldsymbol{\beta}_2$, and $\mathbf{V} = (\mathbf{v}_1, \mathbf{V}_2)$. When the reduced form equations (4.4) are linear, (2.1) and (4.4) has a representation (2.2). In this formulation, Condition (II) is

[5] As a non-trivial example, we take the bias-adjusted TSLS estimator by setting $\lambda_n = K_{2n}/n$ in (2.8) and denote $\hat{\boldsymbol{\beta}}_{2.BTS}$. Then the asymptotic variance of $\hat{\boldsymbol{\beta}}_{2.BTS}$ is greater than $\mathbf{\Phi}^*$ in *Theorem 3* if $0 < c < 1$ and $\left[\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{\Omega}\right]_{22} \geq 0$.

replaced by

$$(\mathbf{II}') \qquad \frac{1}{d_n^2}\mathbf{\Pi}_2^{(n)}(\mathbf{Z})'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{\Pi}_2^{(n)}(\mathbf{Z}) \stackrel{p}{\longrightarrow} \mathbf{\Phi}_{22.1} \ ,$$

where $\mathbf{\Phi}_{22.1}$ is a positive (constant) definite matrix and $d_n \stackrel{p}{\to} \infty$ as $n \to \infty$. We replace (III) by

$$(\mathbf{III})' \qquad \frac{1}{d_n^2}\max_{1\le i\le n}\|\boldsymbol{\pi}_{2i}(\mathbf{z}_i^{(n)})\|^2 \stackrel{p}{\longrightarrow} 0 \ .$$

A possible additional condition (due to nonlinearilty in (4.4)) is

$$(\mathbf{VIII}) \qquad \frac{1}{q_n}\mathbf{\Pi}_2^{(n)}(\mathbf{Z})'\left[\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right]\mathbf{\Pi}_2^{(n)}(\mathbf{Z}) \stackrel{p}{\longrightarrow} \mathbf{O} \ .$$

Condition (VIII) is automatically satisfied in the linear case. It is possible to weaken this condition to some extent with more complications of the resulting analysis.

Three cases are considered. We have already investigated the first case of $d_n = O_p(n^{1/2})$ and $K_{2n} = O(n)$ in Section 3. The asymptotic covariance of the LIML estimator is given by (3.5) in *Theorem 1* or (3.8) in *Theorem 2* under alternative assumptions with $(II)'$ instead of (II).

The second case is the standard large sample asymptotics, which corresponds to the cases of $d_n = O_p(n^{1/2+\delta})$ $(\delta > 0)$, or $d_n = O_p(n^{1/2})$ and $K_{2n}/n = o(1)$. In this case

$$(4.5) \qquad d_n(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1}^{-1}) \ .$$

Theorem 3 is one result in this case, which can be extended directly to the nonlinear model of (2.1) and (4.4).

The third case occurs when $d_n = o_p(n^{1/2})$ and $\sqrt{n}/d_n^2 \to 0$, which may correspond to one case in Hansen et al. (2006) with slightly different normalization and assumptions. In this case

$$(4.6) \qquad \left[\frac{d_n^2}{\sqrt{n}}\right](\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \mathbf{\Psi}^{***})$$

where

$$(4.7) \qquad \mathbf{\Psi}^{***} = \mathbf{\Phi}_{22.1}^{-1}\left\{c_*\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]_{22} + \eta\mathbf{\Gamma}_{44.2}\right\}\mathbf{\Phi}_{22.1}^{-1} \ .$$

The variance (4.7) is simpler than (3.5) because the effects of n dominate the first, the third and the fourth terms of (3.5) in *Theorem 1*. A simple derivation of the asymptotic normality of the LIML estimator will be provided in Section 6 as an illustration.

We now turn to consider the asymptotic optimality of the LIML estimator in the second case ($d_n = n^{1/2}$). In this paper we have focused on the class of estimators in the form of (4.1). We use the proof of *Theorem 4* in Section 6. (See the arguments around (6.53).) For any normalized consistent estimator (4.1) define $\mathbf{e} = \sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$. Then $\mathbf{e} - \hat{\mathbf{e}}^* = o_p(1)$ and

$$\hat{\mathbf{e}}^* = \boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\mathbf{S}\boldsymbol{\beta} \ , \tag{4.8}$$

where $\boldsymbol{\tau}_{11} = (\frac{\partial \phi_k}{\partial g_{11}})$ is a $G_2 \times 1$ vector evaluated at the true vector of parameters, $\mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1$, $\mathbf{G}_1 = \sqrt{n}[(1/n)\mathbf{G} - \text{plim}(1/n)\mathbf{G}]$ and $\mathbf{H}_1 = \sqrt{q_n}[(1/q_n)\mathbf{H} - \text{plim}(1/q_n)\mathbf{H}]$.

Define

$$\hat{\mathbf{e}}_{LI}^* = (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})(\mathbf{I}_{G_2+1} - \frac{1}{\sigma^2}\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'})\mathbf{S}\boldsymbol{\beta} \ . \tag{4.9}$$

Then

$$\hat{\mathbf{e}}^* = \left[\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta}\right]\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta} + \hat{\mathbf{e}}_{LI}^* \ .$$

We notice that $\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta} = \sigma^2\lambda_{1n} + o_p(1)$ in (6.7) and $\lambda_{1n} = \sqrt{n}(\lambda_n - c)$, which is the stochastic part of the smallest characteristic root in the LIML estimation. Then if λ_{1n} and $\hat{\mathbf{e}}_{LI}^*$ are asymptotically uncorrelated, the LIML estimator attains the lower bound of the asymptotic variance. A set of sufficient conditions is either the moment conditions in *Corollary 1* or Condition (VI) in *Theorem 2*. In the more general case of (2.1) and the nonlinear equations (4.4), we summarize our result in *Theorem 5*.

Theorem 5 : For the model of (2.1) and (4.4), assume (I), (II)$'$ and (VIII) with $d_n = n^{1/2}$. Define the class of consistent estimators for $\boldsymbol{\beta}_2$ by (4.1), where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices in (2.4) and (2.5) as $K_{2n} \to \infty$ and $n \to \infty$. Then the lower bound of the asymptotic variance can be attained if and only if

$$\boldsymbol{\tau}_{11} + \frac{1}{\mathcal{E}(\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta})}\mathcal{E}\left[\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta}\left(\hat{\mathbf{e}}_{LI}^* + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\frac{\boldsymbol{\Omega}\boldsymbol{\beta}}{\sigma^2}\right)\right] = o(1) \ . \tag{4.10}$$

From this result we find that it is possible to modify the LIML estimator which can attain the lower bound of the asymptotic variance and we can derive an explicit representation. However, since it depends on the 3rd and 4th order moments of the disturbances in the general case, it is rather complicated and its practical value may be limited. On the other hand, if $(\mathbf{0}, \mathbf{I}_{G_2})\mathcal{E}\left[(\mathbf{I}_{G_2+1} - \frac{1}{\sigma^2}\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}')\mathbf{S}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}\right] = \mathbf{0}$, we can obtain the *key condition*

$$\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\frac{\boldsymbol{\Omega}\boldsymbol{\beta}}{\sigma^2} = o(1) \ . \tag{4.11}$$

Thus the LIML estimator and its variants including the one by Fuller (1977) satisfy (4.11) if we have alternative conditions in *Corollary 1* or *Theorem 2.* Also we find that the TSLS estimator, the GMM estimator and their variants cannot satisfy (4.11) in the first case when $c > 0$.

Although *Theorem 5* formally covers the (first order) asymptotic optimality for the first and the second cases of the parameter sequences on d_n and K_{2n}, it is immediate to extend the result to the third case, but we need additional notations. It is because (4.7) could be regarded as a special case of (3.5) without the first, third and the fourth terms except the normalizations.

4.3 Heteroscedasticity and the asymptotic properties

Recently, there have been some interests on the role of heteroscedasticity with many instruments. Let $\boldsymbol{\Omega}_i = \mathcal{E}(\mathbf{v}_i\mathbf{v}_i'|\mathbf{z}_i^{(n)})$ be the conditional covariance matrix and we assume

$$\textbf{(IX)} \qquad \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Omega}_i \xrightarrow{p} \boldsymbol{\Omega} \ ,$$

where $\boldsymbol{\Omega}$ is a positive definite (constant) matrix. Then in the case when both Conditions (VI) and (IX) hold, the LIML estimator has still some desirable asymptotic properties.

In the more general cases, the distribution of the LIML estimator could be significantly affected by the presence of heteroscedasticity of disturbance terms with many instruments. On this issue, however, there are alternative ways to improve the LIML estimation. The detail of this problem shall be discussed in an another occasion.

5. Concluding Remarks

In this paper, we have developed the large K_2-asymptotic theory when the number of instruments is large in a structural equation of the simultaneous equations system. Although the limited information maximum likelihood (LIML) estimator and the two stage-least squares (TSLS) estimator are asymptotically equivalent in the standard large sample theory, they are asymptotically quite different in the large-K_2 asymptotics. In some recent microeconometric models and models on panel data, it is often a common feature that K_2 is fairly large and this asymptotic theory has some practical relevance. We have shown that the LIML estimator and its variants may have the asymptotic optimality in the large K_2-asymptotics sense. We have given a set of sufficient conditions for the asymptotic normality and the (first order) asymptotic efficiency of the LIML estimator.

As we have suggested in Section 3.2 briefly and in Anderson, Kunitomo and Matsushita (2005) (or Part II of our study), our asymptotic results in this paper (which is Part I of our study) shall give some further reasons why we have the finite sample properties of the alternative estimation methods including the classical LIML and the TSLS estimators, and also the semi-parametric estimation methods of the generalized method of moments (GMM) and the empirical likelihood (EL) estimators.

6 Proof of Theorems

In this section we give the proofs of *Theorems* and the mathematical derivation in Sections 3 and 4.

Proof of Theorem 1 :

Substitution of (2.2) into (2.4) yields

$$\begin{aligned}\mathbf{G} &= (\mathbf{\Pi}_n^{'}\mathbf{Z}^{'} + \mathbf{V}^{'})\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}(\mathbf{Z}\mathbf{\Pi}_n + \mathbf{V}) \\ &= \mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n} + \mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} + \mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V} + \mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n} \; .\end{aligned}$$

Then

$$\begin{aligned}(6.1) \qquad & \mathbf{G} - [\mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n} + K_{2n}\mathbf{\Omega}] \\ = \;& \mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V} + \mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n} + \left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n}\mathbf{\Omega}\right] \; .\end{aligned}$$

By Assumption (II) implies that as $n \longrightarrow \infty$

$$(6.2) \qquad \frac{1}{n}\mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V} \xrightarrow{p} \mathbf{O} \; ,$$

and

$$(6.3) \qquad \frac{1}{n}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n}\mathbf{\Omega}\right] \xrightarrow{p} \mathbf{O} \; .$$

Then as $n \longrightarrow \infty$,

$$(6.4) \qquad \frac{1}{n}\mathbf{G} \xrightarrow{p} \mathbf{G}_0 = \begin{bmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) + c\,\mathbf{\Omega}$$

and

$$(6.5) \qquad \frac{1}{q_n}\mathbf{H} \xrightarrow{p} \mathbf{\Omega} \; .$$

Then $\hat{\boldsymbol{\beta}}_{LI} \xrightarrow{p} \boldsymbol{\beta}$ and $\lambda_n \xrightarrow{p} c$ as $n \to \infty$.

