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

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

November 23, 2009
(Final Version)

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 as well as *old*, and we relate them to results in some recent studies. We have found that the variance of the limiting distribution 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 of some cases recently called *many instruments* and *many weak instruments*.

Key Words

Structural Equation, Simultaneous Equations System, Many Instruments, Many Weak Instruments, Limited Information Maximum Likelihood, Asymptotic Optimality; **JEL Code**: C13, C30

---

*AKM09-11-23-2. This is a revised first part 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, the co-editor and the referees of this journal for some comments to the earlier versions.

†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 a single 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, Jaeger and Baker (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, 1982) 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 Kunitomo (1981, 1982), Morimune (1983) and Bekker (1994), Chao and Swanson (2005), Van Hasselt (2006), Hansen, Hausman, and Newey (2008), and they are presented in a unified notation.

In addition to the LIML and TSLS estimators there are other instrumental variables (IV) methods. See Anderson, Kunitomo, and Sawa (1982) on the earlier studies of the finite sample properties, for instance. Several semiparametric estimation methods have been developed including the generalized method of moments (GMM) estimation and the empirical likelihood (EL) 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.

In this paper we shall give the results on the asymptotic properties of the LIML estimator when the number of instruments is large and we develop *the large-K_2 asymptotic theory* or *the many instruments asymptotics* including so-called the case of *many weak instruments*. The TSLS and the GMM estimators are badly biased and they 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 weak 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), Chao and Swanson (2005, 2006), van Hasselt (2006), van der Ploeg and Bekker (1995), Bekker and van der Ploeg (2005), Chioda and Jansson (2007), Hansen et al. (2008), and Anderson, Kunitomo and Matsushita (2008).

In Section 2 we state the formulation of a linear structural model and the alternative estimation methods of unknown parameters with possibly many instruments. In Section 3 we develop the large-K_2 asymptotics (or *many instruments* asymptotics) and give some results on the asymptotic normality of the LIML estimator when n and K_2 are large. Then we shall present some results on the asymptotic optimality of the LIML estimator in the sense that it attains the lower bound of the asymptotic variance in a class of consistent estimators with *many instruments* under reasonable assumptions. Also we discuss a more general formulation of the models and the related problems. In Section 4 we show that the asymptotic results in Section 3 agree with the finite sample properties of estimators. 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

We first consider the estimation problem of a structural equation in the classical linear simultaneous equations framework [1]. Let a single linear structural equation in an econometric model be

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

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

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

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,

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

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

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

is a positive definite matrix. 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) gives

$$\begin{pmatrix} \boldsymbol{\pi}_{11} & \mathbf{\Pi}_{12} \\ \boldsymbol{\pi}_{21}^{(n)} & \mathbf{\Pi}_{22}^{(n)} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{pmatrix} , \tag{2.3}$$

$u_i = (1, -\boldsymbol{\beta}_2^{'})\mathbf{v}_i = \boldsymbol{\beta}^{'}\mathbf{v}_i$ and $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}$ with $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$.

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

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

[1] We intentionally consider 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 3.3.

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 $\hat{\boldsymbol{\beta}}_{LI}$ 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

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

where $\mathbf{Y}_2 = (\mathbf{y}_i^{'})$ is an $n \times G_2$ matrix. The TSLS estimator 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. In this paper we shall discuss the asymptotic properties of $\hat{\boldsymbol{\beta}}_2$ because of its simplicity although it is straightforward to extend to treat $\sqrt{n}[\hat{\boldsymbol{\beta}}_2^{'} - \boldsymbol{\beta}_2^{'}, \hat{\boldsymbol{\gamma}}_1^{'} - \boldsymbol{\gamma}_1^{'}]$ with some additional notations. The LIML and TSLS estimators and their properties in the general case were originally developed by Anderson and Rubin (1949, 1950). See also Anderson (2005).

3 On Asymptotic Optimality of the LIML Estimator

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 $\mathbf{\Pi}_{2n}$ can depend on n and $n \to \infty$. We first consider the case when

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

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

where $\mathbf{\Phi}_{22.1}$ 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 is often called the case of *many instruments*. These conditions define the rates of growth of the number of incidental parameters. Condition (II) shall be weakened to Condition (II)$'$ where K_{2n} increases with n but at a smaller rate as stated in Section 3.3, which is sometimes called the case of *many weak instruments* [2] .

We first summarize the basic results on the asymptotic distributions as Theorems. Although the present formulation and *Theorem 1* are similar to the corresponding results reported in van Hasselt (2006) and Hansen et al. (2008), we shall give the proofs in Section 6 because the method of our proof gives some insights on the underlying assumptions, its extensions and the asymptotic optimality.

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

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

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

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

[2] A referee has pointed out that Condition (II) is the case of *many strong instruments*.

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.). Suppose that (I) and (II) hold. In addition assume

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

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 $c = 0$, suppose that $\mathcal{E}[\|\mathbf{v}_i\|^2]$ are bounded. Then

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

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 $\mathcal{E}(u^2 \mathbf{w}_2 \mathbf{w}_2^{'}) - \sigma^2 \mathcal{E}(\mathbf{w}_2 \mathbf{w}_2^{'}) = \boldsymbol{\Gamma}_{44.2}$ and assume that $\mathcal{E}[\|\mathbf{v}_i\|^{4+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathcal{E}[\|\boldsymbol{\Pi}_{22}^{(n)'} \mathbf{z}_{in}^{*}\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\mathbf{z}_i^{(n)}$ are stochastic) [3] . Suppose also that 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

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

where

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

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 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 $\mathcal{E}[(\boldsymbol{\beta}^{'}\mathbf{v}_i)^2 \mathbf{v}_i \mathbf{v}_i^{'}] = (1 +$

---

[3] We thank a referee for suggesting the possible improvements on the moment conditions in the previous version.

