A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations with Possibly Many Instruments *

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

February, 2005 (1st Version)
January, 2006 (2nd Version)

Abstract

We compare four different estimation methods for a coefficient of a linear structural equation with instrumental variables. As the classical methods we consider the limited information maximum likelihood (LIML) estimator and the two-stage least squares (TSLS) estimator, and as the semi-parametric estimation methods we consider the maximum empirical likelihood (MEL) estimator and the generalized method of moments (GMM) (or the estimating equation) estimator. We prove several theorems on the asymptotic optimality of the LIML estimator when the number of instruments is large, which are new as well as *old*, and we relate them to the results in some recent studies. Tables and figures of the distribution functions of four estimators are given for enough values of the parameters to cover most of interest. We have found that the LIML estimator has good performance when the number of instruments is large, that is, the micro-econometric models with many instruments in the terminology of recent econometric literature.

Key Words

Finite Sample Properties, Maximum Empirical Likelihood, Generalized Method of Moments, Simultaneous Equations with Possibly Many Instruments, Limited Information Maximum Likelihood, Asymptotic Optimality

*This is a revised version of Discussion Paper CIRJE-F-321 under a slightly different title (Graduate School of Economics, University of Tokyo, February 2005) which was presented at the Econometric Society World Congress 2005 at London (August 2005). We thank Yoichi Arai for some comments to the earlier version.

†Department of Statistics and Department of Economics, Stanford University
‡Graduate School of Economics, University of Tokyo
§Graduate School of Economics, University of Tokyo

1. Introduction

In recent microeconometric applications some econometricians have used many instrumental variables in estimating an important structural equation. It may be partly because it has been possible to use a large number of cross sectional data and instrumental variables. One empirical example of this kind often cited in econometric literatures is Angrist and Krueger (1991) and there are some discussions by Bound et. al. (1995) since then. Because the standard text books in econometrics usually do not cover the feature that the number of instrumental variables is large, it seems that we need to investigate the basic properties of the standard estimation methods of microeconometric models in this situation. This paper will argue that a new light on the estimation of simultaneous equation models actually comes from old wisdoms in the past econometric literatures which have been often ignored and there is a strong message against some econometric methods commonly used in practice.

The study of estimating a single structural equation in econometric models has led to develop several estimation methods as the alternatives to the least squares estimation method. The classical examples in the econometric literature are the limited information maximum likelihood (LIML) method and the instrumental variables (IV) method including the two-stage least squares (TSLS) method. See Anderson and Sawa (1979), and Anderson, Kunitomo, and Sawa (1982) on the studies of their finite sample properties, for instance. As the semi-parametric estimation methods, a generalized method of moments (GMM) estimation, originally proposed by Hansen (1982), has been often used in recent econometric applications. The GMM estimation method is essentially the same as the estimating equation (EE) method originally developed by Godambe (1960) which has been mainly used in statistical applications. Also the maximum empirical likelihood (MEL) method has been proposed and has gotten some attention recently in the statistical and econometric literatures. For sufficiently large sample sizes the LIML and the TSLS estimators have approximately the same distribution in the standard large sample asymptotic theory, but their exact distributions can be quite different for the sample sizes occurring in practice. Also the GMM and the MEL estimators have approximately the same distribution under the more general heteroscedastic disturbances in the standard large sample asymptotic theory, but their exact distributions can be quite different for the sample sizes occurring in practice.

There had been alternative asymptotic theories when the number of instrumental variables is large in estimating structural equations. Kunitomo (1980, 1982), Morimune (1983), and Bekker (1994) were the earlier developers of the large K_2 asymptotic theories in the literatures. There can be some interesting aspects in these asymptotic theories in the context of simultaneous equation models because there are many instrumental variables sometimes used in micro-econometric applications and panel data analyses. The first purpose of this study is to give new results on the asymptotic optimality of the LIML estimator when the number of instruments is large. However, the TSLS and the GMM estimators lose even the consistency in some situations. Our results on the asymptotic optimality 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 microeconometric models with many instruments. There has been a growing literatures on the problem of many instruments in econometric models. We shall try to relate our new (and *old*) results to the recent studies including Donald and Newey (2001), Hahn (2002), Stock and Yogo (2003), Hansen et. al. (2004), Newey (2004), Chao and Swanson (2005), and Bekker and Ploeg (2005).

The second purpose of this study is to give information to determine the small sample properties of the exact cumulative distribution functions (cdf's) of these four different estimators for a wide range of parameter values. We shall pay a special attention on the finite sample properties of alternative estimators when we have *possibly many instruments* in the simultaneous equations. Since it is quite difficult to obtain the exact densities and cdf's of these estimators, the numerical information makes possible the comparison of properties of alternative estimation methods. We intentionally use the classical estimation setting of a linear structural equation when we have a set of instrumental variables in econometric models. It is our intention to make precise comparison of alternative estimation procedures in the possible simplest case which has many applications and it is possible to generalize our formulation into several different directions.

An important approach to the study of the finite sample properties of alternative estimators is to obtain asymptotic expansions of their exact distributions in the normalized forms. As noted before, the leading term of their asymptotic expansions in the standard large sample theory are the same, but the higher-order terms are different. For instance, Fujikoshi et. al. (1982) and their citations for the the LIML and the TSLS estimators, and Kunitomo and Matsushita (2003b) for the MEL and the GMM estimators for the linear structural equation case while Newey and Smith (2004) for the bias and the mean squared errors of estimators in the more general cases. It should be noted, however, that the mean and the mean squared errors of the exact distributions of estimators are not necessarily the same as the mean and the mean squared errors of the asymptotic expansions of the distributions of estimators. In fact the LIML estimator does not possess any moments of positive integer order under a set of reasonable assumptions. Therefore instead of moments we need to investigate the exact cumulative distributions of the LIML, MEL, GMM, and TSLS estimators directly in a systematic way. The problem of non-existence of moments had been already discussed in the econometric literature. For instance, see Mariano and Sawa (1972), Phillips (1980), and Kunitomo and Matsushita (2003a). There have been some recent studies on the computational problem on the MEL estimator by Mittelhammer et. al. (2004) and Guggenberger (2004), which are related to our results.

In Section 2 we state the formulation of models and alternative estimation methods of unknown parameters in the simultaneous equations with possibly many instruments. Then in Section 3 we develop the large K_2 asymptotics and present new results on the asymptotic variance bounds and the asymptotic optimality of the LIML estimator when the number of instruments is large in the simultaneous equations models. They give the pursuasive explanations of the finite sample properties of alternative estimation methods. In Section 4 we shall explain our tables and figures of the finite sample distributions of alternative estimators and discuss their finite sample properties. Then some conclusions will be given in Section 5. The proof of our theorems shall be given in Section 6, and Tables and Figures are gathered in Appendix.

2. Alternative Estimation Methods of a Structural Equation with Possibly Many Instruments

Let a single linear structural equation in the econometric model be given by

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

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and vector of G_2 endogenous variables, $\boldsymbol{z}_{1i}$ is a vector of K_1

included exogenous variables, $\boldsymbol{\theta}^{'} = (\boldsymbol{\gamma}_1^{'}, \boldsymbol{\beta}_2^{'})$ is a vector of $K_1 + G_2$ unknown parameters, and $\{u_i\}$ are mutually independent disturbance terms with $\mathbf{E}(u_i) = 0$ $(i = 1, \cdots, n)$. We assume that (2.1) is the first equation in a system of $(1+G_2)$ structural equations with the vector of $1+G_2$ endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$. The vector of $K(n)$ $(= K_1 + K_2(n))$ $\{\boldsymbol{z}_i(n)\}$ including $\mathbf{z}_{1i}$ is the set of instrumental variables, which satisfy the orthogonal condition $\mathbf{E}[u_i\ \boldsymbol{z}_i(n)] = \mathbf{0}$ $(i = 1, \cdots, n)$. We assume that the reduced form is given by

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi}(n) + \mathbf{V} \ , \tag{2.2}$$

where $\mathbf{Y} = (\mathbf{y}_i^{'})$ is the $n \times (1 + G_2)$ matrix of endogenous variables, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2(n))$ $(= (\mathbf{z}_i^{'}(n)))$ is the $n \times K(n)$ matrix of $(K_1 + K_2(n))$ instrument vectors $\mathbf{z}_i(n)$ $(= (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{'}(n))^{'})$ and $\boldsymbol{\Pi}(n) = (\boldsymbol{\pi}_1^{'}, \boldsymbol{\Pi}_2^{'}(n))^{'}$ is the $(K_1 + K_2(n)) \times (1 + G_2)$ matrix of coefficients. The restrictions on the coefficients can be expressed as $(1, -\boldsymbol{\beta}_2^{'})\boldsymbol{\Pi}^{'}(n) = (\boldsymbol{\gamma}_1^{'}, \mathbf{0}^{'})$ and the last $K_2(n) \times 1$ conditions can be written as $\boldsymbol{\pi}_{21}(n) = \boldsymbol{\Pi}_{22}(n)\boldsymbol{\beta}_2$, where $\boldsymbol{\pi}_{21}(n)$ and $\boldsymbol{\Pi}_{22}(n)$ are $K_2(n) \times 1$ and $K_2(n) \times G_2$ submatrices of $\boldsymbol{\Pi}_2(n)$.

Define the $(1 + G_2) \times (1 + G_2)$ random matrices be

$$\mathbf{G} = \mathbf{Y}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y} \ , \tag{2.3}$$

and

$$\mathbf{H} = \mathbf{Y}^{'}\left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\right)\mathbf{Y} \ , \tag{2.4}$$

where $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}^{'}\mathbf{Z}_{2.1}$, $\mathbf{Z}_{2.1} = \mathbf{Z}_2(n) - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$,

$$\mathbf{Z}_1 = \begin{pmatrix} \mathbf{z}_{11}^{'} \\ \vdots \\ \mathbf{z}_{1n}^{'} \end{pmatrix}, \mathbf{Z}_2(n) = \begin{pmatrix} \mathbf{z}_{21}^{'}(n) \\ \vdots \\ \mathbf{z}_{2n}^{'}(n) \end{pmatrix} \ , \tag{2.5}$$

and

$$\mathbf{A} = \mathbf{Z}^{'}\mathbf{Z} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{2.6}$$

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

$$(\mathbf{G} - \lambda\mathbf{H})\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0} \ , \tag{2.7}$$

where λ is the smallest root of

$$|\mathbf{G} - l\mathbf{H}| = 0 \tag{2.8}$$

and (2.7) corresponds to the minimum of the variance ratio

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

(2.9)

If we replace λ by 0 and omit the second component, we have the TSLS estimator $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ as

$$\mathbf{Y}_2^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y} \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{pmatrix} = \mathbf{0} \tag{2.10}$$

and it also corresponds to the solution of minimizing the numerator of the variance ratio. For the LIML and the TSLS estimators the coefficients of $\boldsymbol{\gamma}_1$ can be estimated by

$$\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Y}\hat{\boldsymbol{\beta}} , \tag{2.11}$$

where $\hat{\boldsymbol{\beta}}$ is either $\hat{\boldsymbol{\beta}}_{LI}$ or $\hat{\boldsymbol{\beta}}_{TS}$.