Define $\mathbf{G}_1$, $\mathbf{H}_1$, λ_{1n}, and $\mathbf{b}_1$ by $\mathbf{G}_1 = \sqrt{n}(\frac{1}{n}\mathbf{G} - \mathbf{G}_0)$, $\mathbf{H}_1 = \sqrt{q_n}(\frac{1}{q_n}\mathbf{H} - \mathbf{\Omega})$, $\lambda_{1n} = \sqrt{n}(\lambda_n - c)$, $\mathbf{b}_1 = \sqrt{n}(\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})$. From (2.8),

$$\begin{aligned}& [\mathbf{G}_0 - c\,\mathbf{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{n}}[\mathbf{G}_1 - \lambda_{1n}\mathbf{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{n}}[\mathbf{G}_0 - c\,\mathbf{\Omega}]\mathbf{b}_1 + \frac{1}{\sqrt{q_n}}[-c\mathbf{H}_1]\boldsymbol{\beta} \\ = \;& o_p(\frac{1}{\sqrt{n}}) \; .\end{aligned}$$

Since $(\mathbf{G}_0 - c\,\mathbf{\Omega})\boldsymbol{\beta} = \mathbf{0}$, (2.8) gives

$$(6.6) \qquad \begin{bmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}_{22.1}\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) = (\mathbf{G}_1 - \lambda_{1n}\mathbf{\Omega} - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) \; .$$

Multiplication of (6.6) on the left by $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2')$ yields

$$\lambda_{1n} = \frac{\boldsymbol{\beta}'(\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}} + o_p(1) \ . \tag{6.7}$$

Also multiplication of (6.6) on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ and substitution for λ_{1n} from (6.6) yields

$$\begin{aligned}(6.8)\ \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) &= \boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{G}_1 - \lambda_{1n}\boldsymbol{\Omega} - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) \\ &= \boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) \ .\end{aligned}$$

By using the relation $\mathbf{V}\boldsymbol{\beta} = \mathbf{u}$, we obtain

$$\begin{aligned}(6.9)\qquad & (\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} \\ = \ & \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{2n}'\mathbf{Z}_{2.1}'\mathbf{u} + \sqrt{c}\frac{1}{\sqrt{K_{2n}}}\left[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{u} - K_{2n}\boldsymbol{\Omega}\boldsymbol{\beta}\right] \\ & -\sqrt{cc_*}\frac{1}{\sqrt{q_n}}\left[\mathbf{V}'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u} - q_n\boldsymbol{\Omega}\boldsymbol{\beta}\right] \ ,\end{aligned}$$

where $K_n + q_n = n$. Since we have the conditional expectation given $\mathbf{Z}$ as

$$\mathcal{E}\left[\boldsymbol{\Pi}_{2n}'\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}'\mathbf{Z}_{2.1}\boldsymbol{\Pi}_{2n}|\mathbf{Z}\right] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\Pi}_{2n}'\mathbf{A}_{22.1}\boldsymbol{\Pi}_{2n} \ ,$$

we apply the central limit theorem with Lindeberg condition to the first term of (6.9). (See *Theorem 1* of Anderson and Kunitomo (1992)). Conditions (II) and (III) imply that $(1/\sqrt{n})\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}'\mathbf{u}$ has a limiting normal distribution with covariance matrix $\sigma^2\boldsymbol{\Phi}_{22.1}$. This proves (i) of *Theorem 1.*

Next we shall consider (ii) of *Theorem 1.* We need to prove that the limiting distribution of $T_n = T_{1n} + \sqrt{c}T_{2n} - \sqrt{c\,c_*}T_{3n}$ is normal by applying a central limit theorem, where $T_{1n} = \mathbf{a}'(1/\sqrt{n})\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}'\mathbf{u}$, $T_{2n} = \mathbf{a}'(1/\sqrt{K_{2n}})\mathbf{W}_2'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{u}$, $T_{3n} = \mathbf{a}'(1/\sqrt{q_n})\mathbf{W}_2'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u}$ for any constant vector $\mathbf{a}$ and

$$\mathbf{W}_2' = (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}]\mathbf{V}' \ .$$

For the second and third terms on the right-hand side of (6.9), we notice that each row vector of $\mathbf{W}_2$ ($\mathbf{w}_{2i} = (\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{v}_i - u_i\mathbf{Cov}(\mathbf{v}_i^{(n)}u_i))/\sigma^2)$) and u_i ($i = 1, \cdots, n$) are uncorrelated and $\mathcal{E}[\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}] = (1/\sigma^2)[\sigma^2\mathbf{\Omega} - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{\Omega}]_{22}$. Thus

$$(6.10) \qquad (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}}]\frac{1}{\sqrt{K_{2n}}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{u} - K_{2n}\mathbf{\Omega}\boldsymbol{\beta}\right]$$
$$= \frac{1}{\sqrt{K_{2n}}}\sum_{i,j=1}^{n}\mathbf{w}_{2i}u_j p_{ij}^{(n)}$$

and

$$(6.11) \qquad (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}}]\frac{1}{\sqrt{q_n}}\left[\mathbf{V}^{'}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\mathbf{u} - q_n\mathbf{\Omega}\boldsymbol{\beta}\right]$$
$$= \frac{1}{\sqrt{q_n}}\sum_{i,j=1}^{n}\mathbf{w}_{2i}u_j[\delta_i^j - q_{ij}^{(n)}] \ ,$$

where $p_{ij}^{(n)} = \mathbf{z}_{in}^{*'}\left[\sum_{k=1}^{n}\mathbf{z}_{kn}^{*}\mathbf{z}_{kn}^{*'}\right]^{-1}\mathbf{z}_{jn}^{*}$, $q_{ij}^{(n)} = \mathbf{z}_i^{(n)'}\left[\sum_{k=1}^{n}\mathbf{z}_k^{(n)}\mathbf{z}_k^{(n)'}\right]^{-1}\mathbf{z}_j^{(n)}$ and $\delta_i^i = 1, \delta_i^j = 0 \ (i \neq j)$. Then the variances of T_{2n} and T_{3n} are

$$\frac{1}{K_{2n}}\mathcal{E}\{[\mathbf{a}^{'}(\sum_{i=1}^{n}\mathbf{w}_{2i}u_i p_{ii}^{(n)} + \sum_{i\neq j}\mathbf{w}_{2i}u_j p_{ij}^{(n)})]^2|\mathbf{Z}\}$$
$$= \frac{1}{K_{2n}}\sum_{i=1}^{n}\mathcal{E}[u_i^2\mathbf{a}^{'}\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}\mathbf{a}p_{ii}^{(n)2}] + \frac{1}{K_{2n}}\sum_{i\neq j}\mathcal{E}(u_j^2)\mathcal{E}(\mathbf{a}^{'}\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}\mathbf{a})p_{ij}^{(n)2} \ ,$$

and

$$\frac{1}{q_n}\mathcal{E}\{[\mathbf{a}^{'}(\sum_{i=1}^{n}\mathbf{w}_{2i}u_i(1 - q_{ii}^{(n)}) - \sum_{i\neq j}\mathbf{w}_{2i}u_j q_{ij}^{(n)}]^2|\mathbf{Z}\}$$
$$= \frac{1}{q_n}\sum_{i=1}^{n}\mathcal{E}[u_i^2\mathbf{a}^{'}\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}\mathbf{a}](1 - 2q_{ii}^{(n)} + q_{ii}^{(n)2}) + \frac{1}{q_n}\sum_{i\neq j}\mathcal{E}(u_i^2)\mathcal{E}(\mathbf{a}^{'}\mathbf{w}_{2j}\mathbf{w}_{2i}^{'}\mathbf{a})q_{ij}^{(n)2} \ .$$

By using the relations $\sum_{i,j=1}^{n} p_{ij}^{(n)2} = K_{2n}$, $\sum_{i,j=1}^{n} q_{ij}^{(n)2} = K_n$ and $\sum_{i=1}^{n}(1 - 2q_{ii}^{(n)} + q_{ii}^{(n)2}) + \sum_{i\neq j} q_{ij}^{(n)2} = q_n$, the limiting variances of T_{2n} and T_{3n} are the limits of

$$(6.12) \qquad \frac{1}{K_{2n}}\mathbf{a}^{'}\left[K_{2n}\sigma^2\mathcal{E}(\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}) + \sum_{i=1}^{n}p_{ii}^{(n)2}\mathbf{\Gamma}_{44.2}\right]\mathbf{a}$$

and

$$(6.13) \qquad \frac{1}{q_n}\mathbf{a}^{'}\left[q_n\sigma^2\mathcal{E}(\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}) + (n - 2K_n + \sum_{i=1}^{n}q_{ii}^{(n)2})\mathbf{\Gamma}_{44.2}\right]\mathbf{a} \ .$$

In order to evaluate the covariances of three terms of T_n, we first notice

$$(6.14) \qquad \mathcal{E}\{[\frac{1}{\sqrt{n}}\mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}^{'}\mathbf{u}][\frac{1}{\sqrt{n}}\mathbf{W}_2^{'}\left(\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'} - c_*(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\right)\mathbf{u}]^{'}|\mathbf{Z}\}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^{*}\left(p_{ii}^{(n)} - c_*(1 - q_{ii}^{(n)})\right)\mathcal{E}(u^2\mathbf{w}_2^{'}) = \mathbf{\Xi}_{3.2}^{(n)} \ (say) \ .$$

Second,

$$
\begin{aligned}
& \mathcal{E}\{[\frac{1}{\sqrt{n}}\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{u}][\frac{1}{\sqrt{n}}\mathbf{W}_2^{'}(\mathbf{I}_n-\mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\mathbf{u}]^{'}|\mathbf{Z}\} \\
= & \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}(u_i^2\mathbf{w}_{2i}\mathbf{w}_{2i}^{'})p_{ii}^{(n)}[1-q_{ii}^{(n)}]+\frac{1}{n}\sum_{i\neq j}^{n}\sigma^2\mathcal{E}(\mathbf{w}_{2i}\mathbf{w}_{2i}^{'})p_{ij}^{(n)}[\delta_i^j-q_{ij}^{(n)}] \\
= & \frac{1}{n}[K_{2n}-\sum_{i=1}^{n}p_{ii}^{(n)}q_{ii}^{(n)}]\mathbf{\Gamma}_{44.2}
\end{aligned}
$$

by using the relations that $\sum_{i,j=1}^{n}p_{ij}^{(n)}\delta_i^j=K_{2n}$ and $\sum_{i,j=1}^{n}p_{ij}^{(n)}q_{ji}^{(n)}=K_{2n}$. Hence we have evaluated each term

$$
\begin{aligned}
\mathcal{E}(T_n^2) = & \ \mathcal{E}(T_{1n}^2)+c\mathcal{E}(T_{2n}^2)+cc_*\mathcal{E}(T_{3n}^2) \\
& +2\sqrt{c}\mathcal{E}(T_{1n}T_{2n})-2\sqrt{cc_*}\mathcal{E}(T_{1n}T_{3n})-2c\sqrt{c_*}\mathcal{E}(T_{2n}T_{3n})\ .
\end{aligned}
$$