$\kappa/3)(\sigma^2\mathbf{\Omega} + 2\mathbf{\Omega}\beta\beta'\mathbf{\Omega})$, where $\mathbf{\Omega} = (\omega_{jk})$, $\mathcal{E}(v_{ji}v_{ki}) = \omega_{jk}$ and κ is the kurtosis of EC($\mathbf{\Omega}$) [4] . 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

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

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

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

A simple example for Condition (VI) is the case when we normalize $(1/n)\mathbf{A}_{22.1} \sim \mathbf{I}_{K_{2n}}$ and $\mathbf{z}_{in}^* \sim N(\mathbf{0}, \mathbf{I}_{K_{2n}})$. 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. These consequences of Condition (VI) imply the following theorem :

Theorem 2 : For $0 \le c < 1$ assume Conditions (I), (II), (III), (VI) and assume that $\mathcal{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathcal{E}[\|\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{in}^*\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\mathbf{z}_i^{(n)}$ are stochastic). Then

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

where

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

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

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

[4] The precise definition of elliptically contoured (EC) distribution has been given by Section 2.7 of Anderson (2003).

that $\sqrt{n}\|\mathbf{\Omega}_i^{(n)}-\mathbf{\Omega}\|$ is bounded and $\sigma^2=\boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}>0$. Suppose

$$(\mathbf{I}^{'}) \qquad \frac{K_{2n}}{n^{\eta}} \longrightarrow c\ (0\le\eta<1,\ 0<c<\infty),$$

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

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

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

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

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

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 $0\le\eta<1/2$,

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

It is possible to interpret the standard large sample theory as a special case of *Theorem 3*. The asymptotic properties 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) 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, the asymptotic lower bound of the covariance matrix is the same as in the case of the large sample asymptotic theory although there are incidental parameters (the number of parameters is growing with the sample size). Chao and Swanson (2005) have considered the consistency of some estimators when K_{2n} is dependent on n and the disturbances are not necessarily normally distributed. Hansen et al. (2008) have obtained the limiting distribution of the LIML estimator for a more general model with non-normal disturbances and different assumptions.

3.2 Asymptotic Optimality with Many Instruments

For the estimation of the vector of structural parameters $\boldsymbol{\beta}$, it seems natural to consider procedures based on the two $(1 + G_2) \times (1 + G_2)$ matrices $\mathbf{G}$ and $\mathbf{H}$, which are sufficient statistics in the classical standard situation. 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 LIML estimator. It contains the modified versions and the combined versions of these estimators including the one proposed by Fuller (1977). (It also includes other estimators which are asymptotically equivalent to these estimators.) Then we have a basic result on the asymptotic optimality of the LIML estimator and its (asymptotically equivalent) modifications, which attains the lower bound of the asymptotic covariance under alternative assumptions in most cases. The proof 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

$$\hat{\boldsymbol{\beta}}_2 = \phi(\mathbf{G},\ \mathbf{H})\ , \tag{3.13}$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of $(1/n)\mathbf{G}$ and $(1/q_n)\mathbf{H}$ 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*, *Theorem 2* or *Theorem 3*,

$$\mathcal{AE}\left[n(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)'\right] \ge \boldsymbol{\Psi}^*\ (\text{or } \boldsymbol{\Psi}^{**})\ , \tag{3.14}$$

and $\boldsymbol{\Psi}^*$ (or $\boldsymbol{\Psi}^{**}$) is given by (3.3), (3.6), (3.8) or (3.9), where the right-hand side of (3.14) is the covariance matrix of the limiting distribution of the normalized estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ for the class of (3.13).

This is the basic result on the asymptotic optimality of the LIML estimator when there are many instruments. 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 not only the form of (3.13). 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 eariler results given by Kunitomo (1981, 1982, 1987) because they assumed that the disturbances are normally distributed. Furthermore, Kunitomo (1987) has investigated the higher order efficiency property of the LIML estimator when $G_2 = 1$, $0 \leq c < 1$. 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] .

There are further comments on recent studies. Bekker (1994) derived the approximate distributions of the TSLS and LIML estimators under the normal disturbances. van der Ploeg and Bekker (1995) and Bekker and van der Ploeg (2005) considered the asymptotic distributions and the asymptotic bound when the data have a group structure with a fixed number m groups and the disturbances are normally distributed. Chioda and Jansson (2007) obtained an asymptotic optimality of the LIML estimator under the normal disturbances when $G_2 = 1$.

Furthermore, Hahn (2002) has derived an asymptotic lower bound when $G_2 = 1, K_1 = 0$ and the disturbances are normally distributed. It is the same as the asymptotic Cramér-Rao lower bound, which can be smaller than $\mathbf{\Psi}^{**}$ under some conditions on the incidental parameters. The trivial case which satisfies such condition is the case when a $K_{2n} \times G_1$ matrix $\mathbf{\Pi}_{22}^{(n)}$ has non-zero finite components. Another example is Condition 1 of Hahn (2002), which corresponds to the case of $0 \leq \eta < 1/2$ in our Theorem 3. These cases could be reduced to the case when $c = 0$ in Theorem 4 and there is no contradiction with our results. The main difference in two approaches come from the fact that we have pursued the asymptotic bound and optimality over the incidental parameters uniformly in some sense [6] and we do

[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 $\left[\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{\Omega}\right]_{22} \geq 0$.