The maximum empirical likelihood (MEL) estimator for the vector of parameters $\boldsymbol{\theta}$ in (2.1) is defined by maximizing the Lagrange form

$$L_{2n}^{*}(\boldsymbol{\nu}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log(np_i) - \mu(\sum_{i=1}^{n} p_i - 1) - n\boldsymbol{\nu}^{'} \sum_{i=1}^{n} p_i\mathbf{z}_i(n) \left[y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}\right] ,$$

where μ and $\boldsymbol{\nu}$ are a scalar and a vector of Lagrange multipliers, and p_i $(i = 1, \cdots, n)$ is the weighted probability function to be chosen. It has been known (see Qin and Lawless (1994) or Owen (1990, 2001)) that the above maximization is the same as maximizing

$$L_{2n}(\boldsymbol{\nu}, \boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\nu}^{'}\mathbf{z}_i \left[y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}\right]\right\} , \tag{2.12}$$

where $\hat{\mu} = n$ and $[n\hat{p}_i]^{-1} = 1 + \boldsymbol{\nu}^{'}\mathbf{z}_i(n)[y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}]$. By differentiating (2.12) with respect to $\boldsymbol{\nu}$ and combining the resulting equation for $\hat{p}_i$ $(i = 1, \cdots, n)$, we have the relations $\sum_{i=1}^{n} \hat{p}_i\mathbf{z}_i(n) \left[y_{1i} - \hat{\boldsymbol{\gamma}}_1^{'}\mathbf{z}_{1i} - \hat{\boldsymbol{\beta}}_2^{'}\mathbf{y}_{2i}\right] = \mathbf{0}$ and

$$\hat{\boldsymbol{\nu}} = [\sum_{i=1}^{n} \hat{p}_i u_i^2(\hat{\boldsymbol{\theta}})\mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} u_i(\hat{\boldsymbol{\theta}})\mathbf{z}_i(n)] , \tag{2.13}$$

where $u_i(\hat{\boldsymbol{\theta}}) = y_{1i} - \hat{\boldsymbol{\gamma}}_1^{'}\mathbf{z}_{1i} - \hat{\boldsymbol{\beta}}_2^{'}\mathbf{y}_{2i}$ and $\hat{\boldsymbol{\theta}}^{'} = (\hat{\boldsymbol{\gamma}}_1^{'}, \hat{\boldsymbol{\beta}}_2^{'})$ is the maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\boldsymbol{\theta}$. Alternatively, the MEL estimator of $\boldsymbol{\theta}$ can be written as the solution of the equations $\hat{\boldsymbol{\nu}}^{'} \sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i(n)[-(\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})] = 0$, which implies

$$(2.14) \qquad [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i(n)^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(n)\ y_{1i}]$$

$$= [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i(n)^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(n)(\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})] \begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.E} \\ \hat{\boldsymbol{\beta}}_{2.E} \end{pmatrix} .$$

On the other hand, the GMM estimator of $\boldsymbol{\theta}^{'} = (\boldsymbol{\gamma}_1^{'}, \boldsymbol{\beta}_2^{'})$ can be given by the solution of the equation

$$(2.15) \qquad [\frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i(n)^{'}][\frac{1}{n}\sum_{i=1}^{n} u_i(\tilde{\boldsymbol{\theta}})^2 \mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(n)\ y_{1i}]$$

$$= [\frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i^{'}][\frac{1}{n}\sum_{i=1}^{n} u_i(\tilde{\boldsymbol{\theta}})^2 \mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(n)(\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})] \begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.G} \\ \hat{\boldsymbol{\beta}}_{2.G} \end{pmatrix} ,$$

where $\tilde{\boldsymbol{\theta}}$ is a consistent initial estimator of $\boldsymbol{\theta}$. By this representation the GMM estimator can be interpreted as the empirical likelihood estimator when we use the fixed probability weight functions as $p_i = 1/n\ (i = 1, \cdots, n)$.

By using the fact that $\log(1+x) \sim x - x^2/2$ for small x and the expression of the Lagrange multiplier vector in (2.13), it may be possible to approximate the criterion function L_{2n} as $-(1/2)$ times

$$L_{3n} = [\sum_{i=1}^{n} \mathbf{z}_i(n)^{'}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})][\sum_{i=1}^{n} \hat{p}_i u_i^2(\boldsymbol{\theta})\mathbf{z}_i(n)\mathbf{z}_i(n)^{'}]^{-1}[\sum_{i=1}^{n} \mathbf{z}_i(n)(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})] .$$

If we treat the disturbance terms as if they were homoscedastic, it may be reasonable to substitute $1/n$ for $\hat{p}_i\ (i = 1, \cdots, n)$ and $\hat{\sigma^2}$ for $\hat{u}_i^2\ (i = 1, \cdots, n)$. Then we have the degrees of freedom times the variance ratio L_{1n}.

Let the normalized error of estimators be in the form of

$$(2.16) \qquad \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix}$$

for $\hat{\boldsymbol{\theta}}' = (\hat{\boldsymbol{\gamma}}_1', \hat{\boldsymbol{\beta}}_2')$ and $\boldsymbol{\theta}' = (\boldsymbol{\gamma}_1', \boldsymbol{\beta}_2')$ is the vector of unknown coefficient parameters. In the standard large sample asymptotic theory we assume that both n and the noncentrality increase while K_2 ($= K_2(n)$) and $\mathbf{\Pi}$ ($= \mathbf{\Pi}(n)$) are fixed. Then under a set of regularity conditions [1], the asymptotic variance-covariance matrix for the GMM and the MEL estimators is given by $(\mathbf{DMC}^{-1}\mathbf{MD}')^{-1}$ while the asymptotic variance-covariance matrix for the LIML and TSLS estimators is given by $(\mathbf{DMD}')^{-1}\mathbf{DCD}'(\mathbf{DMD}')^{-1}$, where $\mathbf{M} = \mathrm{plim}_{n\to\infty}(1/n)\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'$, $\mathbf{C} = \mathrm{plim}_{n\to\infty}(1/n)\sum_{i=1}^{n}u_i^2\mathbf{z}_i\mathbf{z}_i'$, and

$$\mathbf{D} = \begin{bmatrix} (\mathbf{I}_{K_1} \;\; \mathbf{O}) \\ \mathbf{\Pi}_2' \end{bmatrix}, \tag{2.17}$$

provided that the constant matrices $\mathbf{M}$ and $\mathbf{C}$ in the probability limits are positive definite, and the rank condition is given by $\mathbf{rank}(\mathbf{D}) = K_1 + G_2$. The rank condition implies the order condition $K_2 - G_2 \geq 0$, which is the degrees of over-identification.
When $\mathbf{C} = \sigma^2\mathbf{M}$ or the disturbance terms are near homoscedastic, K_2 ($= K_2(n)$) is fixed and $(1/n)\mathbf{A}_{22.1} \xrightarrow{p} \mathbf{M}_{22.1}$ (nonsingular) as $n \to \infty$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2(\mathbf{\Pi}_{22}'\mathbf{M}_{22.1}\mathbf{\Pi}_{22})^{-1}\right] \tag{2.18}$$

for four estimation methods in the standard case, where $\mathbf{M}_{22.1} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$ and we partition the nonsingular matrix $\mathbf{M} = (\mathbf{M}_{ij})$ $(i, j = 1, 2)$.

3 Asymptotic Optimality of the LIML Estimator with Many Instruments

In the recent microeconometric models several important questions on their estimation methods for practical purposes have been raised. First, Staiger and Stock (1997) has introduced the notion of weak instruments. One interpretation of weak instruments is that we have a structural equation in which the noncentrality parameter is not large in comparison with the sample size. Second, Bekker (1994) pointed out that the standard asymptotic theory in econometrics may not be appropriate in practice when the number of instruments is large and the large-K_2 theory may be suited better to applications; see

[1] See Chamberlin (1987) for the GM estimator and Qin and Lawless (1994) for the MEL estimator, respectively.

the earlier studies of Kunitomo (1980) and Morimune (1983). There have been some microeconometric applications when many instruments have been used, but the application of the GMM method gave large biases. Third, Hansen et. al. (2004) have considered the situation when there are many weak instruments and discussed several important issues. These problems have been formulated as the situation when the number of excluded instruments is large ($K_2(n)$ is large in our notation) and comparable to the size of noncentrality parameter. It is interesting to find that it is the same situation which Kunitomo (1982, 1987) investigated under a set of limited assumptions (which could have been removed). In this section we shall develop the asymptotic theory and report new (as well as *old*) optimality results when $K_2(n)$ is dependent on the sample size n and try to relate our results to the recent studies in econometrics.

3.1 Main Results

We first state the limiting distribution of the LIML estimator under a set of alternative assumptions when $K_2(n)$ can be dependent on n and $n \to \infty$. Our result of *Theorem 1* is similar to the one in Newey (2004 unpublished), but our conditions are weaker than his conditions in the sense that we do not require any assumption on the conditional expectations. We will give its proof in Section 6.

Theorem 1 : Let $\{\mathbf{v}_i,\ \mathbf{z}_i(n)\ (i = 1, 2, \cdots)\}$ be a set of independent random vectors. Assume that (2.1) and (2.2) hold with $\mathbf{E}[\mathbf{v}_i] = \mathbf{0}$, $\mathbf{E}[\mathbf{v}_i\mathbf{v}_i^{'}] = \mathbf{\Omega}$, and $\mathbf{E}[\|\mathbf{v}_i\|^4] < \infty$. Suppose that $\mathbf{z}_1(n), \cdots, \mathbf{z}_n(n)$ are independent of $\mathbf{v}_i$ $(i = 1, \cdots, n)$. Define $q(n) = n-(K_1+K_2(n))$, and let $q(n) \longrightarrow \infty$ and $K_2(n) \longrightarrow \infty$ as $n \longrightarrow \infty$. Suppose

$$\textbf{(I)} \qquad \frac{K_2(n)}{q(n)} \longrightarrow c \ \ (0 \leq c < \infty),$$

$$\textbf{(II)} \qquad \frac{1}{K_2(n)}\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n) \stackrel{p}{\longrightarrow} \mathbf{\Phi}_{22.1}\ ,$$

$$\textbf{(III)} \qquad \frac{1}{K_2(n)} \max_{1\leq i\leq n} \|\mathbf{\Pi}_{22}^{'}(n)\mathbf{z}_i^{*}(n)\|^2 \stackrel{p}{\longrightarrow} 0\ ,$$

$$\textbf{(IV)} \qquad \max_{1\leq i\leq n} \mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*} \stackrel{p}{\longrightarrow} 0\ ,$$

where $\mathbf{z}_i^*$ is the i-th row vector of $\mathbf{Z}_{2.1}(n) = \mathbf{Z}_2(n) - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_2(n)$ and $\mathbf{\Phi}_{22.1}$ is a nonsingular constant matrix.