Then we use the relation $c(1+c_*)=c_*$ for the coefficients of two terms of $\mathcal{E}(u_i^2\mathbf{w}_{2i}\mathbf{w}_{2i}^{'})$. Also by using the relation $cc_*(1-c_*)-2cc_*=-c_*^2$ for the coefficients of $\mathbf{\Gamma}_{44.2}$, we find that

$$
\begin{aligned}
& \lim_{n\to\infty}\Bigg[c\frac{n}{K_{2n}}\frac{1}{n}\sum_{i=1}^{n}p_{ii}^{(n)2}+cc_*\frac{1}{q_n}(n-2K_n+\sum_{i=1}^{n}q_{ii}^{(n)2}) \\
& \qquad -2c\sqrt{c_*\frac{n}{K_{2n}}\frac{n}{q_n}}\frac{1}{n}(K_{2n}-\sum_{i=1}^{n}p_{ii}^{(n)}q_{ii}^{(n)})\Bigg] \\
= & \lim_{n\to\infty}\eta_n\ ,
\end{aligned}
$$

where $\eta_n=(1/n)\sum_{i=1}^{n}[p_{ii}^{(n)}+c_*q_{ii}^{(n)}]^2-c_*^2$. By using (6.14), the limiting covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI}-\boldsymbol{\beta}_2)$ is (3.5).

Finally, by using *Lemma 2* below for every constant vector $\mathbf{a}$, we have the asymptotic normality of (3.4) with the asymptotic covariance matrix $\mathbf{\Psi}^*$ and it proves (ii) of *Theorem 1.*

When $G_2=1$, we can use the relation $\sigma^2=\omega_{11}-2\beta_2\omega_{12}+\beta_2^2\omega_{22}$ for $\mathbf{\Omega}=(\omega_{ij})$ to obtain $\sigma^2\omega_{22}-(\omega_{12}-\beta\omega_{22})^2=|\mathbf{\Omega}|$.

Q.E.D

Lemma 1 : Assume Condition (VI) and $c=\lim_{n\to\infty}K_{2n}/n$. Then

$$
(6.15)\quad \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}[p_{ii}^{(n)}+c_*q_{ii}^{(n)}]^2-c_*^2=\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}[p_{ii}^{(n)}-c_*(1-q_{ii}^{(n)})]^2=0\ ,
$$

where $c_* = c/(1-c)$, $p_{ij}^{(n)} = (\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}')_{ij}$ and $q_{ij}^{(n)} = (\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}')_{ij}$.

Proof of Lemma 1 : Note that

$$\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}' = \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1' + \mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'$$

because $\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1'$, $\mathbf{ZC}\left[(\mathbf{ZC})'\mathbf{ZC}\right]^{-1}(\mathbf{ZC})' = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ for any non-singular matrix $\mathbf{C}$ and

$$(\mathbf{Z}_1, \mathbf{Z}_{2.1}) = (\mathbf{Z}_1, \mathbf{Z}_2)\begin{bmatrix} \mathbf{I}_{K_1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{O} & \mathbf{I}_{K_{2n}} \end{bmatrix} .$$

Then the left-hand side of (6.15) is

$$\begin{aligned}
& \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)} + c_*(p_{ii}^{(n)} + (\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1')_{ii})\right]^2 - c_*^2 \\
= & \operatorname*{plim}_{n\to\infty} \frac{1}{(1-c)^2 n}\sum_{i=1}^{n}\left\{\left[p_{ii}^{(n)} + c(\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1')_{ii}\right]^2 - c^2\right\} \\
= & \operatorname*{plim}_{n\to\infty} \frac{1}{(1-c)^2 n}\sum_{i=1}^{n}\left\{p_{ii}^{(n)2} + 2cp_{ii}^{(n)}(\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1')_{ii} + c^2(\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1')_{ii}^2 - c^2\right\} .
\end{aligned}$$

Note $0 \le p_{ii}^{(n)} \le 1$, $0 \le (\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1')_{ii} \le 1$, and $\mathrm{tr}(\mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{Z}_1') = K_1$. Hence the above plimit is $1/(1-c)^2$ times $\operatorname{plim}_{n\to\infty}\sum_{i=n}^{n}[p_{ii}^{(n)}]^2/n - c^2$. However,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)} - c\right]^2 &= \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)2} - 2cp_{ii}^{(n)} + c^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}p_{ii}^{(n)2} - 2c\frac{K_{2n}}{n}c + c^2 \\
&\to \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}p_{ii}^{(n)2} - c^2 .
\end{aligned}$$

Q.E.D.

Lemma 2 : Let $t_{1i}^{(n)} = \mathbf{a}'\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*$ and $t_{2i}^{(n)} = \mathbf{a}'\mathbf{w}_{2i}$ $(i = 1, \cdots, n)$ for any (non-zero) constant vector $\mathbf{a}$. Then as $n \to \infty$,

$$\begin{aligned}
(6.16) \qquad T_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}t_{1i}^{(n)}u_i + \frac{1}{\sqrt{n}}\sum_{i,j=1}^{n}t_{2i}^{(n)}u_j\left[p_{ij}^{(n)} - c_*(\delta_i^j - q_{ij}^{(n)})\right] \\
&\xrightarrow{d} N(\mathbf{0}, \Delta) ,
\end{aligned}$$

where $\Delta = \mathbf{a}^{'}\mathbf{\Phi}_{22.1}\mathbf{\Psi}^{*}\mathbf{\Phi}_{22.1}\mathbf{a}$.

Proof of Lemma 2 : We have already shown that $\mathcal{E}(T_n) = 0$ and $\mathcal{E}[T_n^2] \longrightarrow \Delta$ as $n \longrightarrow \infty$ by Lemma 1. Then, in order to prove a central limit theorem by the standard characteristic function method, it is sufficient to show

$$\mathcal{E}[T_n^3] \longrightarrow 0 \ . \tag{6.17}$$

The third order moment of the second term of $T_n \times \sqrt{n}$, for instance, is

$$\begin{aligned} \mathcal{E}[\sum_{i,j=1}^{n} t_{2i}^{(n)} u_j p_{ij}^{(n)}]^3 &= \sum_{i,i^{'},i^{''},j,j^{'},j^{''}} p_{ij}^{(n)} p_{i^{'}j^{'}}^{(n)} p_{i^{''}j^{''}}^{(n)} \mathcal{E}(t_{2i}t_{2i^{'}}t_{2i^{''}}u_j u_{j^{'}} u_{j^{''}}) \\ &= \sum_{i=i^{'}=i^{''},j=j^{'}=j^{''}} + \sum_{i=i^{'}=j^{''}\neq i^{''}=j=j^{'}} + \sum_{i=i^{'}=j^{'}\neq i^{''}=j=j^{''}} + \sum_{i=i^{''}=j^{''}\neq i^{'}=j=j^{'}} \end{aligned}$$

because each terms of t_{2i} and u_j for any i and j are uncorrelated and other terms except the above summations are zeros. Then we need to evaluate four types of summations. For the first three summations we use the fact that for any i,j $p_{ij}^{(n)} = p_{ji}^{(n)}$, $|p_{ij}^{(n)}| \leq 1$,

$$|\sum_{i,j=1}^{n} p_{ij}^{(n)3}| \leq \sum_{i,j=1}^{n} p_{ij}^{(n)} p_{ji}^{(n)} = \sum_{i=1}^{n} p_{ii}^{(n)} = K_{2n}$$

and

$$|\sum_{i,j=1}^{n} p_{ij}^{(n)2} p_{ji}^{(n)}| \leq \sum_{i,j=1}^{n} p_{ij}^{(n)} p_{ji}^{(n)} = \sum_{i=1}^{n} p_{ii}^{(n)} = K_{2n} \ .$$

For the fourth summation we use *Lemma 3* to obtain $|\sum_{i,j=1}^{n} p_{ii}^{(n)} p_{jj}^{(n)} p_{ij}^{(n)}| \leq K_{2n}$. Because of the boundedness of 6th order moments,

$$(\frac{1}{\sqrt{n}})^3 \mathcal{E}|\sum_{i,j=1}^{n} t_{2i}^{(n)} u_j p_{ij}^{(n)}|^3 \longrightarrow 0$$

as $n \longrightarrow \infty$.

Next we set a projection matrix $\mathbf{D} = (d_{ij}^{(n)}) = (\delta_i^j - q_{ij}^{(n)})$. Then we use the fact that for any i,j, $d_{ij}^{(n)} = d_{ji}^{(n)}$, $|d_{ij}^{(n)}| \leq 1$,

$$|\sum_{i,j=1}^{n} d_{ij}^{(n)3}| \leq \sum_{i,j=1}^{n} d_{ij}^{(n)} d_{ji}^{(n)} = \sum_{i=1}^{n} d_{ii}^{(n)} = n - K_n$$

and

$$|\sum_{i,j=1}^{n} d_{ij}^{(n)2} d_{ji}^{(n)}| \leq \sum_{i,j=1}^{n} d_{ij}^{(n)} d_{ji}^{(n)} = \sum_{i=1}^{n} d_{ii}^{(n)} = n - K_n \ .$$

We also apply similar arguments to the first term under Condition (III) and other cross product terms of $p_{ij}^{(n)}$ and $d_{ij}^{(n)}$, we have the result.

Q.E.D.

Lemma 3 : Let an $n \times n$ matrix $\mathbf{P} = (p_{ij})$ satisfying $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}'$ and rank$(\mathbf{P}) = r \leq n$. Then

$$\sum_{i,j=1}^{n} p_{ii}p_{jj}p_{ij} \leq r \ . \tag{6.18}$$

Proof of Lemma 3 : Since $\mathbf{P}$ is a projection matrix, there exists an orthogonal matrix $\mathbf{C}$ such that

$$\mathbf{P} = \mathbf{C} \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{C}' \ .$$

Then

$$\begin{aligned} \sum_{i,j=1}^{n} p_{ii}p_{jj}p_{ij} &= (p_{11}, \cdots, p_{nn})\, \mathbf{C} \begin{bmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{C}' \begin{pmatrix} p_{11} \\ \vdots \\ p_{nn} \end{pmatrix} \\ &\leq (p_{11}, \cdots, p_{nn})\, \mathbf{C}\mathbf{C}' \begin{pmatrix} p_{11} \\ \vdots \\ p_{nn} \end{pmatrix} \\ &= \sum_{i=1}^{n} p_{ii}^2 \leq \left[\max_{1 \leq j \leq n} p_{jj} \right] \sum_{i=1}^{n} p_{ii} \ . \end{aligned}$$

Since $\mathbf{C}$ is an orthogonal matrix and $0 \leq p_{ii} \leq 1$, we have (6.18).