[6] If we assume the normal disturbances with $\mathbf{\Omega} = \omega^2\mathbf{I}$, the dummy instruments and the j-th row of $\mathbf{\Pi}_n$ $(j = 1, \cdots, K_n)$ as $\boldsymbol{\pi}_j \sim N_{1+G_1}(\mathbf{0}, \mathbf{\Omega}_\pi)$, for instance, we have (so-called) a structural relationship model (see Anderson (1984) for instance). Then the asymptotic bound of estimators is similar to ours. A referee also pointed out the problem of *quality of instruments*, which is certainly related to the problem of choosing instruments. Since the related discussion on the incidental

not have any particular condition on the incidental parameters except (I) and (II). Since we often do not have a prior information on the reduced form coefficients in applications, it may be natural to derive the asymptotic bound with some uniformity condition over the space of reduced form coefficients.

3.3 General formulations of the asymptotic optimality

We can generalize the asymptotic optimality of LIML in several directions. We consider (2.1) and a nonlinear replacement [7] for the last G_2 columns of the reduced form (2.2). We treat (2.1) and

$$\text{(3.15)} \qquad \mathbf{Y}_2 = \mathbf{\Pi}_{2Z}^{(n)} + \mathbf{V}_2 \ ,$$

where $\mathbf{\Pi}_{2Z}^{(n)} = (\boldsymbol{\pi}_{2i}^{'}(\mathbf{z}_i^{(n)}))$ 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 (3.15) are linear, (2.1) and (3.15) has a representation (2.2). In this formulation, Condition (II) is replaced by

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

where $d_n^2 = \mathrm{tr}(\mathbf{\Pi}_{2Z}^{(n)'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{\Pi}_{2Z}^{(n)})$, $\mathbf{\Phi}_{22.1}$ is a positive (constant) definite matrix and $d_n \xrightarrow{p} \infty$ as $n \to \infty$. We replace (III) by

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

A possible additional condition (due to nonlinearilty in (3.15)) is

$$(\mathbf{VII}) \qquad \frac{1}{q_n}\mathbf{\Pi}_{2Z}^{(n)'}\left[\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\right]\mathbf{\Pi}_{2Z}^{(n)} \xrightarrow{p} \mathbf{O} \ .$$

Condition (VII) is automatically satisfied in the linear case. It is possible to weaken this condition such that the probability limit of (VII) is $\mathbf{\Phi}_3$, say. Then we need some additional notations to re-express *Theorem 1-Theorem 4* without changing their essential results. (See Section 3 of Kunitomo (2008).)

Three cases can be considered. We have already investigated the first case of $d_n = O_p(n^{1/2})$ and $K_{2n} = O(n)$. The asymptotic covariance of the LIML estimator

parameters problem and the related issues, which are important, is far beyond the scope of the present paper, we omit the details.

[7] This model is very similar to the one studied by Hansen et al. (2008), which has generalized some results of Anderson et al. (2005).

is given by (3.5) in *Theorem 1* or (3.8) in *Theorem 2* under alternative assumptions with $(\text{II})^{'}$ 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

$$d_n(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \boldsymbol{\Phi}_{22.1}^{-1}) \ . \tag{3.16}$$

Theorem 3 is one result in this case, which can be extended directly to the nonlinear model of (2.1) and (3.15).

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. (2008) with slightly different normalization and assumptions. The next result is an extension of *Theorem 4* and it could be interepreted as an asymptotic optimality of the LIML estimator for *many weak instruments*. The variance (3.19) below is simpler than (3.5) and (3.8) because the effects of n dominate the first, the third and the fourth terms of (3.5) in *Theorem 1*.

Theorem 5 : Assume Conditions $(\text{II})^{'}, (\text{III})^{'}$ and (VII) with $d_n = o_p(n^{1/2})$ and $\sqrt{n}/d_n^2 \to 0$ in *Theorem 2* and *Theorem 4* instead of Conditions (II) and (III). Then

$$\left[\frac{d_n^2}{\sqrt{n}}\right](\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{***}) \tag{3.17}$$

and

$$\mathcal{AE}\left[\left(\frac{d_n^2}{\sqrt{n}}\right)^2 (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)^{'}\right] \geq \boldsymbol{\Psi}^{***} \ , \tag{3.18}$$

where the right-hand side of (3.18) is the covariance matrix of the limiting distribution of the normalized estimator $[d_n^2/\sqrt{n}](\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ for the class of (3.13) and

$$\boldsymbol{\Psi}^{***} = \boldsymbol{\Phi}_{22.1}^{-1}\left\{c_* \left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22}\right\}\boldsymbol{\Phi}_{22.1}^{-1} \ . \tag{3.19}$$

is the covariance matrix of the limiting distribution of the LIML estimator.

3.4 Heteroscedasticity and the asymptotic properties

Recently, there has been some interest 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

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

where $\mathbf{\Omega}$ is a positive definite (constant) matrix. Then in the case when both Conditions (VI) and (VIII) 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. Chao and Swanson (2004) has investigated the JIVE estimation method, and Hausman, J., W. Newey, T. Woutersen, J. Chao and N. Swanson (2007) have proposed the HLIM estimation method. Kunitomo (2008) has also considered the class of MLIML estimators and investigated the problem of asymptotic optimality under heteroscedasticity conditions [8] . The details of these issues shall be discussed in an another occasion.

4 Discussions on Asymptotic Properties and Finite Sample Properties

It is important to investigate the finite sample properties of estimators partly because they are not necessarily similar to their asymptotic properties. One simple example would be the fact that the exact moments of some estimators do not necessarily exist. (In that case it is meaningless to compare the exact MSEs of alternative estimators and their Monte Carlo analogues.) Although we discuss the asymptotic properties of the LIML estimator, we need to investigate their relevance for practical applications.

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, 2008) in a systematic way. As typical examples we present only six figures (Figures 1A-6A) 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.

[8] It includes an extenstion of Sections 3.2 and 3.3 of this paper with the definitions of *weak heteroscedasticity* and *persistent heteroscedasticity*.