Then

$$\left[\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \mathbf{\Psi}^*)\ , \qquad (3.1)$$

where

$$\boldsymbol{\Psi}^{*} = \sigma^2 \mathbf{I}_{G_2} + (1+c)\boldsymbol{\Phi}_{22.1}^{-1/2}\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22}\boldsymbol{\Phi}_{22.1}^{-1/2}\ , \tag{3.2}$$

$\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}$ and $[\ \cdot\ ]_{22}$ is the $G_2 \times G_2$ lower-right corner of $(1+G_2)\times(1+G_2)$ matrix. Alternatively

$$\left[\frac{\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n)}{K_2(n)}\right]\sqrt{K_2(n)}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{**})\ , \tag{3.3}$$

where

$$\boldsymbol{\Psi}^{**} = \sigma^2\boldsymbol{\Phi}_{22.1} + (1+c)\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right]_{22}\ . \tag{3.4}$$

Alternatively

$$\sqrt{K_2(n)}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{***})\ , \tag{3.5}$$

where

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

If $G_2 = 1$, we have $[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}]_{22} = \omega_{11}\omega_{22} - \omega_{12}^2 = |\boldsymbol{\Omega}|$.

When (2.1) and (2.2) hold with $\mathbf{v}_1, \cdots, \mathbf{v}_n$ independently distributed each according to $N(\mathbf{0}, \boldsymbol{\Omega})$, then we do not need *Condition III* and *Condition IV*, which are utilized for the central limit theorems. The asymptotic variance-covariance matrix in (3.1) does not depned on the fourth order moments of disturbances under *Condition IV*.

For the estimation problem of the vector of structural parameters $\boldsymbol{\beta}$, it may be natural to consider a set of statistics of two $(1+G_2)\times(1+G_2)$ random matrices $\mathbf{G}$ and $\mathbf{H}$. Then we shall consider a class of estimators which are some functions of these two random matrices in this section and we have a new result on the asymptotic optimality of the LIML estimator under a set of simplified assumptions. The proof of *Theorem 2* will be given in Section 6.

Theorem 2 : Assume that (2.1) and (2.2) hold and define the class of consistent estimators for $\boldsymbol{\beta}_2$ by

$$\hat{\boldsymbol{\beta}}_2 = \phi(\frac{1}{K_2(n)}\mathbf{G}, \frac{1}{q(n)}\mathbf{H})\ , \tag{3.7}$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices in (3.7) as $K_2(n)\to\infty, q(n)\to\infty\ \ (n\to\infty)$. Then under the assumptions of *Theorem 1*,

$$\left[\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2)\xrightarrow{d} N(\mathbf{0},\mathbf{\Psi})\ , \tag{3.8}$$

where

$$\mathbf{\Psi}\geq\mathbf{\Psi}^{*} \tag{3.9}$$

in the sense of positive definiteness and $\mathbf{\Psi}^{*}$ is given by (3.1).

When the distribution of $\mathbf{V}$ is normal $N(\mathbf{0},\mathbf{\Omega})$ and $\mathbf{Z}$ is exogenous, $(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\mathbf{Y}$ 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 $\mathbf{\Pi}(n)$ and $\mathbf{\Omega}$, the parameters of a model. This implies that of all consistent estimators of $\boldsymbol{\beta}_2$, the LIML estimator suitably normalized has the minimum asymptotic variance. Thus the optimality of $\hat{\boldsymbol{\beta}}_{2.LI}$ extends to the class of all consistent estimators including the MEL estimator (provided that it is consistent) not only the form of (3.7).

The above theorems are the generalized versions of the results given by Kunitomo (1982) or Theorem 3.1 of Kunitomo (1987). Although they assumed that the disturbances are normally distributed and homoscedastic, it is straightforward to extend the above results to the non-normal disturbance cases as we have shown in *Theorem 1* and *Theorem 2*. Thus the essential results on the asymptotic normality as well as the asymptotic optimality of the LIML estimator do not depend on the Gaussianity. Furthermore, Kunitomo (1982) has investigated the higher order efficiency property of the LIML estimator when $G_2=1$, $c=0$ and the disturbances are normally distributed. Chao and Swanson (2005) [2] recently have investigated the consistency issue of instrumental variables methods when $K_2(n)$ is dependent on n and the disturbances are not necessarily normally distributed.

The results for the simplest case when K_2 $(=K_2(n))$ is fixed had been known over several decades since Anderson and Rubin (1950) and the more general results have been even in econometrics textbooks under the name of the standard large sample asymptotic theory for the estimation of simultaneous equations. However, it seems that in the second case, called the large K_2-asymptotic theory, the issue of asymptotic optimality has not been treated in a formal way as we did in this section. 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,

[2] Apparently they were not aware of the eariler studies.

while both the TSLS and the GMM estimators are inconsistent. This is a non-regular situation because the number of incidental parameters increases as $K_2(n)$ increases in the simultaneous equation models.

We also have the asymptotic optimality results of the LIML estimator for the cases even when $K_2(n)$ increases as $n \to \infty$ while $K_2(n)/n \to 0$. In this case the asymptotic lower bound of the covariance matrix is the same as the case of the large sample asymptotic theory. However, the limiting distribution of the LIML estimator can be different from that of the TSLS estimator and we have the next result. (The proof will be given in Section 6.)

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

$$
(\mathbf{I}^{'}) \qquad \frac{K_2(n)}{n^{\eta}} \longrightarrow c\ (0 \leq \eta < 1,\ 0 < c < \infty),
$$

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

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

where $\mathbf{z}_i^{*}(n)^{'}$ is the i-th row vector of $\mathbf{Z}_{2.1}(n) = \mathbf{Z}_2(n) - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_{2.1}(n)$.

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

$$
(3.10) \qquad \left[\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{I}_{G_2})\ ,
$$

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

$$
(3.11) \qquad \left[\frac{\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)}{n}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1})\ .
$$

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

$$
(3.12) \qquad \left[\frac{\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)}{n^{\eta}}\right](\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{p} c(\boldsymbol{\omega}_{21}, \mathbf{\Omega}_{22})\boldsymbol{\beta}\ ,
$$

and when $\eta = 1/2$

$$
(3.13) \qquad \left[\frac{\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)}{n}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{d} N\left[c(\boldsymbol{\omega}_{21}, \mathbf{\Omega}_{22})\boldsymbol{\beta}, \sigma^2\mathbf{\Phi}_{22.1}\right]\ ,
$$

where $(\omega_{21}, \mathbf{\Omega}_{22})$ is the $G_2 \times (1+G_2)$ lower submatrix of $\mathbf{\Omega}$.

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

$$\left[\frac{\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)}{n}\right] \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1}) \ . \tag{3.14}$$

It is possible to interpret the standard large sample theory as a special case of *Theorem 3* except the fact that we have used the noncentrality parameter as the normalization factor instead of $\sqrt{n}$ in (2.18). The asymptotic property of the LIML and TSLS estimators for $\boldsymbol{\gamma}_1$ can be derived from *Theorem 3*. Donald and Newey (2001) (in their *Lemma A.6*) has investigated the asymptotic properties of the LIML estimator when $K_2(n)/n \longrightarrow 0$. Also Stock and Yogo (2003), and Hansen et. al. (2004) have discussed the asymptotic properties of the GMM estimators in some cases of the large-K_2 theory when $0 < \eta < 1/2$. It seems that we need some strong conditions to establish the consistency and the asymptotic normality of the MEL estimator when both $K_2(n)$ and n increase.

Now we use our formulation to investigate the results by Hahn (2002) because he had argued that the LIML estimator is inefficient in his formulation. For this purpose define an $n \times K(n)$ $(K(n) = K_1 + K_2(n)), K_2(n) = K_{21}(n) + K_{22}(n))$ random matrix

$$\mathbf{Z} = (\mathbf{Z}^*, \mathbf{Z}_{22}^{(e)}(n)) = (\mathbf{Z}_1, \mathbf{Z}_{21}^*(n), \mathbf{Z}_{22}^{(e)}(n)) \ , \tag{3.15}$$

where $\mathbf{Z}_2(n) = (\mathbf{Z}_{21}^*(n), \mathbf{Z}_{22}^{(e)}(n))$, $\mathbf{Z}_{21}^*(n)$ is the $n \times K_{21}(n)$ matrix of instruments included in the estimation and $\mathbf{Z}_{22}^{(e)}(n)$ is the $n \times K_{22}(n)$ matrix of instruments excluded in the estimation. Also let two $(1+G_2) \times (1+G_2)$ random matrices be

$$\mathbf{G}^* = \mathbf{Y}^{'}\mathbf{Z}_{2.1}^*\mathbf{A}_{22.1}^{*-1}\mathbf{Z}_{2.1}^{*'}\mathbf{Y} \ , \tag{3.16}$$

and

$$\mathbf{H} = \mathbf{Y}^{'}\left(\mathbf{I}_n - \mathbf{Z}^*(\mathbf{Z}^{*'}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\right)\mathbf{Y} \ , \tag{3.17}$$

where $\mathbf{A}_{22.1}^* = \mathbf{Z}_{2.1}^{*'}\mathbf{Z}_{2.1}^*$, $\mathbf{Z}_{2.1}^* = \mathbf{Z}_{21}(n) - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_{21}(n)$.

Then the subset LIML (SLIML) estimator $\hat{\boldsymbol{\beta}}_{SLI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.SLI}^{'})^{'})$ for the vector of coefficients $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ with a subset of instruments $\mathbf{Z}^{(*)}$ is given by

$$(\mathbf{G}^* - \lambda^*\mathbf{H}^*)\hat{\boldsymbol{\beta}}_{SLI} = \mathbf{0} \ , \tag{3.18}$$

where λ^* is the smallest root of

$$|\mathbf{G}^* - l^*\mathbf{H}^*| = 0 \ . \tag{3.19}$$

If we replace λ^* by 0, we have the subset TSLS (STSLS) estimator with a subset of instruments $\hat{\boldsymbol{\beta}}_{STS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.STS}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$. For the SLIML estimator and STSLS estimator the coefficients of $\boldsymbol{\gamma}_1$ can be estimated by $\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Y}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is either $\hat{\boldsymbol{\beta}}_{SLI}$ or $\hat{\boldsymbol{\beta}}_{STS}$.