Q.E.D.

Proof of Corollary 2 : Consider

$$s_{jk}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} z_{ji} z_{ki} \xrightarrow{p} m_{jk}$$

and $\mathbf{M} = (m_{jk})$. Then

$$\frac{1}{n} \sum_{i=1}^{n} \left[q_{ii}^{(n)2} - c^2 \right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{[\frac{1}{n}\mathbf{z}_i^{(n)'}\mathbf{M}^{-1}\mathbf{z}_i^{(n)} - c] + \frac{K_1}{K_n} + \frac{1}{n}\mathbf{z}_i^{(n)'}[\mathbf{S}^{-1} - \mathbf{M}^{-1}]\mathbf{z}_i^{(n)}\right\}^2$$

$$= (\frac{K_n}{n})^2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K_n}(\mathbf{z}_i^{(n)'}\mathbf{M}^{-1}\mathbf{z}_i^{(n)} - K_n)]^2$$

$$-2(\frac{K_n}{n})^2\frac{1}{n}\sum_{i=1}^{n}[\frac{1}{K_n}(\mathbf{z}_i^{(n)'}\mathbf{M}^{-1}\mathbf{z}_i^{(n)} - K_n)][\frac{1}{K_n}\sum_{j,j'=1}^{K_n} z_{ji}z_{j'i}(s^{jj'} - m^{jj'})]$$

$$+(\frac{K_n}{n})^2\frac{1}{n}\sum_{i=1}^{n}[\frac{1}{K_n}\sum_{j,j'=1}^{K_n} z_{ji}z_{j'i}(s^{jj'} - m^{jj'})]^2 ,$$

where $s^{jk} = (s_{jk}^{(n)})^{-1}$ and $m^{jk} = (m_{jk})^{-1}$. Because

$$(\frac{1}{K_n})^2\sum_{j,j',k,k'=1}^{K_n}(s^{jj'} - m^{jj'})(s^{kk'} - m^{kk'})[\frac{1}{n}\sum_{i=1}^{n} z_{ij}z_{ij'}z_{ik}z_{ik'}]$$

converges to 0 in probability under Condition (VII), we have the result by Lemma 1.

Q.E.D.

Proof of Theorem 3 :

(I) We make use of the fact that $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{Z}_{2.1}(\mathbf{Z}_{2.1}'\mathbf{Z}_{2.1})^{-1}\mathbf{Z}_{2.1}'$ are idempotent of rank K_n and K_{2n}, respectively, and that the boundedness of $\mathbf{E}[v_{ji}^4|\mathbf{z}_i^{(n)}]$ implies a Lindeberg condition

$$\sup_{i=1,\cdots,n}\mathcal{E}\left[\mathbf{v}_i'\mathbf{v}_i\mathbf{I}(\mathbf{v}_i'\mathbf{v}_i > a)|\mathbf{z}_1^{(n)},\cdots,\mathbf{z}_n^{(n)}\right] \xrightarrow{p} 0 \ \ (a \to \infty) . \tag{6.19}$$

We shall refer to *Theorem 1* of Anderson and Kunitomo (1992). Let

$$\begin{aligned}\mathbf{G}_1^* &= \sqrt{n}[\frac{1}{n}\mathbf{G} - \frac{1}{n}\mathbf{\Pi}_{2n}'\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}] \\ &= \frac{1}{\sqrt{n}}\mathbf{\Pi}_{2n}'\mathbf{Z}_{2.1}'\mathbf{V} + \frac{1}{\sqrt{n}}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} .\end{aligned} \tag{6.20}$$

Since the matrix $\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}$ is positive definite and $\mathcal{E}[\mathbf{v}_i^{(n)}\mathbf{v}_i^{(n)'}|z_i^{(n)}]$ is bounded, there is a (constant) $\bar{\mathbf{\Omega}}$ such that

$$\begin{aligned}\mathcal{E}[\frac{1}{\sqrt{n}}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}] &= \mathcal{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{\Omega}_i^{(n)}p_{ii}^{(n)}] \\ &\leq \frac{K_{2n}}{\sqrt{n}}\bar{\mathbf{\Omega}} \longrightarrow \mathbf{O} .\end{aligned} \tag{6.21}$$

Then

$$\mathbf{G}_1^*\boldsymbol{\beta} - \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} = \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0}\ . \tag{6.22}$$

For the LIML estimator (2.8) implies

$$(\mathbf{0}, \mathbf{I}_{G_2})\left[\frac{1}{n}\boldsymbol{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H}\right]\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} = \mathbf{0}\ . \tag{6.23}$$

By using the facts that $(1/\sqrt{n})\mathbf{G}_1^* \xrightarrow{p} \mathbf{O}$, $\lambda_n \xrightarrow{p} 0$ and $[1/q_n]\mathbf{H} \xrightarrow{p} \boldsymbol{\Omega}$, we have

$$\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\text{plim}_{n\to\infty}\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} = \mathbf{0}\ ,$$

which implies $\text{plim}_{n\to\infty}\hat{\boldsymbol{\beta}}_{2.LI} = \boldsymbol{\beta}_2$ because $\boldsymbol{\Phi}_{22.1}$ is positive definite. Then again (2.8) implies

$$\sqrt{n}\left[\frac{1}{n}\boldsymbol{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H}\right]\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0}. \tag{6.24}$$

Lemma 4 : Let λ_n $(n > 2)$ be the smallest root of (2.9). For $0 < \nu < 1 - \eta$ and $0 \leq \eta < 1$,

$$n^{\nu}\lambda_n \xrightarrow{p} 0 \tag{6.25}$$

as $n \to \infty$.

Proof of Lemma 4

Write

$$\begin{aligned} \lambda_n &= \min_{\mathbf{b}}\frac{\mathbf{b}^{'}\frac{1}{n}\mathbf{G}\mathbf{b}}{\mathbf{b}^{'}\frac{1}{q_n}\mathbf{H}\mathbf{b}} \\ &\leq \frac{q_n}{n}\frac{\boldsymbol{\beta}^{'}\mathbf{G}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{H}\boldsymbol{\beta}} = \frac{q_n}{n}\frac{\boldsymbol{\beta}^{'}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{V}^{'}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\mathbf{V}\boldsymbol{\beta}}\ . \end{aligned} \tag{6.26}$$

By using the boundedness of the fourth order moments of $\mathbf{v}_i$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{v}_i^{'} \xrightarrow{p} \boldsymbol{\Omega}\ . \tag{6.27}$$

Also $n^{-(1-\nu)}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} \xrightarrow{p} \mathbf{O}$ by using the similar arguments as (6.22). Then

$$n^{\nu}\lambda_n \leq \left[\frac{q_n}{n}\right] \frac{n^{-(1-\nu)}\boldsymbol{\beta}'\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}}{n^{-1}\boldsymbol{\beta}'\mathbf{V}'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{V}\boldsymbol{\beta}} \xrightarrow{p} 0 \tag{6.28}$$

as $n \to \infty$. **Q.E.D.**

Due to *Lemma 4*, $\sqrt{n}\,\lambda_n \xrightarrow{p} 0$ when $0 \leq \eta < 1/2$ (and the asymptotic distributions of the LIML and TSLS estimators are equivalent). Then

$$(\mathbf{0}, \mathbf{I}_{G_2})\frac{1}{n}\mathbf{\Pi}_{2n}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}^{(n)}\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2})\mathbf{G}_1^*\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \,. \tag{6.29}$$

We notice that

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Omega}_i^{(n)} \otimes \mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\mathbf{z}_{in}^{*'}\mathbf{\Pi}_{22}^{(n)} - \mathbf{\Omega} \otimes \mathbf{\Phi}_{22.1} \\ = & \frac{1}{n}\sum_{i=1}^{n}(\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}) \otimes \mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\mathbf{z}_{in}^{*'}\mathbf{\Pi}_{22}^{(n)} \\ & + \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Omega} \otimes \left[\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\mathbf{z}_{in}^{*'}\mathbf{\Pi}_{22}^{(n)} - \mathbf{\Phi}_{22.1}\right] \xrightarrow{p} \mathbf{O} \end{aligned}$$

because Condition (II') and

$$\begin{aligned} & \|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}) \otimes \mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\mathbf{z}_{in}^{*'}\mathbf{\Pi}_{22}^{(n)}\| \\ \leq & \max_{1 \leq i \leq n}\|\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}\|\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\mathbf{z}_{in}^{*'}\mathbf{\Pi}_{22}^{(n)}\| \xrightarrow{p} 0 \,. \end{aligned}$$

Then by applying the central limit theorem (see *Theorem 1* of Anderson and Kunitomo (1992)) to $(1/\sqrt{n})\mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}$, we obtain the limiting normal distribution $N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1})$. This proves (i) of *Theorem 4* for $0 \leq \eta < 1/2$.

(II) We consider the asymptotic distribution of the LIML estimator when $1/2 \leq \eta < 1$. By using the argument of (6.23) and the fact that $\lambda_n \xrightarrow{p} 0$, we have $\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2 \xrightarrow{p} \mathbf{0}$. By multiplying $\boldsymbol{\beta}'$ from the left to (6.40), we have

$$\begin{aligned} & \boldsymbol{\beta}'\left\{\sqrt{n}[\frac{K_{2n}}{n} - \lambda_n]\mathbf{\Omega} + \frac{1}{\sqrt{n}}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}}[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} - K_{2n}\mathbf{\Omega}] \right. \\ & \left. -\lambda_n\sqrt{\frac{n}{q_n}}\mathbf{H}_1\right\} \times \left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0} \,. \end{aligned}$$

Lemma 5 : For $0 \leq \eta < 1$,

$$\sqrt{n}\left[\lambda_n - \frac{K_{2n}}{n}\right] \xrightarrow{p} 0 \tag{6.30}$$

as $n \longrightarrow \infty$.

Proof of Lemma 5 : The result follows from (6.25), (6.26) and $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} > 0$. **Q**.**E**.**D**.