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 $\delta^2 = \mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}^{(n)}/\omega_{22}$. In addition to the LIML estimator, we have added the distribution functions of the TSLS and OLS(ordinary least squares) estimators for comparisons in some cases. (See Anderson et al. (2008) for the details.) The figures (Figures 1A-6A) show the estimated cdf of alternative estimators in the standard form, that is,

$$\frac{\sqrt{n}}{\sigma}\mathbf{\Phi}_{22.1}^{1/2}\left(\hat{\beta}_2 - \beta_2\right) . \tag{4.1}$$

By using (3.3) the limiting distributions of the LIML and TSLS 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-6A, and they are traced by the dashed curves. In some figures 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 cdfs of the estimators 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.) The distributions of the TSLS and OLS estimators have significant bias with many instruments.

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 a typical example we give Table 1A on the cdf of t-ratio

$$t(\hat{\beta}_{2.LI}) = \frac{\sqrt{n}(\hat{\beta}_{2.LI} - \beta_2)}{s(\hat{\beta}_{2.LI})} , \tag{4.2}$$

which is constructed by the LIML estimation, where $s^2(\hat{\beta}_{2.LI})$ is the estimator of the variance. The formulas (3.3), (3.4), (3.5), (3.6) and (3.8) are used. (Matsushita (2006) has investigated the finite sample properties of t-ratios and derived their asymptotic expansions of their distribution functions in a systematic way.) We can see that the effects of many instruments on the cdf of the null distributions of t-ratios are often significant while 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 among the effects of (3.5) in *Theorem 1* are not substaitial 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 (1981, 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. These observations agree with the recent studies reported by Kunitomo and Matsushita (2008), Hansen et al. (2008) and Anderson et al. (2005, 2008).

5. Concluding Remarks

In this paper, we have discussed the asymptotic optimality 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 with *many instruments* or *many weak instruments*. 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. (See Hsiao and Tahmiscioglu (2008), for instance.) We have shown that the LIML estimator and its variants have often the asymptotic optimality with many instruments.

In practical applications it is often not easy to identify whether many instruments are *weak* or *strong* such as Conditions (II) or Condition $(\mathrm{II})^{'}$. The TSLS estimator (and hence the GMM method) crucially depends on whether there are many instruments or not in any forms, while the LIML estimator does not. The only additional cost is that we need to use the asymptotic covariance, which can be slightly larger than the standard one. In this sense the LIML estimator has some asymptotic robustness.

The asymptotic optimality results in this paper give some further reasons why we have the finite sample properties of the alternative estimation methods including the classical LIML and TSLS estimators. The LIML estimator is also quite attractive over the semi-parametric estimation methods of the generalized method of moments (GMM) and the empirical likelihood (EL) estimators in the situations with *many instruments* or *many weak instruments*.

6 Proof of Theorems

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

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

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

Condition (II) implies that as $n \longrightarrow \infty$

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

and

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

Then as $n \longrightarrow \infty$,

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

and

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

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

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

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

Since $(\mathbf{G}_0 - c\,\boldsymbol{\Omega})\boldsymbol{\beta} = \mathbf{0}$ and $\hat{\boldsymbol{\beta}}_{LI}^{'} = (1, -\hat{\boldsymbol{\beta}}_{2.LI}^{'})$, (2.8) gives

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

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

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

Also multiplication of (6.6) on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ and substitution for λ_{1n} from (6.7) yields

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

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

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

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

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

we apply the central limit theorem with Lindeberg condition to the first term of (6.9). (See *Theorem 1* of Anderson and Kunitomo (1992) for instance). 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 $T_{1n} = \mathbf{a}^{'}(1/\sqrt{n})\mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}^{'}\mathbf{u}$, $T_{2n} = \mathbf{a}^{'}(1/\sqrt{K_{2n}})\mathbf{W}_2^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{u}$, $T_{3n} = \mathbf{a}^{'}(1/\sqrt{q_n})\mathbf{W}_2^{'}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\mathbf{u}$ for any constant vector $\mathbf{a}$ and

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

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

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

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

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

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

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

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 the central limit theorem (CLT) in *Lemma 3* 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.* **Q.E.D**

The next two lemmas are the results of straightforward evaluations on projection matrices, and we have omitted their derivations.

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 $q_{ij}^{(n)} = (\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}')_{ij}$.

Lemma 2 : 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

$$\sum_{i,j=1}^{n} p_{ii}p_{jj}p_{ij} \leq r \ . \tag{6.16}$$

Let also

$$\mathbf{B} = (b_{ij}) = \mathbf{Z}_{2.1}(\mathbf{Z}_{2.1}'\mathbf{Z}_{2.1})^{-1}\mathbf{Z}_{2.1}' - c_*[\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] \ . \tag{6.17}$$

Then

$$\sum_{i,j=1}^{n} b_{ii}b_{jj}b_{ij} = O(n) \ . \tag{6.18}$$

Lemma 3 : 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 $\mathbf{a}$. As $n \to \infty$,

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} t_{1i}^{(n)}u_i + \frac{1}{\sqrt{n}}\sum_{i,j=1}^{n} t_{2i}^{(n)}u_j \left[p_{ij}^{(n)} - c_*(\delta_i^j - q_{ij}^{(n)})\right] \\ &\xrightarrow{d} N(\mathbf{0}, \Delta) \ , \end{aligned} \tag{6.19}$$

where $\Delta = \mathbf{a}'\mathbf{\Phi}_{22.1}\mathbf{\Psi}^*\mathbf{\Phi}_{22.1}\mathbf{a}$ or $\Delta = \mathbf{a}'\mathbf{\Phi}_{22.1}\mathbf{\Psi}^{**}\mathbf{\Phi}_{22.1}\mathbf{a}$ in Theorems 1 and 2.

Proof of Lemma 3 : We first consider the case when $\mathbf{z}_i^{(n)}$ are the sequence of non-stochastic variables.