Theorem 4 : Let $\{\mathbf{v}_i,\ \mathbf{z}_i(n)\ (i = 1, 2, \cdots)\}$ be a sequence of independent random vectors. Assume that (2.1) and (2.2) hold with $\mathbf{E}[\mathbf{v}_i|\mathbf{z}_i(n)] = \mathbf{0}$ $(a.s.)$ and $\mathbf{E}[\mathbf{v}_i\mathbf{v}_i^{'}|\mathbf{z}_i(n)] = \boldsymbol{\Omega}_i(n)$ $(a.s.)$ is a function of $\mathbf{z}_i(n)$, say, $\boldsymbol{\Omega}_i[n, \mathbf{z}_i(n)]$. The further assumptions on $(\mathbf{v}_i, \mathbf{z}_i(n))$ are that $\max_{1\le i\le n} \mathbf{E}[v_{ij}^4|\mathbf{z}_i(n)]$ $(\mathbf{v}_i = (v_{ij}))$ are bounded, there exists a constant matrix $\boldsymbol{\Omega}$ such that $\sqrt{n}\max_{1\le i\le n} \|\boldsymbol{\Omega}_i(n) - \boldsymbol{\Omega}\|$ is bounded and $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} > 0$. Let $K_{21}(n) \longrightarrow \infty$ and $K_2(n) \longrightarrow \infty$ as $n \longrightarrow \infty$. Suppose further

$$(\mathbf{I}^*) \qquad \frac{K_{21}(n)}{\sqrt{n}} \longrightarrow 0 \ ,$$

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

$$(\mathbf{III}^*) \qquad \sqrt{n}\left[\frac{1}{n}\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{Z}_2^{'}(n)(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{Z}_2(n)\boldsymbol{\Pi}_{22}(n) - \boldsymbol{\Phi}_{22.1}\right] \stackrel{p}{\longrightarrow} \mathbf{O} \ ,$$

$$(\mathbf{IV}^*) \qquad \frac{1}{n}\max_{1\le i\le n} \|\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{Z}^{'}\mathbf{Z}^*(\mathbf{Z}^{*'}\mathbf{Z}^*)^{-1}\mathbf{z}_i^*(n)\|^2 \stackrel{p}{\longrightarrow} 0 \ ,$$

where $\mathbf{P}_{Z^*} = \mathbf{Z}^*(\mathbf{Z}^{*'}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}$, $\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}$, $\mathbf{z}_i^*(n)$ is the i-th row $(K_1 + K_{21}(n))$ vector of $\mathbf{Z}^*(n)$ and $\boldsymbol{\Phi}_{22.1}$ is a nonsingulr (constant) matrix.
Then

$$\left[\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_{2.SLI} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \sigma^2\mathbf{I}_{G_2}) \tag{3.20}$$

and

$$\left[\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_{2.STS} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \sigma^2\mathbf{I}_{G_2}) \ , \tag{3.21}$$

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

Hahn (2002) has considered the special case when $G_2 = 1$, $\mathbf{Z}_1 = \mathbf{O}$ and the disturbance terms are normally distributed. Our conditons $(I)^*$-$(III)^*$ are identical to his Condition 1 in this case and we need some additional conditions to use the central limit theorem. Because we have assumed the condition $K_{21}(n)/\sqrt{n} \longrightarrow 0$ in the present situation, $\sqrt{n}\lambda^* \xrightarrow{p} 0$ and then we have the standard efficiency bound as a simple consequence of our results. Although Hahn (2002) has argued that the LIML estimator is inefficient, our result of *Theorem 4* shows that the SLIML estimator has an asymptotic optimality if we define it in a natural way.

3.2 Discussions

Our results give some new light on the practical use of estimation methods in microeconometric models with many instruments. Since the LIML estimator has the asymptotic optimal properties when the number of instruments is large, our results in this section give the explanations of the finite sample properties of the LIML and MEL estimators. Furthermore, we shall pay an attention to the fundamental relationship betweeen the simultaneous equation sysytem and the linear functional relationship model, which gives us an important interpretation on the asymptotic behaviors of alternative estimation methods including the LIML, MEL, TSLS, and GMM estimators when there are many incidental parameters. The errors-in-variables model in the econometric literature and the linear functional relationship model in the statistical literature are mathematically equivalent to the simultaneous equations model considered here [(2.1) and (2.2)]. Such a model can be defined as follows.

Let the observed $(1+G_2)$-component vector $\mathbf{X}_{\alpha j}$ $(\alpha = 1, \cdots, K_2(n); j = 1, \cdots, m)$ be modelled as

$$\mathbf{X}_{\alpha j} = \boldsymbol{\xi}_\alpha + \boldsymbol{V}_{\alpha j} \ , \tag{3.22}$$

where $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{K_2(n)}$ are *incidental parameters*, $\mathbf{V}_{\alpha j}$ are unobserved random vectors distributed as $N(\mathbf{0}, \boldsymbol{\Omega})$, and m is the number of repeated measurements. The assumed linear relationship among $\boldsymbol{\xi}_\alpha$ is

$$\boldsymbol{\xi}_\alpha^{'}\boldsymbol{\beta} = \mathbf{0} \ , \ \alpha = 1, \cdots, K_2(n) \ . \tag{3.23}$$

Then (3.22) can be written as $\mathbf{X} = \mathbf{Z\Pi} + \mathbf{V}$, where $mK_2(n) = n$ and

$$(3.24) \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_{11}^{'} \\ \vdots \\ \mathbf{X}_{1m}^{'} \\ \mathbf{X}_{21}^{'} \\ \vdots \\ \mathbf{X}_{K_2(n)m}^{'} \end{bmatrix}, \ \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \ \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11}^{'} \\ \vdots \\ \mathbf{V}_{1m}^{'} \\ \mathbf{V}_{21}^{'} \\ \vdots \\ \mathbf{V}_{K_2(n)m}^{'} \end{bmatrix},$$

$$\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\xi}_1^{'} \\ \boldsymbol{\xi}_2^{'} \\ \vdots \\ \boldsymbol{\xi}_{K_2(n)}^{'} \end{bmatrix}.$$

The linear relationship (3.24) implies that the rank of $\mathbf{\Pi}$ is G_2 . The estimator of $\boldsymbol{\xi}_\alpha$ is $\bar{\mathbf{x}}_\alpha = (1/m)\sum_{j=1}^m \mathbf{X}_{\alpha j}$; the estimator of $\mathbf{\Pi}^{'} = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{K_2(n)})$ of unrestricted rank is $(\bar{\mathbf{x}}_1, \cdots, \bar{\mathbf{x}}_{K_2(n)})$; further

$$(3.25) \qquad \mathbf{G} = m \sum_{\alpha=1}^{K_2(n)} \bar{\mathbf{x}}_\alpha \bar{\mathbf{x}}_\alpha^{'} \, , \mathbf{H} = \sum_{\alpha=1}^{K_2(n)} \sum_{j=1}^{m} (\mathbf{x}_{\alpha j} - \bar{\mathbf{x}}_\alpha)(\mathbf{x}_{\alpha j} - \bar{\mathbf{x}}_\alpha)^{'} \, .$$

If $\boldsymbol{\beta}$ is normalized such that $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$, the maximum likelihood estimator of $\boldsymbol{\beta}$ under the normal disturbances is defined by (2.7). Also the least squares estimator by regressing the first component of $\bar{\mathbf{x}}_\alpha$ on other variables corresponds to the TSLS estimator defined by (2.10). The information matrix for $\boldsymbol{\beta}$ (or the noncentrality parameter in the structural equation estimation) under the assumption of the homoscedasticity and normality for the disturbance terms can be rewritten as

$$(3.26) \qquad \mathbf{\Theta}_m(\beta) = \frac{m}{\sigma^2} \sum_{i=1}^{K_2(n)} \boldsymbol{\xi}_i \boldsymbol{\xi}_i^{'} \, ,$$

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

The relation between the estimation problem of structural equations in econometrics and the linear functional relationships model including statistical factor analysis have been investigated by Anderson (1976, 1984). (See Sections 12 and 13 of Anderson (2003) for the details.) In the econometric literature there have been several earlier studies including Kunitomo (1980, 1982), Morimune (1983), and Bekker (1994). Anderson (1976, 1984) first showed that the TSLS estimation in the simultaneous equation models is mathematically equivalent to the least squares method in the linear functional relationship models given by (3.22) and (3.23). Bekker and Ploeg (2005) have devepoled the group asymptotics, which deals with the problems of heteroscedastic disturbances and non-identical repeated measurements, is related to our formulation.

These observations of this section give the persuasive reasons why we have finite sample properties of the LIML, MEL, TSLS, and GMM estimators as we shall see in the next section.

4. Evaluation of Exact Distribution Functions and Tables

4.1 Parameterization

The evaluation method of the cdf's of estimators we have used in this study is based on the simulation method. (See Anderson et. al. (1982) and Kunitomo and Matsushita (2003a).) In order to describe our evaluation method, we use the classical notation of Anderson et. al. (1982) for the ease of comparison except the sample size being n and we concentrate on the comparison of the estimators of the coefficient parameter of the endogenous variable when $G_2 = 1$ for the ease of interpretation. To specify the exact distributions of estimators we use the *key parameters* used by Anderson et. al. (1982) in the study of the finite sample properties of the LIML and TSLS estimators in the classical parametric framework We shall investigate the exact finite sample distributions of the normalized estimator as

$$(4.1) \qquad \frac{1}{\sigma}\left[\mathbf{\Pi}_{22}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}\right]^{1/2}(\hat{\beta}_2 - \beta_2) \ .$$

The distributions of (4.1) for the LIML estimator and TSLS estimator depend only on the key parameters used by Anderson et. al. (1982) which are K_2(fixed), $n-K$(fixed),

$$(4.2) \qquad \delta^2 = \frac{\mathbf{\Pi}_{22}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}}{\omega_{22}} \ ,$$

and

$$(4.3) \qquad \alpha = \frac{\omega_{22}\beta_2 - \omega_{12}}{|\mathbf{\Omega}|^{1/2}} = \frac{\sqrt{\omega_{22}}}{\sqrt{\omega_{11.2}}}(\beta_2 - \frac{\omega_{12}}{\omega_{22}}) \ .$$

Here ω_{12}/ω_{22} is the regression coefficient of v_{1i} on v_{2i} and $\omega_{11.2}$ is the conditional variance of v_{1i} given v_{2i} . The parameter α can be interpreted intuitively by transforming it into

$\tau = -\alpha/\sqrt{1+\alpha^2}$. Then we can rewrite $\tau = -\alpha/\sqrt{1+\alpha^2} = (\omega_{12} - \omega_{22}\beta_2)/[\sigma\sqrt{\omega_{22}}]$, which is the correlation coefficient between two random variables u_i and v_{2i} (or y_{2i}) and it is the coefficient of simultaneity in the structural equation of the simultaneous equations system. The numerator of the noncentrality parameter δ^2 represents the additional explanatory power due to $\mathbf{y}_{2i}$ over $\mathbf{z}_{1i}$ in the structural equation and its denominator is the error variance of $\mathbf{y}_{2i}$. Hence the noncentrality δ^2 determines how well the equation is defined in the simultaneous equations system, and $n-K$ is the number of degree of freedom of $\mathbf{H}$ which estimates $\mathbf{\Omega}$ in the LIML method; it is not relevant to the TSLS method.