Multiply (6.24) on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ to obtain

$$\begin{aligned}
&(\mathbf{0}, \mathbf{I}_{G_2})\sqrt{n}\left\{\left[\frac{1}{n}\boldsymbol{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{2n} + \frac{K_{2n}}{n}\boldsymbol{\Omega}\right]\right.\\
&+\frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\boldsymbol{\Pi}_{2n} + \frac{1}{\sqrt{n}}(\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n})\boldsymbol{\Omega})\right]\\
&\left.-\lambda_n\frac{1}{q_n}\mathbf{H}\right\} \times \left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0}\ .
\end{aligned}$$

We consider the asymptotoic behavior of the quadratic term

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n}\boldsymbol{\Omega}\right]\\
=\ &\frac{1}{\sqrt{n}}\left[\sum_{i,j=1}^{n} p_{ij}^{(n)}\left(\mathbf{v}_i\mathbf{v}_j^{'} - \delta_i^j\boldsymbol{\Omega}_i^{(n)}\right)\right] + \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} p_{ii}^{(n)}\left(\boldsymbol{\Omega}_i^{(n)} - \boldsymbol{\Omega}\right)\right]\ ,
\end{aligned}$$

where δ_i^j is the indicator function ($\delta_i^i = 1$ and $\delta_i^j = 0$ $(i \neq j)$). For any constant vectors $\mathbf{a}$ and $\mathbf{b}$, there exists a positive constant M_1 such that

$$\begin{aligned}
&\frac{1}{n}\mathcal{E}\left[\sum_{i,j=1}^{n} p_{ij}^{(n)} \times \mathbf{a}^{'}(\mathbf{v}_i^{(n)}\mathbf{v}_j^{(n)'} - \delta_i^j\boldsymbol{\Omega}_i^{(n)})\mathbf{b}\right]^2\\
=\ &\frac{1}{n}\mathcal{E}\left[\sum_{i=1}^{n} p_{ii}^{(n)2}[\mathbf{a}^{'}(\mathbf{v}_i\mathbf{v}_i^{'} - \boldsymbol{\Omega}_i^{(n)})\mathbf{b}]^2\right.\\
&\left.+\sum_{i\neq j} p_{ij}^{(n)2}[\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j\mathbf{b}]^2 + \sum_{i\neq j} p_{ij}^{(n)2}[\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j^{'}\mathbf{b}\mathbf{a}^{'}\mathbf{v}_j\mathbf{v}_i^{'}\mathbf{b}]\right]\\
\leq\ &M_1\frac{K_{2n}}{n} \longrightarrow 0
\end{aligned}$$

because the conditional moments of v_{ji}^4 are bounded, $\sum_{i=1}^{n} p_{ii}^{(n)} = K_{2n}$ and $\sum_{i=1}^{n} p_{ii}^{(n)2} \leq K_{2n}$. Since

$$\| \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} p_{ii}^{(n)} \left(\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega} \right) \right] \| \leq \left[\sqrt{n} \max_{1 \leq i \leq n} \| \mathbf{\Omega}_i^{(n)} - \mathbf{\Omega} \| \right] \frac{K_{2n}}{n} ,$$

we find

$$\frac{1}{\sqrt{n}} \left[\mathbf{V}' \mathbf{Z}_{2.1} \mathbf{A}_{22.1}^{-1} \mathbf{Z}_{2.1}' \mathbf{V} - K_{2n} \mathbf{\Omega} \right] \xrightarrow{p} \mathbf{O} \tag{6.31}$$

when $0 \leq \eta < 1$.

We now use (6.23), (6.25) and the fact that

$$\left[\frac{1}{n} \mathbf{\Pi}_{2n}' \mathbf{A}_{22.1} \mathbf{\Pi}_{2n} + \frac{K_{2n}}{n} \mathbf{\Omega} - \lambda_n \frac{1}{q_n} \mathbf{H} \right] \boldsymbol{\beta} = o_p(\frac{1}{\sqrt{n}}) .$$

By multiplying the preceding equation out to separate the terms with factor $\boldsymbol{\beta}$ and with the factor $\sqrt{n}\,(\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})$, we have

$$(\mathbf{0}, \mathbf{I}_{G_2}) \left[\frac{1}{n} \mathbf{\Pi}_{2n}' \mathbf{A}_{22.1} \mathbf{\Pi}_{2n} \sqrt{n} (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}) + \frac{1}{\sqrt{n}} \mathbf{\Pi}_{2n}' \mathbf{Z}_{2.1}' \mathbf{V} \boldsymbol{\beta} \right] \xrightarrow{p} \mathbf{0} , \tag{6.32}$$

which is equivalent to

$$\mathbf{\Phi}_{22.1} \ \sqrt{n} (\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) - \frac{1}{\sqrt{n}} \mathbf{\Pi}_{22}^{(n)'} \mathbf{Z}_{2.1}' \mathbf{V} \boldsymbol{\beta} \xrightarrow{p} \mathbf{0} . \tag{6.33}$$

By applying the CLT to the second term of (6.33) as (I), we complete the proof of (i) of *Theorem 3* for the LIML estimator of $\boldsymbol{\beta}$ when $1/2 \leq \eta < 1$.

(III) Next, we shall investigate the asymptotic property of the TSLS estimator. If we substitute λ_n for 0 in (2.8), we have the TSLS estimator. Then we find that the limiting distribution of the TSLS estimator is the same as the LIML estimator when $0 \leq \eta < 1/2$.

When $\eta = 1/2$, however, we have

$$\mathbf{G}_1^* \boldsymbol{\beta} - \left[c \mathbf{\Omega} \boldsymbol{\beta} + \frac{1}{\sqrt{n}} \mathbf{\Pi}_{2n}' \mathbf{Z}_{2.1}' \mathbf{V} \boldsymbol{\beta} \right] \xrightarrow{p} \mathbf{O} . \tag{6.34}$$

We set $\hat{\boldsymbol{\beta}}_{TS}' = (1, -\hat{\boldsymbol{\beta}}_{2.TS}')$, which is the solution of (2.11). By evaluating each term of

$$(\mathbf{0}, \mathbf{I}_{G_2}) \sqrt{n} \left[\frac{1}{n} \mathbf{\Pi}_{2n}' \mathbf{A}_{22.1} \mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}} \mathbf{G}_1^* \right] \left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}) \right] = \mathbf{0} ,$$

we have

$$\left[\frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}_{TS}-\boldsymbol{\beta})-(\mathbf{0},\mathbf{I}_{G_2})\mathbf{G}_1^*\boldsymbol{\beta}=o_p(1)\ . \tag{6.35}$$

Then the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS}-\boldsymbol{\beta}_2)$ is the same as that of $\mathbf{\Phi}_{22.1}^{-1}(\mathbf{0},\mathbf{I}_{G_2})\mathbf{G}_1^*\boldsymbol{\beta}$. By using $(1/\sqrt{n})\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}\xrightarrow{p}c\mathbf{\Omega}\boldsymbol{\beta}$ and applying the CLT as (I), we have the result for the TSLS estimator of $\boldsymbol{\beta}$ when $\eta=1/2$.

When $1/2<\eta<1$, we notice

$$\begin{aligned} & n^{1-\eta}\left[\frac{1}{n}\mathbf{G}-\frac{1}{n}\mathbf{\Pi}_{2n}'\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}\right]\boldsymbol{\beta} \\ = & \frac{K_{2n}}{n^{\eta}}\mathbf{\Omega}\boldsymbol{\beta}+\frac{1}{n^{\eta}}\mathbf{\Pi}_{2n}'\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}+\frac{1}{n^{\eta}}\left[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}-K_{2n}\mathbf{\Omega}\right]\boldsymbol{\beta}. \end{aligned} \tag{6.36}$$

Because the last two terms of the right-hand side of (6.36) except the first term are of the order $o_p(n^{-\eta})$, we have

$$n^{1-\eta}\left[\frac{1}{n}\mathbf{G}-\frac{1}{n}\mathbf{\Pi}_{2n}'\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}\right]\boldsymbol{\beta}\xrightarrow{p}c\mathbf{\Omega}\boldsymbol{\beta} \tag{6.37}$$

as $n\longrightarrow\infty$. Hence by using the similar arguments as (I) for the TSLS estimator of $\boldsymbol{\beta}$,

$$(\mathbf{0},\mathbf{I}_{G_2})\frac{1}{n}\mathbf{\Pi}_{2n}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}^{(n)}\times n^{1-\eta}(\hat{\boldsymbol{\beta}}_{2.TS}-\boldsymbol{\beta}_2)-(\mathbf{0},\mathbf{I}_{G_2})c\mathbf{\Omega}\boldsymbol{\beta}\xrightarrow{p}\mathbf{0} \tag{6.38}$$

and we complete the proof of (ii) of *Theorem 3* for the TSLS estimator when $1/2\leq\eta<1$.

Q.E.D.

Proof of Theorem 4 : We set the vector of true parameters $\boldsymbol{\beta}'=(1,-\boldsymbol{\beta}_2')=(1,-\beta_2,\cdots,-\beta_{1+G_2})$. An estimator of the vector $\boldsymbol{\beta}_2$ is composed of

$$\hat{\beta}_k=\phi_k(\frac{1}{n}\mathbf{G},\frac{1}{q_n}\mathbf{H})\ \ (k=2,\cdots,1+G_2)\ . \tag{6.39}$$

For the estimator to be consistent we need the conditions

$$\beta_k=\phi_k\left[\begin{pmatrix}\boldsymbol{\beta}_2'\\ \mathbf{I}_{G_2}\end{pmatrix}\mathbf{\Phi}_{22.1}\left(\boldsymbol{\beta}_2,\mathbf{I}_{G_2}\right)+c\mathbf{\Omega},\mathbf{\Omega}\right]\ \ (k=2,\cdots,1+G_2) \tag{6.40}$$

as identities in $\boldsymbol{\beta}_2$, $\mathbf{\Phi}_{22.1}$, and $\mathbf{\Omega}$. Let a $(1+G_2)\times(1+G_2)$ matrix

$$\mathbf{T}^{(k)}=\left(\frac{\partial\phi_k}{\partial g_{ij}}\right)=(\tau_{ij}^{(k)})\ \ (k=2,\cdots,1+G_2;i,j=1,\cdots,1+G_2) \tag{6.41}$$

evaluated at the probability limits of (6.43). We write a $(1+G_2)\times(1+G_2)$ matrix $\mathbf{\Theta}$ $(=(\theta_{ij}))$

$$\mathbf{\Theta}=\begin{pmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\mathbf{\Phi}_{22.1}\left(\boldsymbol{\beta}_2,\mathbf{I}_{G_2}\right)=\begin{bmatrix}\boldsymbol{\beta}_2^{'}\mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2 & \boldsymbol{\beta}_2^{'}\mathbf{\Phi}_{22.1}\\ \mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2 & \mathbf{\Phi}_{22.1}\end{bmatrix},$$

where $\mathbf{\Phi}_{22.1}=(\rho_{m,l})$ $(m,l=2,\cdots,1+G_2)$, $(\mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2)_l=\sum_{j=2}^{1+G_2}\beta_j\rho_{lj}$ $(l=2,\cdots,1+G_2)$, $(\boldsymbol{\beta}_2^{'}\mathbf{\Phi}_{22.1})_m=\sum_{i=2}^{1+G_2}\beta_i\rho_{im}$ $(m=2,\cdots,1+G_2)$, and $\boldsymbol{\beta}_2^{'}\mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2=\sum_{i,j=2}^{1+G_2}\rho_{ij}\beta_i\beta_j$.
By differentiating each components of $\mathbf{\Theta}$ with respect to β_j $(j=1,\cdots,G_2)$, we have

$$\frac{\partial\mathbf{\Theta}}{\partial\beta_j}=(\frac{\partial\theta_{lm}}{\partial\beta_j}) \ , \tag{6.42}$$