Set $s_{1i}^{(n)} = (1/\sqrt{n})t_{1i}^{(n)}u_i$, $s_{2i}^{(n)} = (1/\sqrt{n})t_{2i}^{(n)}u_ib_{ii}$, $s_{3i}^{(n)} = (1/\sqrt{n})u_i\sum_{j=1}^{i-1}t_{2j}^{(n)}b_{ji}$, $s_{4i}^{(n)} = (1/\sqrt{n})t_{2i}^{(n)}\sum_{j=1}^{i-1}u_jb_{ij}$ and $b_{ij} = p_{ij}^{(n)} - c_*(\delta_i^j - q_{ij}^{(n)})$ $(i, j = 1, \cdots, n)$. Let $\mathcal{F}_{n,i}$ be the $\sigma-$field generated by the random variables $u_j, \mathbf{v}_j$ $(j \leq i, i \leq n)$ and $\mathcal{F}_{n,0}$ be the initial $\sigma-$field. Then $T_n = \sum_{i=1}^{n} X_{ni}$ can be decomposed $X_{ni} = s_{1i}^{(n)} + s_{2i}^{(n)} + s_{3i}^{(n)} + s_{4i}^{(n)}$ and $\mathcal{E}[X_{ni}|\mathcal{F}_{n,i-1}] = 0$ $(i = 1, \cdots, n)$. Since each term X_{ni} $(i = 1, \cdots, n)$ are martingale difference sequences, by direct calculations we find

$$\begin{aligned} \mathcal{E}\left[X_{ni}^2|\mathcal{F}_{n,i-1}\right] &= \frac{1}{n}[t_{1i}^{(n)}]^2\sigma^2 + \frac{1}{n}b_{ii}^2\mathcal{E}[t_{2i}^{(n)}u_i]^2 + \frac{1}{n}\sigma^2[\sum_{j=1}^{i-1}b_{ij}t_{2j}]^2 + \frac{1}{n}\mathcal{E}(t_{2i}^2)[\sum_{j=1}^{i-1}b_{ij}u_j]^2 \\ &\quad + \frac{2}{n}t_{1i}^{(n)}b_{ii}\mathcal{E}[t_{2i}^{(n)}u_i^2] + \frac{2}{n}\sigma^2 t_{1i}^{(n)}\sum_{j=1}^{i-1}b_{ij}t_{2j} \\ &\quad + \frac{2}{n}\mathcal{E}(u_i^2t_{2i})b_{ii}\sum_{j=1}^{i-1}b_{ij}t_{2j} + \frac{2}{n}\mathcal{E}(u_it_{2i}^2)b_{ii}\sum_{j=1}^{i-1}b_{ij}u_j. \end{aligned}$$

Then we apply a martingale central limit theorem to $T_n = \sum_{i=1}^{n} X_{ni}$. The most important step is to show

$$\text{(6.20)} \qquad \frac{1}{n}\sum_{i=2}^{n}\left[(\sum_{j=1}^{i-1} b_{ij}t_{2j})^2 - \mathcal{E}(\sum_{j=1}^{i-1} b_{ij}t_{2j})^2\right] \xrightarrow{p} 0 \ ,$$

$$\text{(6.21)} \qquad \frac{1}{n}\sum_{i=2}^{n}\left[\mathcal{E}(t_{2i}^2)(\sum_{j=1}^{i-1} b_{ij}u_j)^2 - \mathcal{E}(t_{2i}^2)\mathcal{E}(\sum_{j=1}^{i-1} b_{ij}u_j)^2\right] \xrightarrow{p} 0 \ ,$$

$$\text{(6.22)} \qquad \frac{1}{n}\sum_{i=1}^{n}\left[t_{1i}^{(n)} b_{ii}\mathcal{E}(t_{2i}^{(n)}u_i^2) - \mathbf{a}'\mathbf{\Xi}_{3.2}\mathbf{a}\right] \xrightarrow{p} 0 \ ,$$

and

$$\text{(6.23)} \qquad \frac{1}{n}\sum_{i=2}^{n}\sigma^2 t_{1i}^{(n)}\sum_{j=1}^{i-1} b_{ij}t_{2j} \xrightarrow{p} 0 \ ,$$

$$\text{(6.24)} \qquad \frac{1}{n}\sum_{i=2}^{n}\mathcal{E}(u_i^2 t_{2i})b_{ii}\sum_{j=1}^{i-1} b_{ij}t_{2j} \xrightarrow{p} 0 \ ,$$

$$\text{(6.25)} \qquad \frac{1}{n}\sum_{i=2}^{n}\mathcal{E}(u_i t_{2i}^2)b_{ii}\sum_{j=1}^{i-1} b_{ij}u_j \xrightarrow{p} 0 \ .$$

Under the assumptions in Theorem 1, it is straightforward but tedious to show these relations by using Lemma 2. We only give an illustration. We evaluate

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{n}\sum_{i=2}^{n} t_{1i}(\sum_{j=i-1}^{n} b_{ij}t_{2i})\right]^2 &\leq (\frac{1}{n})^2\mathcal{E}\left[\sum_{i,i'=2}^{n} t_{1i}t_{1i}'\sum_{j=1}^{n} b_{ij}b_{i'j}\right]\mathcal{E}(t_{2i}^2) \\
&= (\frac{1}{n})^2\mathcal{E}\left[\sum_{i,i'=2}^{n} t_{1i}t_{1i'}b_{ii'}\right]\mathcal{E}(t_{2i}^2)
\end{aligned}$$

because we can utilize $\mathbf{Z}_{2.1}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}) = \mathbf{O}$. Then we have (6.23).