4.2 Simulation Procedures

By using a set of Monte Carlo simulations we can obtain the empirical cdf's of estimators for the coefficient of the endogenous variable in the structural equation as follows. We generate a set of random numbers by using the two-equation system

$$y_{1i} = \gamma_1^{(0)} z_{1i} + \beta_2^{(0)} y_{2i} + u_i \ , \tag{4.4}$$

and

$$y_{2i} = \mathbf{z_i}' \boldsymbol{\pi}_2^{(0)} + v_{2i} \ , \tag{4.5}$$

where $\mathbf{z}_i \sim N(0, \mathbf{I}_K),\ u_i \sim N(0,1),\ v_{2i} \sim N(0,1)\ (i = 1, \cdots, n).$ (We set the true values of parameters $\gamma_1^{(0)} = \beta_2^{(0)} = 0$ and we have controlled the values of δ^2 by choosing a value of c and setting the $(1+K_2)$-vector $\boldsymbol{\pi}_2^{(0)} = c(1, \cdots, 1)'$. Inorder to examine whether our results strongly depend on the specific values of parameters $\gamma_1^{(0)} = \beta_2^{(0)} = 0$, however we have done several simulations for the values of $\gamma_1^{(0)} \neq 0$ and $\beta_2^{(0)} \neq 0$.) For each simulation we generated a set of random variables from the disturbance terms and exogenous variables. In the simulation the number of repetitions were 5,000 and we consider the representative situations including the corresponding cases of earlier studies.

In order to investigate the effects of nonnormal disturbances on the distributions of estimators, we used many non-normal distributions, but we only report two cases when the distributions of the disturbances are skewed or fat-tailed. As the first case we have generated a set of random variables $(y_{1i}, y_{2i}, \mathbf{z}_i)$ by using (4.4), (4.5), and $u_i = -(\chi_i^2(3) - 3)/\sqrt{6}$, and $\chi_i^2(3)$ are χ^2-random variables with 3 degrees of freedom. As the second case, we took the t-distribution with 5 degrees of freedom for the disturbance terms. Also in order to investigate the effects of heteroscedastic disturbances on the distributions of estimators, we took one example from Hayashi (2000) as an important one with $u_i = \|\mathbf{z}_i\| u_i^*$ $(i = 1, \cdots, n)$, and u_i^* $(i = 1, \cdots, n)$ are homoscedastic disturbance terms. In this case the covariance matrix $\mathbf{C}$ is not necessarily the same as $\sigma^2 \mathbf{M}$ and the asymptotic variance-covariance matrix for the LIML and TSLS estimators could be slightly larger than those of the MEL and GMM estimators in the standard large sample theory.

The empirical cdf's of estimators are consistent for the corresponding true cdf's. In addition to the empirical cdf's we have used a smoothing technique of cubic splines to estimate the cdf's and their percentile points. The distributions are tabulated in the standardized terms because this form of tabulation makes comparisons and interpolation easier. The tables includes the three quartiles, the 5 and 95 percentiles and the interquartile range of the distribution for each case, which are summarized in Tables of Appendix.

To evaluate the accuracy of our estimates based on the Monte Carlo experiments, we compared the empirical and exact cdf's of the Two-Stage Least Squares (TSLS) estimator, which corresponds to the GMM estimator given by (2.15) when $\hat{u}_i^2$ is replaced by a constant (namely σ^2), that is, the variance-covariance matrix is homoscedastic and known. The exact distribution of the TSLS estimator has been studied and tabulated extensively by Anderson and Sawa (1979). We do not report the details of our results, but we have found that the differences are less than 0.005 in most cases and the maximum difference between the exact cdf and its estimate is about 0.008. Hence our estimates of the cdf's are quite accurate and we have enough accuracy to two digits. This does not necessarily mean that the simulated moments such as the mean and the mean squared error in simulations are reliable as indicated in Introduction.

It has been known that there is a non-trivial computational problem on the MEL estimation when the noncentrality parameter is extremely near to zero. (See Mittelhammer et. al. (2004), for instance.) Therefore we have made figures to the extent that we did not have any problem in the numerical convergences. Incidentally we have found that some of our findings on the behavior of the MEL estimator were also pointed out by Guggenberger (2004).

4.3 Distributions of the MEL and LIML Estimators

For $\alpha = 0$, the densities of the LIML and MEL estimators are close to symmetric. As α increases there is some slight asymmetry, but the median is very close to zero. For given

α, K_2, and n, the lack of symmetry decreases as δ^2 increases. For given α, δ^2, and n, the asymmetry increases with K_2. The main finding from tables is that the distributions of the MEL and LIML estimators are roughly symmetric around the true parameter value and they are almost median-unbiased. This finite sample property holds even when K_2 is fairly large. At the same time, their distributions have relatively long tails. As $\delta^2 \to \infty$, the distributions approach $N(0, 1)$; however, for small values of δ^2 there is an appreciable probability outside of 3 or 4 ASD(asymptotic standard deviation)'s. (When δ^2 is extremely small, we cannot ignore the tail probabilities for practical purposes. See Table 9 and Figures 17 and 18.) As δ^2 increases, the spread of the normalized distribution decreases. Also the distribution of the LIML estimator has slightly tighter tails than that of the MEL estimator. For given α,K_2, and δ^2, the spread decreases as n increases and it tends to increase with K_2 and decrease with α.

4.4 Distributions of the GMM and TSLS Estimators

We have included tables of the distributions of the GMM and TSLS estimators. However, since they are quite similar in most cases, we have included the distribution of the GMM estimator only in many figures. The most striking feature of the distributions of the GMM and TSLS estimators is that they are skewed towards the left for $\alpha > 0$ (and towards the right for $\alpha < 0$), and the distortion increases with α and K_2 . The MEL and LIML estimators are close to median-unbiased in each case while the GMM and TSLS estimators are biased. As K_2 increases, this bias becomes more serious; for $K_2 = 10$ and $K_2 = 30$, the median is less than -1.0 ASD's. If K_2 is large, the GMM and TSLS estimators substantially underestimate the true parameter. This fact definitely favors the MEL and LIML estimators over the GMM and TSLS estimators. However, when K_2 is as small as 3, the GMM and TSLS estimators are very similar to the MEL and its distribution has tighter tails.

The distributions of the MEL and LIML estimators approach normality faster than the distribution of the GMM and TSLS estimators, due primarily to the bias of the latter. In particular when $\alpha \neq 0$ and $K_2 = 10, 30$, the actual 95 percentiles of the GMM estimator are substantially different from 1.96 of the standard normal. This implies that the conventional hypothesis testing about a structural coefficient based on the normal approximation to the distribution is very likely to seriously underestimate the actual significance. The 5 and

95 percentiles of the MEL and LIML estimators are much closer to those of the standard normal distribution even when K_2 is large.

We should note that these observations on the distributions of the MEL estimator and the GMM estimator are anologous to the earlier findings on the distributions of the LIML estimator and the TSLS estimator by Anderson et. al. (1982) and Morimune (1983) under the normal disturbances in the same setting of the linear simultaneous equations system.

4.5 Effects of Normality and Heteroscedasticity

Because the distributions of estimators depend on the distributions of the disturbance terms, we have investigated the effects of nonnormality and heteroscedasticity of disturbances. Among many tables we show only two tables and figures as the representative ones. From our tables the comparison of the distributions of four estimators are approximately valid even if the distributions of disturbances are different from normal and they are heteroscedasitic in the sense we have specified above. Thus the effects of heteroscedastic disturbances on the exact distributions of alternative estimators are not large in our setting.

5. Conclusions

First, the distributions of the MEL and GMM estimators are asymptotically equivalent in the sense of the limiting distribution in the standard large sample asymptotic theory, but their exact distributions are substantially different in finite samples. The relation of their distributions are quite similar to the distributions of the LIML and TSLS estimators. The MEL and LIML estimators are to be preferred to the GMM and TSLS estimators if K_2 is large. In some microeconometric models and models on panel data, it is often a common feature that K_2 is fairly large. For such situations we have shown that the LIML estimator has the asymptotic optimality in the large K_2-asymptotics sense. It seems that we need some stronger conditions for the MEL estimator, but its finite sample properties are often similar to the corresponding LIML estimator.

Second, the large-sample normal approximation in the large K_2 asymptotic theory is relatively accurate for the MEL and LIML estimators except the cases when we have extremely small noncentrality parameter. Hence the usual methods with asymptotic standard deviations give reasonable inferences except some extreme cases. On the other hand, for the GMM and TSLS estimators the sample size should be very large to justify the use of procedures based on the normality when K_2 is large, in particular.

Third, it is recommended to use the probability of concentration as a criterion of comparisons because the LIML estimator does not possess any moments of positive integer orders and hence we expect to have some large absolute values of their bias and mean squared errors of estimators in the Monte Carlo simulations unless we impose some restrictions on the parameter space which make it a compact set. In order to make fair comparisons of alternative estimators in a linear structural equation we need to use their culumative distribution functions and the concentration of probability. This is the reason why we directly considered the finite sample distribution functions of alternative estimation methods.

To summarize the most important conclusion from the study of small sample distributions of four alternative estimators is that the GMM and TSLS estimators can be badly biased in some cases and in that sense their use is risky. The MEL and LIML estimator, on the other hand, may have a little more variability with some chance of extreme values, but its distribution is centered at the true parameter value. The LIML estimator has tighter tails than those of the MEL estimator and in this sense the former would be attractive to the latter. Besides the computational burden for the LIML estimation is not heavy.

It is interesting that the LIML estimation was initially invented by Anderson and Rubin (1949). Other estimation methods including the TSLS, the GMM, and the MEL estimation methods have been developed with several different motivations and purposes. Now we have some practical situations in econometric applications where the LIML estimation has clear advantage over other estimation methods. It may be fair to say that *a new light has come from old wisdoms* in econometrics.