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 $\frac{\partial\theta_{lm}}{\partial\beta_j}=0$ $(l,m=2,\cdots,1+G_2)$.
Hence

$$\mathbf{tr}\left(\mathbf{T}^{(k)}\frac{\partial\mathbf{\Theta}}{\partial\beta_j}\right)=2\tau_{11}^{(k)}\sum_{i=2}^{1+G_2}\rho_{ji}\beta_i+2\sum_{i=2}^{1+G_2}\rho_{ji}\tau_{ji}^{(k)}=\delta_j^k \ , \tag{6.43}$$

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

$$\mathbf{T}^{(k)}=\begin{bmatrix}\tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)'}\\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)}\end{bmatrix} . \tag{6.44}$$

Then (6.47) is represented as

$$2\tau_{11}^{(k)}\mathbf{\Phi}_{22.1}\boldsymbol{\beta}+2\mathbf{\Phi}_{22.1}\boldsymbol{\tau}_2^{(k)}=\boldsymbol{\epsilon}_k \ , \tag{6.45}$$

where $\boldsymbol{\epsilon}_k^{'}=(0,\cdots,0,1,0,\cdots,0)$ with 1 in the k-th place and zeros in other elements.
Since $\mathbf{\Phi}_{22.1}$ is positive definite, we solve (6.49) as

$$\boldsymbol{\tau}_2^{(k)}=\frac{1}{2}\mathbf{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k-\tau_{11}^{(k)}\boldsymbol{\beta}_2 \ . \tag{6.46}$$

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

$$\frac{\partial\mathbf{\Theta}}{\partial\rho_{ii}}=(\frac{\partial\theta_{lm}}{\partial\rho_{ii}}) \ , \tag{6.47}$$

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 $i \neq j$

$$\frac{\partial\mathbf{\Theta}}{\partial\rho_{ij}} = (\frac{\partial\theta_{lm}}{\partial\rho_{ij}}) \ , \tag{6.48}$$

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 we have the representation

$$\mathrm{tr}\left(\mathbf{T}^{(k)}\frac{\partial\mathbf{\Theta}}{\partial\rho_{ij}}\right) = \begin{cases} \beta_i^2\tau_{11}^{(k)} + 2\tau_{1i}^{(k)}\beta_i + \tau_{ii}^{(k)} & (i = j) \\ \\ 2\beta_i\beta_j\tau_{11}^{(k)} + 2\tau_{1j}^{(k)}\beta_i + 2\tau_{1i}^{(k)}\beta_j + 2\tau_{ij}^{(k)} & (i \neq j) \end{cases} . \tag{6.49}$$

In the matrix form we have a simple relation as

$$\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} + \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} + \mathbf{T}_{22}^{(k)} = \mathbf{O} \ . \tag{6.50}$$

Then we have the representation

$$\begin{aligned} \mathbf{T}_{22}^{(k)} &= -\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} - \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} \\ &= \tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \frac{1}{2}\left[\mathbf{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\epsilon}_k^{'}\mathbf{\Phi}_{22.1}^{-1}\right] \ . \end{aligned}$$

Next we consider the role of the second matrix in (6.43). By differentiating (6.43) with respect to ω_{ij} $(i, j = 1, \cdots, 1 + G_2)$, we have the condition

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

evaluated at the probability limit of (6.43). Let

$$\mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1 = \begin{bmatrix} s_{11} & \mathbf{s}_2^{'} \\ \mathbf{s}_2 & \mathbf{S}_{22} \end{bmatrix} \ . \tag{6.51}$$

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

$$
\begin{aligned}
\sum_{g,h=1}^{1+G_2} \tau_{gh}^{(k)} s_{gh} &= \tau_{11}^{(k)} s_{11} + 2\boldsymbol{\tau}_2^{(k)'} \mathbf{s}_2 + \mathbf{tr}\left[\mathbf{T}_{22}^{(k)} \mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)} s_{11} + \left(\boldsymbol{\epsilon}_k^{'} \boldsymbol{\Phi}_{22.1}^{-1} - 2\tau_{11}^{(k)} \boldsymbol{\beta}_2^{'}\right) \mathbf{s}_2 + \mathbf{tr}\left[\left(\tau_{11}^{(k)} \boldsymbol{\beta}_2 \boldsymbol{\beta}_2^{'} - \boldsymbol{\Phi}_{22.1}^{-1} \boldsymbol{\epsilon}_k \boldsymbol{\beta}_2^{'}\right) \mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)} \left[s_{11} - 2\boldsymbol{\beta}_2^{'} \mathbf{s}_2 + \boldsymbol{\beta}_2^{'} \mathbf{S}_{22} \boldsymbol{\beta}_2\right] + \boldsymbol{\epsilon}_k^{'} \boldsymbol{\Phi}_{22.1}^{-1} (\mathbf{s}_2 - \mathbf{S}_{22} \boldsymbol{\beta}_2) \\
&= \tau_{11}^{(k)} \boldsymbol{\beta}^{'} \mathbf{S} \boldsymbol{\beta} + \boldsymbol{\epsilon}_k^{'} \boldsymbol{\Phi}_{22.1}^{-1} (\mathbf{s}_2, \mathbf{S}_{22}) \boldsymbol{\beta} \ .
\end{aligned}
$$

Let

$$
\boldsymbol{\tau}_{11} = \begin{bmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{bmatrix} \tag{6.52}
$$

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

$$
\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11} \boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\right] \mathbf{S} \boldsymbol{\beta} \ . \tag{6.53}
$$

Since the asymptotic variance-covariance matrix of $\mathbf{S}\boldsymbol{\beta}$ has been obtained by the proof of *Theorem 1*, *Theorem 2* and *Lemma 6* below, we have

$$
\begin{aligned}
& \mathcal{E}\left[\hat{\mathbf{e}}\, \hat{\mathbf{e}}^{'}\right] \\
= & \left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}) \boldsymbol{\Omega} \boldsymbol{\beta}) \boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega} \boldsymbol{\beta} \boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'} \boldsymbol{\Omega} \boldsymbol{\beta}})\right] \\
& \times \mathcal{E}[\mathbf{S} \boldsymbol{\beta} \boldsymbol{\beta}^{'} \mathbf{S}] \times \left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}) \boldsymbol{\Omega} \boldsymbol{\beta}) \boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega} \boldsymbol{\beta} \boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'} \boldsymbol{\Omega} \boldsymbol{\beta}})\right]^{'} \\
= & \ \boldsymbol{\Psi}^* + \mathcal{E}\left[(\boldsymbol{\beta}^{'} \mathbf{S} \boldsymbol{\beta})^2\right] \left[\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}) \frac{1}{\sigma^2} \boldsymbol{\Omega} \boldsymbol{\beta}\right] \left[\boldsymbol{\tau}_{11}^{'} + \frac{1}{\sigma^2} \boldsymbol{\beta}^{'} \boldsymbol{\Omega} \begin{pmatrix} \mathbf{0}^{'} \\ \boldsymbol{\Phi}_{22.1}^{-1} \end{pmatrix}\right] + o(1) \ ,
\end{aligned}
$$

where $\boldsymbol{\Psi}^*$ has been given by *Corollary 1*, *Theorem 2* or *Corollary 2*.

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

$$
\boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}) \boldsymbol{\Omega} \boldsymbol{\beta} \ . \tag{6.54}
$$

Hence we have completed the proof of *Theorem 4*.

Q.E.D.

Lemma 6 : Under the assumptions of *Theorem 2*,

$$(6.55)\qquad (\mathbf{0},\mathbf{I}_{G_2})[\mathbf{I}_{G_2+1}-\frac{\boldsymbol{\Omega\beta\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega\beta}}]\mathcal{E}\left[\mathbf{S}\boldsymbol{\beta\beta}'\mathbf{S}\boldsymbol{\beta}|\mathbf{Z}\right]=o_p(1)\ .$$

Proof of Lemma 6 : We need to evaluate each term of

$$\frac{1}{n}\mathcal{E}\left\{\left[\mathbf{u}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{u}-c_*\mathbf{u}'(\mathbf{I}_n-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u}\right]\right.$$
$$\left.\times\left[\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}'\mathbf{u}+\mathbf{W}_2'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{u}-c_*\mathbf{W}_2'(\mathbf{I}_n-\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u}\right]|\mathbf{Z}\right\}\ ,$$

where $\mathbf{W}_2'=\mathbf{V}_2'-(\mathbf{0},\mathbf{I}_{G_2})\boldsymbol{\Omega\beta}\mathbf{u}'/\sigma^2$.
By using the similar calculations as (6.12)-(6.14) on the third and fourth order moments, it is equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\left(p_{ii}^{(n)}-c_*(1-q_{ii}^{(n)})\right)\mathcal{E}(u_i^3)+\frac{1}{n}\sum_{i=1}^{n}\left(p_{ii}^{(n)}-c_*(1-q_{ii}^{(n)})\right)^2\mathcal{E}(u_i^3\mathbf{w}_{2i})\ .$$

Then by using *Lemma 1*, we have the desired result.
Q.E.D

Proof of (4.6) and (4.7) : We use the arguments in a parallel way to the proof of *Theorem 1*. In the nonlinear case we set

$$\mathbf{G}=\boldsymbol{\Pi}_{2n}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\boldsymbol{\Pi}_{2n}+\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}+\boldsymbol{\Pi}_{2n}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}+\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\boldsymbol{\Pi}_{2n}$$

and

$$(6.56)\qquad \begin{aligned}\mathbf{H}\ =\ &\boldsymbol{\Pi}_{2n}'[\mathbf{I}_n-\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\boldsymbol{\Pi}_{2n}+\mathbf{V}'[\mathbf{I}_n-\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V}\\ &+\boldsymbol{\Pi}_{2n}'[\mathbf{I}_n-\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V}+\mathbf{V}'[\mathbf{I}_n-\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\boldsymbol{\Pi}_{2n}\ ,\end{aligned}$$

where $\boldsymbol{\Pi}_{2n}=\boldsymbol{\Pi}_2^{(n)}(Z)[\boldsymbol{\beta},\mathbf{I}_{G_2}]$ and $\boldsymbol{\Pi}_2^{(n)}(Z)$ is given by (4.4).
Because of Condition (VIII), $(1/q_n)\mathbf{H}-(1/q_n)\mathbf{V}'[\mathbf{I}_n-\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V}=o_p(1)$, then the essential arguments of the proof of *Theorem 1* hold. In the third case, however, we notice that the noncentrality term (i.e. the first term) of $(1/n)\mathbf{G}$ is of smaller order than the second term $(1/n)\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}$. Hence in this case because $(1/n)\mathbf{G}\xrightarrow{p}c\boldsymbol{\Omega}$ and $(1/q_n)\mathbf{H}\xrightarrow{p}\boldsymbol{\Omega}$, we find

$$(6.57)\qquad |c\boldsymbol{\Omega}-\mathrm{plim}\lambda_n\boldsymbol{\Omega}|=0$$

and hence $\text{plim}\lambda_n = c$. Then we consider

$$(6.58) \qquad \frac{n}{d_n^2}\left[(\frac{1}{n}\mathbf{G} - c\mathbf{\Omega}) - (\lambda_n - c)\mathbf{\Omega} - c(\frac{1}{q_n}\mathbf{H} - \mathbf{\Omega})\right] \text{plim}\boldsymbol{\beta}_{LI} = o_p(1) \ .$$

By evaluating each terms as in the proof of *Theorem 1*,

$$(6.59) \qquad \begin{bmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \ \text{plim}\boldsymbol{\beta}_{LI} = o_p(1)$$

and thus $\hat{\boldsymbol{\beta}}_{LI} \xrightarrow{p} \boldsymbol{\beta}$ as $n \to \infty$.