Finally, we apply the martingale CLT as Theorem 3.5 of Hall and Heyde (1980). We set $V_n = \sum_{i=1}^{n}\mathcal{E}\left[X_{ni}^2|\mathcal{F}_{n,i-1}\right]$. Then by utilizing that for any $\xi > 0$ and $\nu > 0$, $\sum_{i=1}^{n}\mathcal{E}[(X_{ni})^2 I(|X_{ni}| \geq \xi)] \leq (1/\xi)^{\nu}\sum_{i=1}^{n}\mathcal{E}[X_{ni}^{2+\nu}]$, we can show their (3.33), (3.34) and (3.36) under Assumptions of Theorem 1. Thus we have the result by using the moment condition $\mathcal{E}[\|\mathbf{v}_i\|^{4+\epsilon}] < \infty$. When Condition (**VI**) holds, we only need $\mathcal{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ because $\sum_{i=1}^{n} s_{2i}^{(n)} \xrightarrow{p} 0$. When $\mathbf{z}_i^{(n)}$ are stochastic, we utilize the additional moment condition to show the conditions of Theorem 3.5 of Hall and Heyde (1980). **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 $\sup_{1\leq i\leq n}\mathcal{E}\left[\mathbf{v}_i^{'}\mathbf{v}_i\mathbf{I}(\mathbf{v}_i^{'}\mathbf{v}_i > a)|\mathbf{z}_1^{(n)},\cdots,\mathbf{z}_n^{(n)}\right] \xrightarrow{p} 0 \;\; (a\rightarrow\infty)$. Let

$$
\begin{aligned}
\mathbf{G}_1^* &= \sqrt{n}[\frac{1}{n}\mathbf{G} - \frac{1}{n}\mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}] \\
&= \frac{1}{\sqrt{n}}\mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\ . \qquad (6.26)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathcal{E}[\frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}] &= \mathcal{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{\Omega}_i^{(n)}p_{ii}^{(n)}] \qquad (6.27)\\
&\leq \frac{K_{2n}}{\sqrt{n}}\bar{\mathbf{\Omega}} \longrightarrow \mathbf{O}
\end{aligned}
$$

when $0<\eta<1/2$. Then

$$
\mathbf{G}_1^*\boldsymbol{\beta} - \frac{1}{\sqrt{n}}\mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} = \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0}\ . \qquad (6.28)
$$

For the LIML estimator (2.8) implies

$$
(\mathbf{0},\mathbf{I}_{G_2})\left[\frac{1}{n}\mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H}\right]\begin{pmatrix}1\\ -\hat{\boldsymbol{\beta}}_{2.LI}\end{pmatrix} = \mathbf{0}\ . \qquad (6.29)
$$

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

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

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

$$
\sqrt{n}\left[\frac{1}{n}\mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n} + \frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H}\right]\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0}. \qquad (6.30)
$$

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

$$
n^{\nu}\lambda_n \xrightarrow{p} 0 \qquad (6.31)
$$

as $n\rightarrow\infty$. (ii) For $0\leq\eta<1$,

$$
\sqrt{n}\left[\lambda_n - \frac{K_{2n}}{n}\right] \xrightarrow{p} 0 \qquad (6.32)
$$

as $n \longrightarrow \infty$.

Proof of Lemma 4 : Write

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

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

$$
(6.34) \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{v}_i' \xrightarrow{p} \boldsymbol{\Omega} .
$$

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 (6.28). Then

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

as $n \to \infty$. The result (ii) follows from (6.31), (6.34) and $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta} > 0$. **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

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

We notice that

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

because Condition (II$'$) the conditions imposed on $\boldsymbol{\Omega}_i^{(n)}$ $(i = 1, \cdots, n)$.
Then by applying CLT to $(1/\sqrt{n})\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}$, we obtain the limiting normal distribution $N(\mathbf{0}, \sigma^2\boldsymbol{\Phi}_{22.1})$. This proves (i) of *Theorem 3* 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 (6.29) 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 (6.30), we have

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

Multiply (6.30) on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ to obtain

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

We consider the asymptotic behavior of the quadratic term

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

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

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

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

$$
\frac{1}{\sqrt{n}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n}\mathbf{\Omega}\right] \xrightarrow{p} \mathbf{O} \tag{6.37}
$$

when $0 \leq \eta < 1$. We can use (6.30) 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

$$
(\mathbf{0}, \mathbf{I}_{G_2})\left[\frac{1}{n}\mathbf{\Pi}_{2n}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{2n}\sqrt{n}(\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}) + \frac{1}{\sqrt{n}}\mathbf{\Pi}_{2n}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta}\right] \xrightarrow{p} \mathbf{0}\ , \tag{6.38}
$$

which is equivalent to $\mathbf{\Phi}_{22.1}\,\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) - \frac{1}{\sqrt{n}}\mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0}$. By applying the CLT to the second term, 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

$$\mathbf{G}_1^* \boldsymbol{\beta} - \left[c\boldsymbol{\Omega}\boldsymbol{\beta} + \frac{1}{\sqrt{n}} \boldsymbol{\Pi}_{2n}^{'} \mathbf{Z}_{2.1}^{'} \mathbf{V}\boldsymbol{\beta} \right] \xrightarrow{p} \mathbf{O} \ . \tag{6.39}$$

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

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

we have

$$\left[\frac{1}{n} \boldsymbol{\Pi}_{22}^{(n)'} \mathbf{A}_{22.1} \boldsymbol{\Pi}_{2n} \right] \sqrt{n}(\hat{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}) - (\mathbf{0}, \mathbf{I}_{G_2}) \mathbf{G}_1^* \boldsymbol{\beta} = o_p(1) \ . \tag{6.40}$$