6 Proof of Theorems

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

Proof of Theorem 1 :

By substituting (2.2) into (2.3) and using the similar arguments for partitioned matrices as *Theorem A.3.3* of Anderson (2003), we have

$$\begin{aligned}\mathbf{G} &= (\mathbf{\Pi}'(n)\mathbf{Z}' + \mathbf{V}')\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'(\mathbf{Z}\mathbf{\Pi}(n) + \mathbf{V}) \\ &= \mathbf{\Pi}_2'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n) + \mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} + \mathbf{\Pi}_2'(n)\mathbf{Z}_{2.1}'\mathbf{V} + \mathbf{V}'\mathbf{Z}_{2.1}\mathbf{\Pi}_2(n) \ .\end{aligned}$$

Then we rewrite

$$\begin{aligned}&\mathbf{G} - [\mathbf{\Pi}_2'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n) + K_2(n)\mathbf{\Omega}] \\ = \ &\mathbf{\Pi}_2'(n)\mathbf{Z}_{2.1}'\mathbf{V} + \mathbf{V}'\mathbf{Z}_{2.1}\mathbf{\Pi}_2(n) + \left[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} - K_2(n)\mathbf{\Omega}\right] \ .\end{aligned} \tag{6.1}$$

By using Assumptions (II) and (IV) when $K_2(n) \longrightarrow \infty$,

$$\frac{1}{K_2(n)}\mathbf{\Pi}_2'(n)\mathbf{Z}_{2.1}'\mathbf{V} \xrightarrow{p} \mathbf{O} \ , \tag{6.2}$$

and

$$\frac{1}{K_2(n)}\left[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} - K_2(n)\mathbf{\Omega}\right] \xrightarrow{p} \mathbf{O} \ . \tag{6.3}$$

Then as $K_2(n) \to \infty$, we have the convergence in probability as

$$\frac{1}{K_2(n)}\mathbf{G} \xrightarrow{p} \mathbf{G}_0 = \begin{bmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) + \mathbf{\Omega} \tag{6.4}$$

and

$$\frac{1}{q(n)}\mathbf{H} \xrightarrow{p} \mathbf{\Omega} \ . \tag{6.5}$$

For the LIML estimation we set the smallest characteristic root and its associated vector as $|(1/K_2(n))\mathbf{G} - \lambda(n)(1/q(n))\mathbf{H}| = 0$ and

$$[\frac{1}{K_2(n)}\mathbf{G} - \lambda(n)\frac{1}{q(n)}\mathbf{H}]\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0} \ . \tag{6.6}$$

Then the probability limit of the LIML estimator $\hat{\boldsymbol{\beta}}_{LI} = (1, -\hat{\boldsymbol{\beta}}_{2.LI}')'$ is $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$ as $n \to \infty$ and $\lambda(n) \xrightarrow{p} \lambda_0$, where

$$\lambda_0\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta} = \boldsymbol{\beta}'\mathbf{G}_0\boldsymbol{\beta} \ . \tag{6.7}$$

Let $\hat{\mathbf{G}}_1 = \sqrt{K_2(n)}[(1/K_2(n))\mathbf{G} - \mathbf{G}_0]$, $\lambda_1 = \sqrt{K_2(n)}[\lambda(n) - \lambda_0]$, $\hat{\mathbf{b}}_1 = \sqrt{K_2(n)}[\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}]$, $\hat{\mathbf{H}}_1 = \sqrt{q(n)}[(1/q(n))\mathbf{H} - \mathbf{\Omega}]$. Then we can write $\hat{\mathbf{b}}_1 = (-1)(\mathbf{0}, \mathbf{I}_{G_2})'\sqrt{K_2(n)}[\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}]$. By substituing the random variables $\hat{\mathbf{G}}_1$, $\hat{\mathbf{H}}_1$, and λ_1 into (6.6), the resulting relation becomes

$$\begin{aligned}&[\mathbf{G}_0 - \lambda_0\mathbf{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{K_2(n)}}[\hat{\mathbf{G}}_1 - \lambda_1\mathbf{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{K_2(n)}}[\mathbf{G}_0 - \lambda_0\mathbf{\Omega}]\hat{\mathbf{b}}_1 + \frac{1}{\sqrt{q(n)}}[-\lambda_0\hat{\mathbf{H}}_1]\boldsymbol{\beta} \\ = \ &o_p(\frac{1}{\sqrt{K_2(n)}}) \ .\end{aligned}$$

Then by ignoring the higher order terms and using the fact $\lambda_0 = 1$, we shall consider the modified estimator $\mathbf{e}^*_{LI}(\boldsymbol{\beta})$ which satisfies

$$(6.8) \qquad [\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega}] \begin{pmatrix} \mathbf{0}' \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{e}^*_{LI}(\beta) = [\hat{\mathbf{G}}_1 - \lambda_1 \boldsymbol{\Omega}]\boldsymbol{\beta} - \sqrt{c}\hat{\mathbf{H}}_1 \boldsymbol{\beta} \; .$$

By defining the normalized (LIML) random vector $\hat{\mathbf{e}}_{LI}(\boldsymbol{\beta}) = \sqrt{K_2(n)}[\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2]$, we can show that $\mathbf{e}^*_{LI}(\beta) = \hat{\mathbf{e}}_{LI}(\beta) + o_p(1)$. By multiplying $(\mathbf{0}, \mathbf{I}_{G_2})$ and $\boldsymbol{\beta}'$ from the left-hand-side of (6.8), we have the relations

$$(6.9) \qquad (\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega}) \begin{pmatrix} \mathbf{0}' \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{e}^*_{LI}(\boldsymbol{\beta}) = (\mathbf{0}, \mathbf{I}_{G_2})(\hat{\mathbf{G}}_1 - \lambda_1 \boldsymbol{\Omega} - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta} \; ,$$

and

$$(6.10) \qquad \boldsymbol{\beta}'(\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega}) \begin{pmatrix} \mathbf{0}' \\ \mathbf{I}_{G_2} \end{pmatrix} \mathbf{e}^*_{LI}(\boldsymbol{\beta}) = \boldsymbol{\beta}'(\hat{\mathbf{G}}_1 - \lambda_1 \boldsymbol{\Omega} - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta} \; .$$

Since $(\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega})\boldsymbol{\beta} = \mathbf{0}$ and $(\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega})(\mathbf{0}, \mathbf{I}_{G_2})' = \boldsymbol{\Phi}_{22.1}$, we can simplify these relations as

$$\lambda_1 = \frac{\boldsymbol{\beta}'(\hat{\mathbf{G}}_1 - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}} \; ,$$

and then

$$(6.11) \qquad \mathbf{e}^*_{LI}(\boldsymbol{\beta}) = \left[(\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{G}_0 - \lambda_0 \boldsymbol{\Omega}) \begin{pmatrix} \mathbf{0}' \\ \mathbf{I}_{G_2} \end{pmatrix} \right]^{-1} \left[(\mathbf{0}, \mathbf{I}_{G_2})(\hat{\mathbf{G}}_1 - \lambda_1 \boldsymbol{\Omega} - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta} \right]$$

$$= \Phi^{-1}_{22.1}(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}](\hat{\mathbf{G}}_1 - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta}.$$

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

$$(6.12) \qquad \begin{aligned} & (\hat{\mathbf{G}}_1 - \sqrt{c}\hat{\mathbf{H}}_1)\boldsymbol{\beta} \\ = \; & \frac{1}{\sqrt{K_2(n)}} \boldsymbol{\Pi}'_2(n)\mathbf{Z}'_{2.1}\mathbf{u} + \frac{1}{\sqrt{K_2(n)}} \left[\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}^{-1}_{22.1}\mathbf{Z}'_{2.1}\mathbf{u} - K_2(n)\boldsymbol{\Omega}\boldsymbol{\beta} \right] \\ & -\sqrt{c}\frac{1}{\sqrt{q(n)}} \left[\mathbf{V}'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{u} - q(n)\boldsymbol{\Omega}\boldsymbol{\beta} \right] \; , \end{aligned}$$

where $K(n) + q(n) = n$. Then the asymptotic distributions of each terms on the right-hand side are normal by applying the central limit theorem with the Lindeberg conditions. (See *Theorem 1* of Anderson and Kunitomo (1992).) In order to obtain the asymptotic covariance matrix of (6.12), we use the conditional expectation given $\mathbf{Z}$ as

$$\mathbf{E}\left[\boldsymbol{\Pi}'_2(n)\mathbf{Z}'_{2.1}\mathbf{V}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{V}'\mathbf{Z}_{2.1}\boldsymbol{\Pi}_2(n)|\mathbf{Z} \right] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\Pi}'_2(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n) \; .$$

Then by using Conditions (II) and (III) in *Theorem 1*, we have the asymptotic normality for the first term of (6.12) and its asymptotic covariance matrix is given by

$$\sigma^2 \begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \ .$$

For the second and third terms of (6.12) on the right-hand side, we notice that the G_2 vector $\mathbf{w}_{2i} = \mathbf{v}_{2i} - u_i \mathbf{Cov}(\mathbf{v}_{2i}u_i)/\sigma^2$ and u_i $(i = 1, \cdots, n)$ are uncorrelated and

$$\text{(6.13)} \qquad \mathbf{E}[\mathbf{w}_{2i}\mathbf{w}_{2i}^{'}] = \frac{1}{\sigma^2}(\mathbf{0}, \mathbf{I}_{G_2})[\sigma^2\boldsymbol{\Omega} - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}] \begin{pmatrix} \mathbf{0}^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \ .$$

Then we have the representation

$$\text{(6.14)} \qquad (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}] \frac{1}{\sqrt{K_2(n)}} \left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{u} - K_2(n)\boldsymbol{\Omega}\boldsymbol{\beta} \right]$$
$$= \frac{1}{\sqrt{K_2(n)}} \left[\mathbf{V}_2^{'} - (\mathbf{0}, \mathbf{I}_{G_2})\frac{\boldsymbol{\Omega}\boldsymbol{\beta}}{\sigma^2}\mathbf{u}^{'} \right] \mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{u} \ .$$

By using *Lemma 1* below, we have the asymptotic normality of (6.14) and its asymptotic covariance matrix is given by $\sigma^2\times$ (6.13). By using the similar aruguments to the third term of (6.12) on the right-hand side, we find that the asymptotic covariance matrix of the normalized LIML estimator is given by $\boldsymbol{\Psi}^*$.
When $G_2 = 1$, we can use the relation $\sigma^2 = \omega_{11} - 2\beta_2\omega_{12} + \beta_2^2\omega_{22}$ for $\boldsymbol{\Omega} = (\omega_{ij})$ to obtain $\sigma^2\omega_{22} - (\omega_{12} - \beta\omega_{22})^2 = |\boldsymbol{\Omega}|$. **Q.E.D**

Lemma 1 : Let (u_i, w_i) $(i = 1, \cdots, n)$ be a sequence of independent random variables with $w_i = \mathbf{a}^{'}\mathbf{w}_{2i}$ for any non-zero vector $\mathbf{a}$. Assume $\mathbf{E}(u_i) = \mathbf{E}(w_i) = 0$, $\mathbf{E}(u_iw_i) = 0$, $\mathbf{E}(u_i^2) = \sigma_u^2$, $\mathbf{E}(w_i^2) = \sigma_w^2$, $\mathbf{E}(u_i^4) < \infty$, and $\mathbf{E}(w_i^4) < \infty$. Set $T(n) = \sum_{i,j=1}^n a_{ij}(n)u_iw_j$ and $a_{ij}(n) = \mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_j^*/\sqrt{K_2(n)}$. Then under Condition (IV) we have

$$\text{(6.15)} \qquad T(n) \xrightarrow{w} N[0, \sigma_u^2\sigma_w^2] \ .$$

Proof of Lemma 1 : First we evaluate the diagonal elements of $T(n)$. Since the fourth order moments are bounded, there exits a positive contant M_1 such that

$$\mathbf{E}[\sum_{i=1}^n a_{ii}(n)u_iw_i|\mathbf{Z}]^2 \le M_1 \left[\max_{1\le i\le n} \mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^* \right] \frac{1}{K_2(n)} \left[\sum_{i=1}^n \mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^* \right]$$
$$\xrightarrow{p} 0 \ .$$