For the asymptotic normality of the LIML estimator, we use the similar arguments as (6.6)-(6.8) in the proof of *Theorem 1*. In the present case, the equation corresponding to (6.8) becomes

$$(6.60) \ (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} = \mathbf{\Phi}_{22.1}\frac{d_n^2}{\sqrt{n}}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) + o_p(1) \ ,$$

where $\mathbf{G}_1$ and $\mathbf{H}_1$ are defined in a similar way as the proof of *Theorem 1*. Because $d_n^2/n \to 0$, the first term of (6.9) converges to zero vector and $\mathbf{\Xi}_{3.2} = \mathbf{O}$ as $n \to \infty$. Then we have the result. **Q.E.D.**

References

[1] Anderson, T.W. (1976), "Estimation of Linear Functional Relationships : Approximate Distributions and Connections to Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society*, B, Vol. 38, 1-36.

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

[3] Anderson, T.W. (2005), "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics*, Vol. 127, 1-16.

[4] Anderson, T.W. and N. Kunitomo (1992), "Asymptotic Distributions of Regression and Autoregression Coefficients with Martingale Difference Disturbances," *Journal of Multivariate Analysis*, Vol. 40, 221-243.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Angrist, J. D. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, Vol. 106, 979-1014.

[10] Bekker, P.A. (1994), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, Vol. 63, 657-681.

[11] Bekker, P.A. and J. Ploeg (2005), "Instrumental Variables Estimation Based on Group Data," *Statistica Neerlandica*, Vol. 59, 3, 239-267.

[12] Bound, J., D.A. Jaeger, and R.M. Baker (1995), "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, Vol. 90, 443-450.

[13] Chao, J. and N. Swanson (2005), "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, Vol. 73, 1673-1692.

[14] Donald, S. and W. Newey (2001), "Choosing The Number of Instruments," *Econometrica*, Vol. 69-5, 1161-1191.

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

[16] Hahn, J. (2002), "Optimal Inference with Many Instruments," *Econometric Theory*, 18, 140-168.

[17] Hansen, C. J. Hausman, and Newey, W. K. (2004, 2006), "Estimation with Many Instrumental Variables," Unpublished Manuscript.

[18] Hasselt, M. V. (2006), "Estimation and Inference with Many Instruments," Unpublished Manuscript.

[19] Hayashi, F. (2000) *Econometrics*, Princeton University Press.

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

[21] Kunitomo,N. (1982), "Asymptotic Efficiency and Higher Order Efficiency of the Limited Information Maximum Likelihood Estimator in Large Econometric Models," Technical Report No. 365, Institute for Mathematical Studies in the Social Sciences, Stanford University.

[22] Kunitomo, N. (1987), "A Third Order Optimum Property of the ML Estimator in a Linear Functional Relationship Model and Simultaneous Equation System in Econometrics," *Annals of the Institute of Statistical Mathematics*, Vol.39, 575-591.

[23] Matsushita, Y. (2006), "t-Tests in a Structural Equation with Many Instruments," Discussion Paper CIRJE-F-399, Graduate School of Economics, University of Tokyo.

[24] Morimune, K. (1983), "Approximate Distribiutions of k-class Estimators When the Degree of Overidentification is Large Compared With Sample Size," *Econometrica*, Vol.51-3, 821-841.

[25] Staiger, D. and Stock, J. (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, Vol. 65, 557-586.

[26] Stock, J. and Yogo, M. (2005), "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," In *Identification and Inference for Econometric Models*, Cambridge University Press (edited by D. Andrews and J. Stock).

APPENDIX : Figures and Tables

In Figures the distribution functions of the LIML estimator are shown with the large sample normalization. The limiting distributions for the LIML estimator in the standard large asymptotics are $N(0,1)$ as $n \to \infty$, which are denoted as "o" while the limiting distributions for the LIML estimator in the large K_2 asymptotics are $N(0,a)$ ($a \geq 1$), which are denoted by the dashed curves and "x". The parameter α stands for the normalized coefficient of an endogenous variable and the details of numerical computation method are given in Anderson et al. (2005).

The tables of t-ratios include the 5, 10, 90 and 95 percentiles in one-side or two-sides, of the null-distributions for each case.

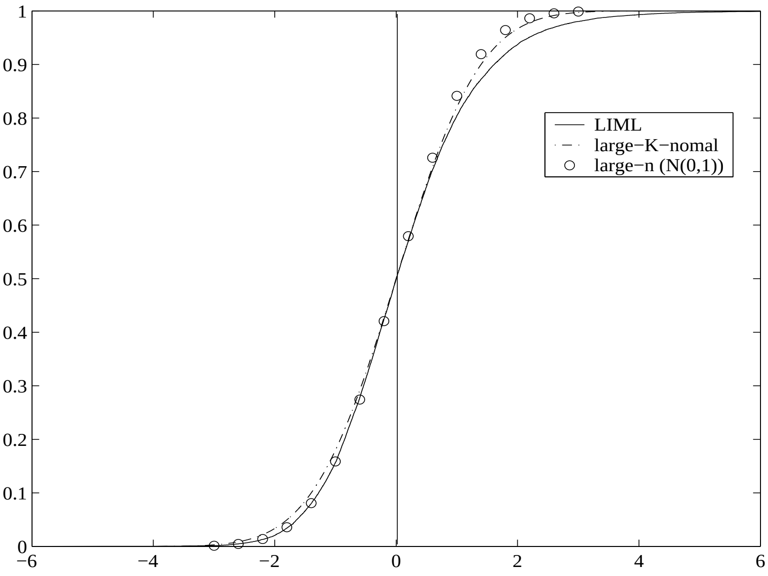

Figure 1A: CDF of Standardized LIML estimator and approximations: $n-K=30,\ K_2=5,\ \alpha=0.5,\ \delta^2=30,\ u_i=N(0,1)$

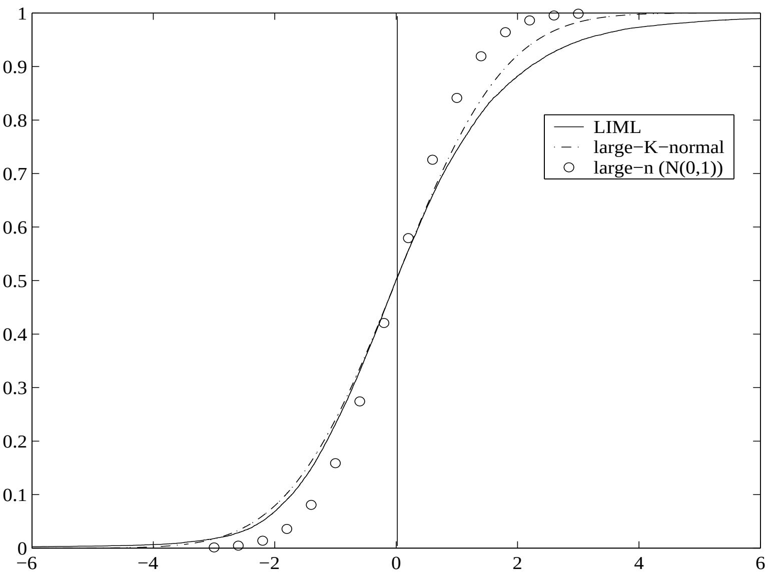

Figure 2A: CDF of Standardized LIML estimator and approximations: $n-K=30,\ K_2=30,\ \alpha=0.5,\ \delta^2=50,\ u_i=N(0,1)$

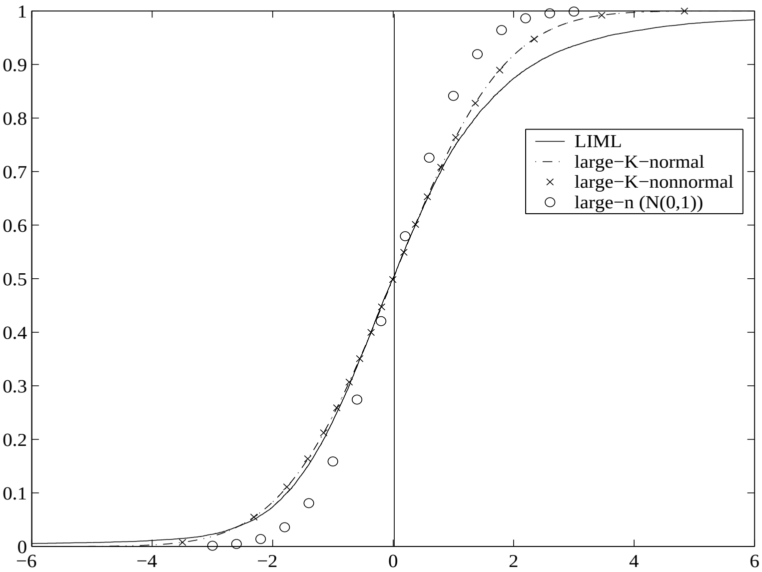

Figure 3A: CDF of Standardized LIML estimator and approximations: $n - K = 100,\ K_2 = 30,\ \alpha = 0.5,\ \delta^2 = 30,\ u_i = (\chi^2(3) - 3)/\sqrt{6}$

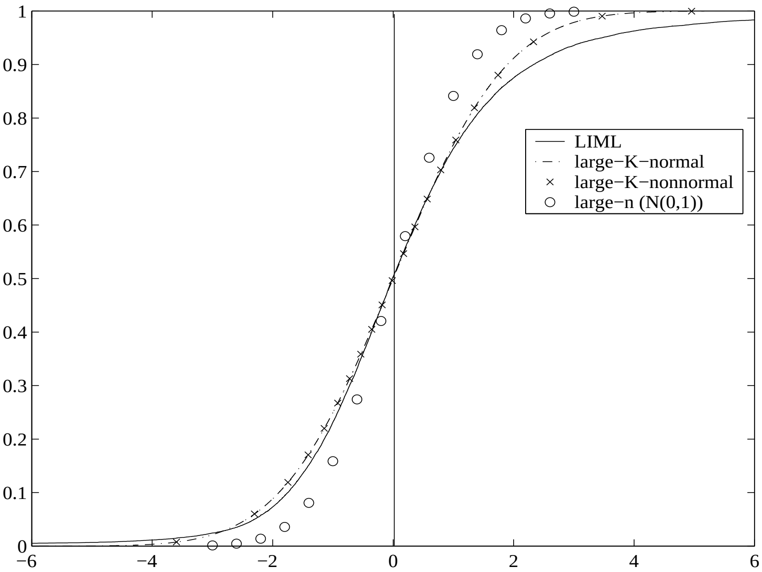

Figure 4A: CDF of Standardized LIML estimator and approximations: $n - K = 100,\ K_2 = 30,\ \alpha = 0.5,\ \delta^2 = 30,\ u_i = t(5)$