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

$$\begin{aligned} & n^{1-\eta} \left[\frac{1}{n} \mathbf{G} - \frac{1}{n} \boldsymbol{\Pi}_{2n}^{'} \mathbf{A}_{22.1} \boldsymbol{\Pi}_{2n} \right] \boldsymbol{\beta} \\ = & \ \frac{K_{2n}}{n^\eta} \boldsymbol{\Omega}\boldsymbol{\beta} + \frac{1}{n^\eta} \boldsymbol{\Pi}_{2n}^{'} \mathbf{Z}_{2.1}^{'} \mathbf{V}\boldsymbol{\beta} + \frac{1}{n^\eta} \left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_{2n}\boldsymbol{\Omega} \right] \boldsymbol{\beta}. \end{aligned} \tag{6.41}$$

Because the last two terms of the right-hand side of (6.41) except the first term are of the order $o_p(n^{-\eta})$, we have

$$n^{1-\eta} \left[\frac{1}{n} \mathbf{G} - \frac{1}{n} \boldsymbol{\Pi}_{2n}^{'} \mathbf{A}_{22.1} \boldsymbol{\Pi}_{2n} \right] \boldsymbol{\beta} \xrightarrow{p} c\boldsymbol{\Omega}\boldsymbol{\beta} \tag{6.42}$$

as $n \longrightarrow \infty$. Hence by using the similar arguments as (i),

$$(\mathbf{0}, \mathbf{I}_{G_2}) \frac{1}{n} \boldsymbol{\Pi}_{2n}^{'} \mathbf{A}_{22.1} \boldsymbol{\Pi}_{22}^{(n)} \times n^{1-\eta}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2}) c\boldsymbol{\Omega}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \tag{6.43}$$

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 $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'}) = (1, -\beta_2, \cdots, -\beta_{1+G_2})$. We write

$$\hat{\beta}_k = \phi_k\left(\frac{1}{n}\mathbf{G}, \frac{1}{q_n}\mathbf{H}\right) \ \ (k = 2, \cdots, 1 + G_2) \ . \tag{6.44}$$

For the estimator to be consistent we need the conditions

$$\beta_k = \phi_k \left[\begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_{22.1} \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2} \right) + c\boldsymbol{\Omega}, \boldsymbol{\Omega} \right] \ \ (k = 2, \cdots, 1 + G_2) \tag{6.45}$$

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

$$\mathbf{T}^{(k)} = \left(\frac{\partial \phi_k}{\partial g_{ij}}\right) = (\tau_{ij}^{(k)}) \;\; (k=2,\cdots,1+G_2; i,j=1,\cdots,1+G_2) \tag{6.46}$$

evaluated at the probability limits of (6.45). We write a $(1+G_2)\times(1+G_2)$ matrix $\boldsymbol{\Theta}$ $(=(\theta_{ij}))$

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

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

$$\frac{\partial \boldsymbol{\Theta}}{\partial \beta_j} = (\frac{\partial \theta_{lm}}{\partial \beta_j}) , \tag{6.47}$$

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

$$\mathbf{tr}\left(\mathbf{T}^{(k)}\frac{\partial\boldsymbol{\Theta}}{\partial\beta_j}\right) = 2\tau_{11}^{(k)}\sum_{i=2}^{1+G_2}\rho_{ji}\beta_i + 2\sum_{i=2}^{1+G_2}\rho_{ji}\tau_{ji}^{(k)} = \delta_j^k , \tag{6.48}$$

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

$$\mathbf{T}^{(k)} = \begin{bmatrix} \tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)'} \\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix} . \tag{6.49}$$

Then (6.48) is represented as

$$2\tau_{11}^{(k)}\boldsymbol{\Phi}_{22.1}\boldsymbol{\beta} + 2\boldsymbol{\Phi}_{22.1}\boldsymbol{\tau}_2^{(k)} = \boldsymbol{\epsilon}_k , \tag{6.50}$$

where $\boldsymbol{\epsilon}_k^{'} = (0,\cdots,0,1,0,\cdots,0)$ with 1 in the k-th place and zeros in other elements. Since $\boldsymbol{\Phi}_{22.1}$ is positive definite, we solve (6.50) as

$$\boldsymbol{\tau}_2^{(k)} = \frac{1}{2}\boldsymbol{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k - \tau_{11}^{(k)}\boldsymbol{\beta}_2 . \tag{6.51}$$

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

$$\frac{\partial\boldsymbol{\Theta}}{\partial\rho_{ii}} = (\frac{\partial\theta_{lm}}{\partial\rho_{ii}}) , \tag{6.52}$$

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 $i \neq j$

$$\frac{\partial\boldsymbol{\Theta}}{\partial\rho_{ij}} = (\frac{\partial\theta_{lm}}{\partial\rho_{ij}}) , \tag{6.53}$$

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

$$\text{(6.54)} \quad \mathrm{tr}\left(\mathbf{T}^{(k)}\frac{\partial \mathbf{\Theta}}{\partial \rho_{ij}}\right) = \begin{cases} \beta_i^2\tau_{11}^{(k)} + 2\tau_{1i}^{(k)}\beta_i + \tau_{ii}^{(k)} & (i = j) \\ \\ 2\beta_i\beta_j\tau_{11}^{(k)} + 2\tau_{1j}^{(k)}\beta_i + 2\tau_{1i}^{(k)}\beta_j + 2\tau_{ij}^{(k)} & (i \neq j) \end{cases} .$$

In the matrix form we have a simple relation as

$$\text{(6.55)} \qquad \tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} + \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} + \mathbf{T}_{22}^{(k)} = \mathbf{O} \ .$$

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 (6.44). By differentiating (6.45) 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 limits. Let

$$\text{(6.56)} \qquad \mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1 = \begin{bmatrix} s_{11} & \mathbf{s}_2^{'} \\ \mathbf{s}_2 & \mathbf{S}_{22} \end{bmatrix} \ .$$