Next let $T^{'}(n) = \sum_{i \neq j} a_{ij}(n) u_i w_j$. Then

$$\text{(6.16)} \qquad \begin{aligned} \mathbf{E}[T^{'}(n)^2|\mathbf{Z}] &= \sigma_u^2 \sigma_w^2 \frac{1}{K_2(n)} \sum_{i \neq j} \mathbf{z}_i^{*'} \mathbf{A}_{22.1}^{-1} \mathbf{z}_j^* \mathbf{z}_j^{*'} \mathbf{A}_{22.1}^{-1} \mathbf{z}_i^* \\ &= \sigma_u^2 \sigma_w^2 \left[1 - \frac{1}{K_2(n)} \sum_{i=1}^{n} (\mathbf{z}_i^{*'} \mathbf{A}_{22.1}^{-1} \mathbf{z}_i^*)^2 \right] \xrightarrow{p} \sigma_u^2 \sigma_w^2 \ . \end{aligned}$$

Also because the fourth order moments are bounded, there exists a positive constant M_2 such that

$$\begin{aligned} \mathbf{E}[|T^{'}(n)|^3|\mathbf{Z}] &\leq M_2 \sum_{i \neq j, i^{'} \neq j^{'}, i^{''} \neq j^{''}} |a_{ij}(n) a_{i^{'}j^{'}}(n) a_{i^{''}j^{''}}(n)| \\ &\leq M_2 \left[\max_{1 \leq i \leq n} \mathbf{z}_i^{*'} \mathbf{A}_{22.1}^{-1} \mathbf{z}_i^* \right]^{3/2} \left[\frac{1}{K_2(n)} \right]^{3/2} \left[\sum_{i=1}^{n} \mathbf{z}_i^{*'} \mathbf{A}_{22.1}^{-1} \mathbf{z}_i^* \right]^{3/2} \\ &\xrightarrow{p} 0 \ , \end{aligned}$$

where we have used the Cauchy-Schwartz inequalities in the above evaluation. Hence by using the standard arguments on the characteristic function of $T(n)$, we have $T(n) \xrightarrow{w} N(0, \sigma_u^2 \sigma_w^2)$. **Q.E.D**

Proof of Theorem 2 :

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

$$\text{(6.17)} \qquad \hat{\beta}_i = \phi_i(\frac{1}{K_2(n)} \mathbf{G}, \frac{1}{q(n)} \mathbf{H}) \ \ (i = 2, \cdots, 1 + G_2) \ .$$

For the estimator to be consistent we need the conditions

$$\text{(6.18)} \qquad \beta_i = \phi_i \left[\begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_{22.1} \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2} \right) + \boldsymbol{\Omega}, \boldsymbol{\Omega} \right] \ \ (i = 2, \cdots, 1 + G_2)$$

as identities in $\boldsymbol{\beta}_2$, $\boldsymbol{\Phi}_{22.1}$, and $\boldsymbol{\Omega}$.

Let a $(1 + G_2) \times (1 + G_2)$ matrix

$$\text{(6.19)} \qquad \mathbf{T}^{(k)} = \left(\frac{\partial \phi_k}{\partial g_{ij}} \right) = (\tau_{ij}^{(k)}) \ \ (k = 2, \cdots, 1 + G_2; i, j = 1, \cdots, 1 + G_2)$$

evaluated at the probability limits of (6.18). We write a $(1 + G_2) \times (1 + G_2)$ matrix $\boldsymbol{\Theta}$ $(= (\theta_{ij}))$

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

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

$$\frac{\partial \mathbf{\Theta}}{\partial \beta_j} = (\frac{\partial \theta_{lm}}{\partial \beta_j}) \; , \tag{6.20}$$

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

$$\mathbf{tr} \left(\mathbf{T}^{(k)} \frac{\partial \mathbf{\Theta}}{\partial \beta_j} \right) = 2 \tau_{11}^{(k)} \sum_{i=2}^{1+G_2} \rho_{ji} \beta_i + 2 \sum_{i=2}^{1+G_2} \rho_{ji} \tau_{ji}^{(k)} = \delta_j^k \; , \tag{6.21}$$

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

$$\mathbf{T}^{(k)} = \begin{bmatrix} \tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)'} \\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix} . \tag{6.22}$$

Then (6.21) is represented as

$$2 \tau_{11}^{(k)} \mathbf{\Phi}_{22.1} \boldsymbol{\beta} + 2 \mathbf{\Phi}_{22.1} \boldsymbol{\tau}_2^{(k)} = \boldsymbol{\epsilon}_k \; , \tag{6.23}$$

where $\boldsymbol{\epsilon}_k' = (0, \cdots, 0, 1, 0, \cdots, 0)$ with 1 in the k-th place and zeros in other elements.
Since $\mathbf{\Phi}_{22.1}$ is positive definite, we solve (6.23) as

$$\boldsymbol{\tau}_2^{(k)} = \frac{1}{2} \mathbf{\Phi}_{22.1}^{-1} \boldsymbol{\epsilon}_k - \tau_{11}^{(k)} \boldsymbol{\beta}_2 \; . \tag{6.24}$$

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

$$\frac{\partial \mathbf{\Theta}}{\partial \rho_{ii}} = (\frac{\partial \theta_{lm}}{\partial \rho_{ii}}) \; , \tag{6.25}$$

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

$$\frac{\partial \mathbf{\Theta}}{\partial \rho_{ij}} = (\frac{\partial \theta_{lm}}{\partial \rho_{ij}}) \; , \tag{6.26}$$

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

$(2 \le l, m \le 1 + G_2)$.

Then we have the representation

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

In the matrix form we have a simple relation as

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

Then we have the representation

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

Next we consider the role of the second matrix in (6.17). By differentiating (6.18) with respect to ω_{ij} $(i, j = 1, \cdots, 1 + G_2)$, we have the condition

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

evaluated at the probability limit of (6.18). Let

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

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

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

Let

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

and we consider the asymptotic behavihor of the normalized estimator $\sqrt{K_2(n)}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ as

$$\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\right]\mathbf{S}\boldsymbol{\beta}\ . \tag{6.31}$$

Since the asymptotic variance-covariance matrix of $\mathbf{S}\boldsymbol{\beta}$ has been obtained from the proof of *Theorem 1*, we have

$$\begin{aligned}
&\mathbf{E}\left[\hat{\mathbf{e}}\,\hat{\mathbf{e}}^{'}\right] \\
= &\left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\right]\left[\sigma^2\begin{pmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) + (1+c)\left(\sigma^2\boldsymbol{\Omega} + \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right)\right] \\
&\times\left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\right]^{'} \\
= &\ 2(1+c)\sigma^4\boldsymbol{\tau}_{11}\boldsymbol{\tau}_{11}^{'} \\
&+\left[2(1+c)\sigma^2\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\begin{pmatrix}\mathbf{0}^{'}\\ \boldsymbol{\Phi}_{22.1}^{-1}\end{pmatrix}\right] + \left[2(1+c)\sigma^2\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\begin{pmatrix}\mathbf{0}^{'}\\ \boldsymbol{\Phi}_{22.1}^{-1}\end{pmatrix}\right]^{'} \\
&+\left(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1}\right)\left[\sigma^2\begin{pmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) + (1+c)\left(\sigma^2\boldsymbol{\Omega} + \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\right)\right]\begin{pmatrix}\mathbf{0}^{'}\\ \boldsymbol{\Phi}_{22.1}^{-1}\end{pmatrix} \\
= &\ \boldsymbol{\Psi}^{*} + 2(1+c)\left[\sigma^2\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta}\right]\left[\sigma^2\boldsymbol{\tau}_{11}^{'} + \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\begin{pmatrix}\mathbf{0}^{'}\\ \boldsymbol{\Phi}_{22.1}^{-1}\end{pmatrix}\right]\quad ,
\end{aligned}$$

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

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

$$\boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta}\ . \tag{6.32}$$

Hence we have completed the proof of *Theorem 2*.

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 $K_2(n)$, respectively, and that the boundedness of $\mathbf{E}[v_{ij}^4|\mathbf{z}_i(n)]$ implies a Lindeberg condition

$$\sup_{i=1,\cdots,n}\mathbf{E}\left[\mathbf{v}_i^{'}\mathbf{v}_i\mathbf{I}(\mathbf{v}_i^{'}\mathbf{v}_i > a)|\mathbf{z}_1(n), \cdots, \mathbf{z}_n(n)\right] \xrightarrow{p} 0\ \ (a \to \infty)\ . \tag{6.33}$$

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

When $0 \leq \eta < 1/2$, we set

$$
\begin{aligned}
\hat{\mathbf{G}}_1^*(n) &= \sqrt{n}[\frac{1}{n}\hat{\mathbf{G}} - \frac{1}{n}\mathbf{\Pi}_2^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n)] \\
&= \frac{1}{\sqrt{n}}\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_{2.1}^{'}\mathbf{V} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{\Pi}_2(n) + \frac{1}{\sqrt{n}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\right] \ .
\end{aligned} \tag{6.34}
$$

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

$$
\begin{aligned}
\mathbf{E}[\frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}] &= \mathbf{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{\Omega}_i(n)\mathbf{z}_i^{*'}(n)\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*}(n)] \\
&\leq \frac{K_2(n)}{\sqrt{n}}\bar{\mathbf{\Omega}} \longrightarrow \mathbf{O} \ .
\end{aligned} \tag{6.35}
$$

Then

$$
\hat{\mathbf{G}}_1^*(n)\boldsymbol{\beta} - \frac{1}{\sqrt{n}}\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} \stackrel{p}{\rightarrow} \mathbf{0} \ . \tag{6.36}
$$

Lemma 2 : Let $\lambda(n)$ be the smallest characteristic root of

$$
|\frac{1}{n}\mathbf{G} - l^*\frac{1}{q(n)}\mathbf{H}| = 0 \ .
$$

For $0 < \nu < 1 - \eta$ and $0 \leq \eta < 1$,

$$
n^{\nu}\lambda(n) \stackrel{p}{\longrightarrow} 0 \tag{6.37}
$$

as $n \rightarrow \infty$.