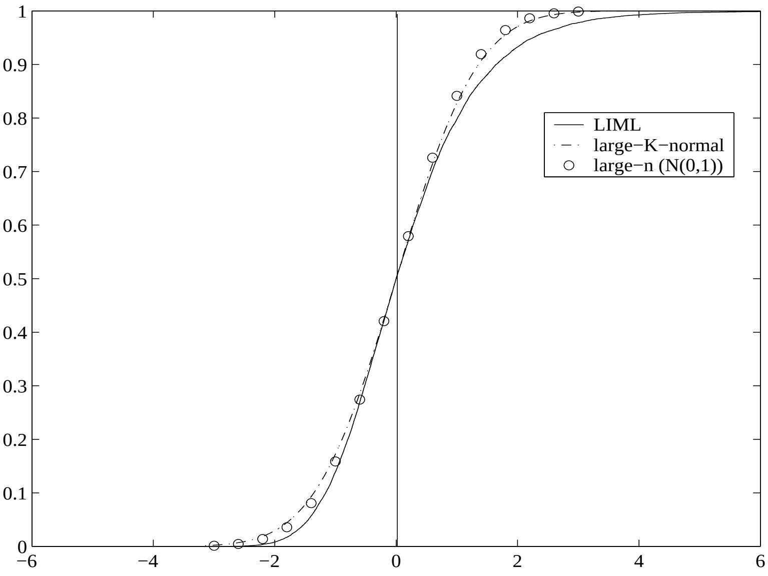

Figure 5A: CDF of Standardized LIML estimator and approximations: $n - K = 30,\ K_2 = 5,\ \alpha = 1,\ \delta^2 = 30,\ u_i = N(0,1)$

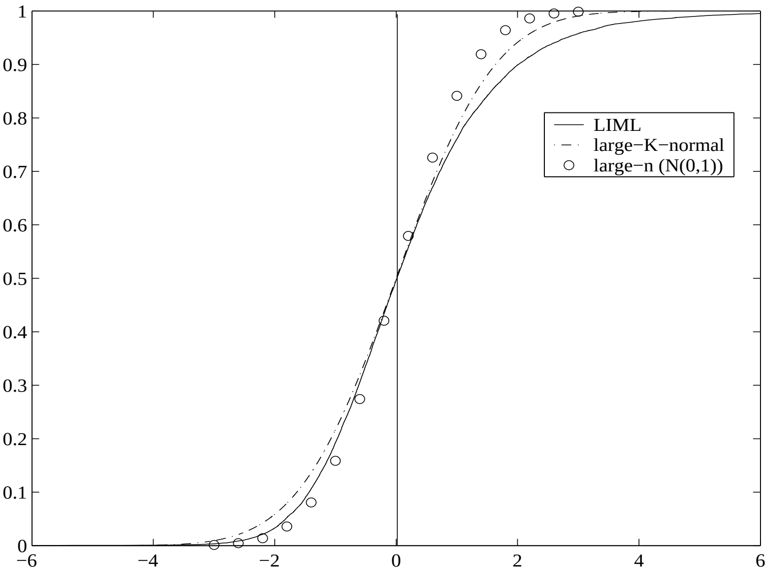

Figure 6A: CDF of Standardized LIML estimator and approximations: $n - K = 30,\ K_2 = 30,\ \alpha = 1,\ \delta^2 = 50,\ u_i = N(0,1)$

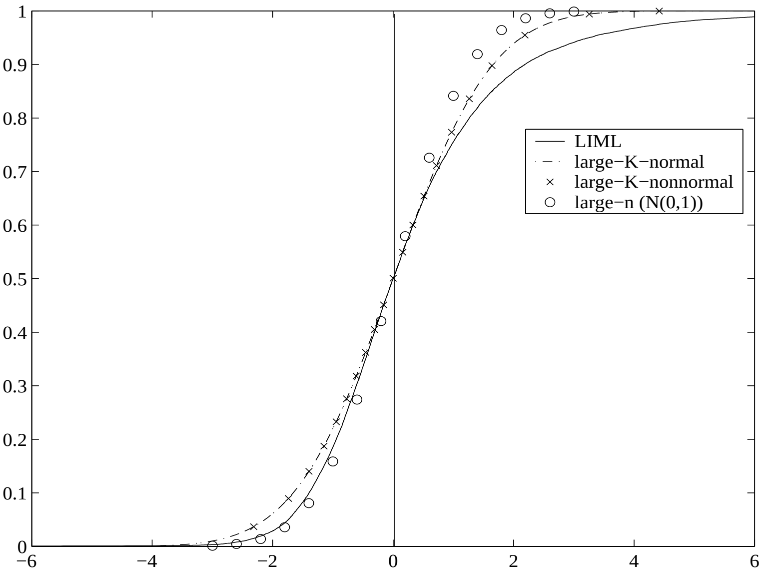

Figure 7A: CDF of Standardized LIML estimator and approximations: $n - K = 100,\ K_2 = 30,\ \alpha = 1,\ \delta^2 = 30,\ u_i = (\chi^2(3) - 3)/\sqrt{6}$

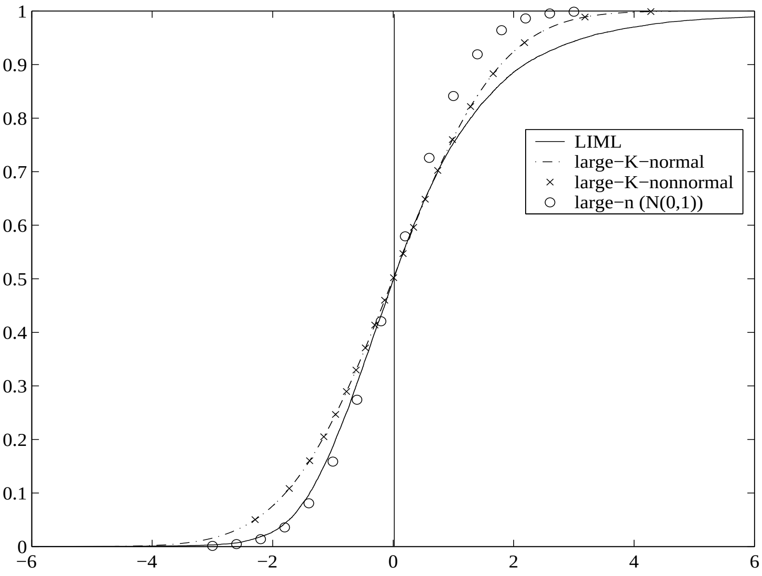

Figure 8A: CDF of Standardized LIML estimator and approximations: $n - K = 100,\ K_2 = 30,\ \alpha = 1,\ \delta^2 = 30,\ u_i = t(5)$

	normal	$t_{large-n}$	$t_{large-K}$	$t^{ellip}_{large-K}$	$t^{nonnormal}_{large-K}$
X05	-1.65	-2.67	-1.86	-1.86	-1.86
X10	-1.28	-2.07	-1.44	-1.44	-1.43
MEDN	0	-0.01	-0.01	-0.01	-0.01
X90	1.28	1.42	1.01	1.01	1.01
X95	1.65	1.68	1.24	1.24	1.23
$P(t < z_{05})$	5.0%	15.2%	7.4%	7.3%	7.4%
$P(t > z_{95})$	5.0%	5.7%	0.9%	1.0%	1.0%
$P(\|t\| > z_{975})$	5.0%	13.0%	4.4%	4.4%	4.4%
$P(\|t\| > z_{95})$	10.0%	20.8%	8.4%	8.3%	8.4%

Table 1B: Null distributions of t-ratios:
$n - K = 100,\ K_2 = 30,\ \delta^2 = 30,\ \alpha = 0.5,\ u_i = (\chi^2(3) - 3)/\sqrt{6}$

	normal	$t_{large-n}$	$t_{large-K}$	$t^{ellip}_{large-K}$	$t^{nonnormal}_{large-K}$
X05	-1.65	-2.58	-1.81	-1.81	-1.81
X10	-1.28	-2.01	-1.40	-1.40	-1.40
MEDN	0	0.00	0.00	0.00	0.00
X90	1.28	1.42	1.01	1.01	1.01
X95	1.65	1.68	1.23	1.23	1.23
$P(t < z_{05})$	5.0%	14.4%	6.8%	6.8%	6.8%
$P(t > z_{95})$	5.0%	5.5%	0.9%	0.9%	0.9%
$P(\|t\| > z_{975})$	5.0%	12.3%	3.9%	3.9%	4.0%
$P(\|t\| > z_{95})$	10.0%	20.0%	7.7%	7.6%	7.7%

Table 2B: Null distributions of t-ratios:
$n - K = 100,\ K_2 = 30,\ \delta^2 = 30,\ \alpha = 0.5,\ u_i = t(5)$

	normal	$t_{large-n}$	$t_{large-K}$	$t^{ellip}_{large-K}$	$t^{nonnormal}_{large-K}$
X05	-1.65	-2.61	-2.01	-2.01	-2.01
X10	-1.28	-1.94	-1.50	-1.49	-1.50
MEDN	0	0.01	0.00	0.00	0.00
X90	1.28	1.21	0.95	0.95	0.95
X95	1.65	1.42	1.12	1.12	1.12
$P(t < z_{05})$	5.0%	13.8%	8.2%	8.2%	8.2%
$P(t > z_{95})$	5.0%	1.8%	0.2%	0.2%	0.2%
$P(\|t\| > z_{975})$	5.0%	10.3%	5.4%	5.3%	5.3%
$P(\|t\| > z_{95})$	10.0%	15.6%	8.4%	8.4%	8.4%

Table 3B: Null distributions of t-ratios:
$n - K = 100,\ K_2 = 30,\ \delta^2 = 30,\ \alpha = 1,\ u_i = (\chi^2(3) - 3)/\sqrt{6}$

	normal	$t_{large-n}$	$t_{large-K}$	$t^{ellip}_{large-K}$	$t^{nonnormal}_{large-K}$
X05	-1.65	-2.60	-2.02	-2.02	-2.02
X10	-1.28	-1.96	-1.51	-1.51	-1.51
MEDN	0	0.00	0.00	0.00	0.00
X90	1.28	1.23	0.95	0.95	0.95
X95	1.65	1.43	1.13	1.13	1.13
$P(t < z_{05})$	5.0%	13.6%	8.4%	8.4%	8.5%
$P(t > z_{95})$	5.0%	1.8%	0.2%	0.2%	0.2%
$P(\|t\| > z_{975})$	5.0%	10.2%	5.5%	5.5%	5.5%
$P(\|t\| > z_{95})$	10.0%	15.4%	8.7%	8.6%	8.6%

Table 4B: Null distributions of t-ratios:
$n - K = 100,\ K_2 = 30,\ \delta^2 = 30,\ \alpha = 1,\ u_i = t(5)$