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

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

Let

$$\text{(6.57)} \qquad \boldsymbol{\tau}_{11} = \begin{bmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{bmatrix}$$

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

$$\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11} \boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}) \right] \mathbf{S}\boldsymbol{\beta} \; . \tag{6.58}$$

Since the asymptotic variance-covariance matrix of $\mathbf{S}\boldsymbol{\beta}$ has been obtained by the proof of *Theorem 1*, *Theorem 2* and *Lemma 5* 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 *Theorem 1* or *Theorem 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

$$\boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta} \;\; . \tag{6.59}$$

Hence we have completed the proof of *Theorem 4.* : **Q.E.D.**

Lemma 5 : Under the assumptions of *Theorem 2*,

$$(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}]\mathcal{E}\left[\mathbf{S}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta}|\mathbf{Z}\right] = o_p(1) \; . \tag{6.60}$$

Proof of Lemma 5 : We need to evaluate each term of

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

where $\mathbf{W}_2^{'} = \mathbf{V}_2^{'} - (\mathbf{0}, \mathbf{I}_{G_2})\boldsymbol{\Omega}\boldsymbol{\beta}\mathbf{u}^{'}/\sigma^2$.

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 Theorem 5 We use the arguments in a parallel way to the proof of *Theorem 1*. In the nonlinear case we set

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

and

$$\begin{aligned}(6.61)\qquad \mathbf{H} &= \mathbf{\Pi}_{2n}^{(z)'}[\mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{\Pi}_{2n}^{(z)} + \mathbf{V}'[\mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V} \\ &\quad +\mathbf{\Pi}_{2n}^{(z)'}[\mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V} + \mathbf{V}'[\mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{\Pi}_{2n}^{(z)} ,\end{aligned}$$

where $\mathbf{\Pi}_{2n}^{(z)} = \mathbf{\Pi}_{2Z}^{(n)}[\boldsymbol{\beta}, \mathbf{I}_{G_2}]$ and $\mathbf{\Pi}_{2Z}^{(n)}$ is given by (3.15).
Because of Condition (VII), $(1/q_n)\mathbf{H} - (1/q_n)\mathbf{V}'[\mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}']\mathbf{V} = o_p(1)$, 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 $(1/n)\mathbf{G} \xrightarrow{p} c\mathbf{\Omega}$ and $(1/q_n)\mathbf{H} \xrightarrow{p} \mathbf{\Omega}$, we find

$$(6.62)\qquad |c\mathbf{\Omega} - \text{plim}\lambda_n\mathbf{\Omega}| = 0$$

and hence $\text{plim}\lambda_n = c$. Then by using (2.8) we consider

$$(6.63)\qquad \frac{n}{d_n^2}\left[(\frac{1}{n}\mathbf{G} - c\mathbf{\Omega}) - (\lambda_n - c)\mathbf{\Omega} - c(\frac{1}{q_n}\mathbf{H} - \mathbf{\Omega})\right] \text{plim}\hat{\boldsymbol{\beta}}_{LI} = o_p(1) .$$

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

$$(6.64)\qquad \begin{bmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \text{ plim}\hat{\boldsymbol{\beta}}_{LI} = o_p(1)$$

and thus $\hat{\boldsymbol{\beta}}_{LI} \xrightarrow{p} \boldsymbol{\beta}$ as $n \to \infty$ because $\mathbf{\Phi}_{22.1}$ is nonsingular.
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

$$(6.65)\,(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} = \mathbf{\Phi}_{22.1}\frac{d_n^2}{\sqrt{n}}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) + o_p(1) ,$$

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. The proof of optimality is similar to Theorem 4, which is omitted. **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. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, Vol. 12, 1-45.

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

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

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

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

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

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

[14] 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.

[15] Chao, J. and N. Swanson (2004), "Asymptotic Distributions of JIVE in a Heteroscedastic IV Regression with many Instruments," working paper.

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

[17] Chao, J. and N. Swanson (2006), "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Instruments," in *Econometric Theory and Practice*, ed. by D. Corbae, S.Durlauf, and B. Hansen , Cambridge University Press.

[18] Chioda, J. and Jansson (2007), "Optimal Invariant Inference when the Number of Instruments is Large," Unpublished Manuscript.

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

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

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

[22] Hall, P. and C. Heyde (1980), *Martingale Limit Theory and its Applications*, Academic Press.

[23] Hansen, C. J. Hausman, and Newey, W. K. (2008), "Estimation with Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26-4, 398-422.

[24] Hausman, J., W. Newey, T. Woutersen, J. Chao and N. Swanson (2007), "Instrumental Variables Estimation with Heteroscedasticity and Many Instruments," Unpublished Manuscript.

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

[26] Hsiao, C. and A.K. Tahmiscioglu (2008), "Estimation of Dynamic Panel Data Models with Both Individual and Time Specific Effects," Unpublished Manuscript.

[27] 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.

[28] Kunitomo, N. (1981), "Asymptotic Optimality of the Limited Information Maximum Likelihood Estimator in Large Econometric Models," *The Economic Studies Quarterly*, Vol. XXXII-3, 247-266.

[29] 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.

[30] 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.

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

[32] Kunitomo,N. and Matsushita, Y. (2008), "Asymptotic Expansions and Higher Order Properties of Semi-Parametric Estimators in a Linear Simultaneous Equations," forthcoming in *Journal of Multivariate Analysis*.

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

[34] 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.

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

[36] 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).

[37] van der Ploeg, J. and Bekker, P.A. (1995), "Efficiency Bounds for Instrumental Variables Estimators under Group-Saymptotics," Unpublished Manuscript.

[38] van Hasselt, M. (2006), "Many Instruments Asymptotic Approximations under Non-normal Distributions," University of Western Ontario, Working Paper.