Proof of Lemma 2

Write

$$
\begin{aligned}
\lambda(n) &= \min_{\mathbf{b}} \frac{\mathbf{b}^{'}\frac{1}{n}\mathbf{G}\mathbf{b}}{\mathbf{b}^{'}\frac{1}{q(n)}\mathbf{H}\mathbf{b}} \\
&\leq \frac{q(n)}{n}\frac{\boldsymbol{\beta}^{'}\mathbf{G}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{H}\boldsymbol{\beta}} = \frac{q(n)}{n}\frac{\boldsymbol{\beta}^{'}\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{V}^{'}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'})\mathbf{V}\boldsymbol{\beta}} \ .
\end{aligned} \tag{6.38}
$$

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

$$
\frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i\mathbf{v}_i^{'} \stackrel{p}{\rightarrow} \mathbf{\Omega} \ . \tag{6.39}
$$

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 (6.35). Then

$$n^{\nu}\lambda(n) \leq \left[\frac{q(n)}{n}\right] \frac{n^{-(1-\nu)}\boldsymbol{\beta}'\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}}{n^{-1}\boldsymbol{\beta}'\mathbf{V}'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{V}\boldsymbol{\beta}} \xrightarrow{p} 0 \tag{6.40}$$

as $n \to \infty$.

Q.E.D.

We consider

$$(\mathbf{0}, \mathbf{I}_{G_2}) \left[\frac{1}{n}\boldsymbol{\Pi}_2'(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n) + \frac{1}{\sqrt{n}}\hat{\mathbf{G}}_1^*(n) - \lambda(n)\frac{1}{q(n)}\mathbf{H}\right] \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} = \mathbf{0} \tag{6.41}$$

for the LIML estimator. By using the facts that $(1/\sqrt{n})\hat{\mathbf{G}}_1^* \xrightarrow{p} \mathbf{O}$, $\lambda(n) \xrightarrow{p} 0$ and $[1/q(n)]\mathbf{H} \xrightarrow{p} \boldsymbol{\Omega}$, we have

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

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

$$\sqrt{n}\left[\frac{1}{n}\boldsymbol{\Pi}_2'(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n) + \frac{1}{\sqrt{n}}\hat{\mathbf{G}}_1^*(n) - \lambda(n)\frac{1}{q(n)}\mathbf{H}\right]\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0}. \tag{6.42}$$

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

$$(\mathbf{0}, \mathbf{I}_{G_2})\frac{1}{n}\boldsymbol{\Pi}_2'(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n)\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2})\hat{\mathbf{G}}_1^*(n)\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} . \tag{6.43}$$

We notice that

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

because Condition (II) and

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

Then by applying the central limit theorem (see *Theorem 1* of Anderson and Kunitomo (1992)) to $(1/\sqrt{n})\mathbf{\Pi}_{22}^{'}(n)\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta}$, we obtain the limiting normal distribution $N(\mathbf{0},\sigma^2\mathbf{\Phi}_{22.1})$. This proves (i) of *Theorem 3* for $0 \leq \eta < 1/2$.

(II) We consider the asymptotoic behavior of the quadratic term

$$
\begin{aligned}
(6.44) \qquad & \frac{1}{\sqrt{n}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_2(n)\mathbf{\Omega}\right] \\
= \quad & \frac{1}{\sqrt{n}}\left[\sum_{i,j=1}^{n}\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_j^{*}\left(\mathbf{v}_i\mathbf{v}_j^{'} - \delta_i^j\mathbf{\Omega}_i(n)\right)\right] + \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*}\left(\mathbf{\Omega}_i(n) - \mathbf{\Omega}\right)\right] ,
\end{aligned}
$$

where δ_i^j is the indicator function ($\delta_i^i = 1$ and $\delta_i^j = 0$ $(i \neq j)$). For any (constant vectors) $\mathbf{a}$ and $\mathbf{b}$, there exists M_3 $(M_3 > 0)$ such that

$$
\begin{aligned}
& \frac{1}{n}\mathbf{E}\left[\sum_{i,j=1}^{n}\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_j^{*} \times \mathbf{a}^{'}(\mathbf{v}_i\mathbf{v}_j^{'} - \delta_i^j\mathbf{\Omega}_i(n))\mathbf{b}\right]^2 \\
= \quad & \frac{1}{n}\mathbf{E}\Bigg[\sum_{i=1}^{n}[\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*}]^2[\mathbf{a}^{'}(\mathbf{v}_i\mathbf{v}_i^{'} - \mathbf{\Omega}_i(n))\mathbf{b}]^2 \\
& + \sum_{i\neq j}[\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_j^{*}]^2[\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j\mathbf{b}]^2 + \sum_{i\neq j}[\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_j^{*}]^2[\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j^{'}\mathbf{b}\mathbf{a}^{'}\mathbf{v}_j\mathbf{v}_i^{'}\mathbf{b}]\Bigg] \\
\leq \quad & M_3\frac{K_2(n)}{n} \longrightarrow 0
\end{aligned}
$$

because the conditional moments of v_{ij}^4 are bounded, $\sum_{i=1}^{n}\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*} = K_2(n)$ and $\sum_{i=1}^{n}(\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*})^2 \leq K_2(n)$. Since

$$
\|\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\mathbf{z}_i^{*'}\mathbf{A}_{22.1}^{-1}\mathbf{z}_i^{*}\left(\mathbf{\Omega}_i(n) - \mathbf{\Omega}\right)\right]\| \leq \left[\sqrt{n}\max_{1\leq i\leq n}\|\mathbf{\Omega}_i(n) - \mathbf{\Omega}\|\right]\frac{K_2(n)}{n} ,
$$

we find

$$
(6.45) \qquad \frac{1}{\sqrt{n}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_2(n)\mathbf{\Omega}\right] \stackrel{p}{\longrightarrow} \mathbf{O}
$$

when $0 \leq \eta < 1$.

Next, we shall investigate the asymptotic property of the TSLS estimator. If we substitute $\lambda(n)$ for 0 in (6.6), 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

$$
(6.46) \qquad \hat{\mathbf{G}}_1^{*}(n)\boldsymbol{\beta} - \left[c\mathbf{\Omega}\boldsymbol{\beta} + \frac{1}{\sqrt{n}}\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta}\right] \stackrel{p}{\longrightarrow} \mathbf{O} .
$$

We set $\hat{\boldsymbol{\beta}}_{TS}^{'} = (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'})$, which is the solution of (2.10). By evaluating each terms of

$$(\mathbf{0}, \mathbf{I}_{G_2})\sqrt{n}\left[\frac{1}{n}\boldsymbol{\Pi}_2^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n) + \frac{1}{\sqrt{n}}\hat{\mathbf{G}}_1^{*}(n)\right]\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta})\right] = \mathbf{0} \ ,$$

we have

$$\left[\frac{1}{n}\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n)\right]\sqrt{n}(\hat{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}) - (\mathbf{0}, \mathbf{I}_{G_2})\hat{\mathbf{G}}_1^{*}(n)\boldsymbol{\beta} = o_p(1) \ . \tag{6.47}$$

Then the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2)$ is the same as that of $\boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})\mathbf{G}_1^{*}(n)\boldsymbol{\beta}$. 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

$$\begin{aligned} & n^{1-\eta}\left[\frac{1}{n}\mathbf{G} - \frac{1}{n}\boldsymbol{\Pi}_2^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n)\right]\boldsymbol{\beta} \\ = \ & \frac{K_2(n)}{n^{\eta}}\boldsymbol{\Omega}\boldsymbol{\beta} + \frac{1}{n^{\eta}}\boldsymbol{\Pi}_2^{'}(n)\mathbf{Z}_{2.1}^{'}\mathbf{V}\boldsymbol{\beta} + \frac{1}{n^{\eta}}\left[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_2(n)\boldsymbol{\Omega}\right]\boldsymbol{\beta}. \end{aligned} \tag{6.48}$$

Because the last two terms of the right-hand side of (6.48) except the first term are of the order $o_p(n^{-\eta})$, we have

$$n^{1-\eta}\left[\frac{1}{n}\mathbf{G} - \frac{1}{n}\boldsymbol{\Pi}_2^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_2(n)\right]\boldsymbol{\beta} \xrightarrow{p} c\boldsymbol{\Omega}\boldsymbol{\beta} \tag{6.49}$$

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

$$(\mathbf{0}, \mathbf{I}_{G_2})\frac{1}{n}\boldsymbol{\Pi}_2^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n) \times n^{1-\eta}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2})c\boldsymbol{\Omega}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \tag{6.50}$$

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

(III) We consider the asymptotic property of the LIML estimator when $1/2 \leq \eta < 1$. By using the argument of (6.41) 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 (6.41), we have

$$\begin{aligned} & \boldsymbol{\beta}^{'}\left\{\sqrt{n}[\frac{K_2(n)}{n} - \lambda(n)]\boldsymbol{\Omega} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}\mathbf{Z}_{2.1}\boldsymbol{\Pi}_2(n) + \frac{1}{\sqrt{n}}[\mathbf{V}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{V} - K_2(n)\boldsymbol{\Omega}] \right. \\ & \left. -\lambda(n)\sqrt{\frac{n}{q(n)}}\hat{\mathbf{H}}_1\right\} \times \left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0} \ . \end{aligned}$$

Then by (6.37), (6.41) and $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} > 0$, we have the following result, which is summarized as a lemma.

Lemma 3 : For $0 \leq \eta < 1$,

$$\sqrt{n}\left[\lambda(n) - \frac{K_2(n)}{n}\right] \xrightarrow{p} 0 \tag{6.51}$$

as $n \longrightarrow \infty$.

Multiply (6.42) on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ to obtain

$$
\begin{aligned}
&(\mathbf{0}, \mathbf{I}_{G_2})\sqrt{n}\left\{\left[\frac{1}{n}\mathbf{\Pi}_2'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n) + \frac{K_2(n)}{n}\mathbf{\Omega}\right]\right.\\
&+\frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\mathbf{\Pi}_2'(n)\mathbf{Z}_{2.1}'\mathbf{V} + \frac{1}{\sqrt{n}}\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{\Pi}_2(n) + \frac{1}{\sqrt{n}}(\mathbf{V}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{V} - K_2(n)\mathbf{\Omega})\right]\\
&\left.-\left[\lambda^*(n)\right]\frac{1}{q(n)}\mathbf{H}\right\} \times \left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta})\right] = \mathbf{0} \ .
\end{aligned}
$$

We now use (6.37), (6.41) and the fact that

$$
\left[\frac{1}{n}\mathbf{\Pi}_2'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n) + \frac{K_2(n)}{n}\mathbf{\Omega} - \lambda(n)\frac{1}{q(n)}\mathbf{H}\right]\boldsymbol{\beta} = o_p(\frac{1}{\sqrt{n}}) \ .
$$

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

$$
(6.52) \qquad (\mathbf{0}, \mathbf{I}_{G_2})\left[\frac{1}{n}\mathbf{\Pi}_2'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_2(n)\sqrt{n}(\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}) + \frac{1}{\sqrt{n}}\mathbf{\Pi}_2'(n)\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta}\right] \xrightarrow{p} \mathbf{0} \ ,
$$

which is equivalent to

$$
(6.53) \qquad \left[\mathbf{\Pi}_{22}'(n)\mathbf{M}_{22.1}\mathbf{\Pi}_{22}(n)\right]\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI} - \boldsymbol{\beta}_2) - \frac{1}{\sqrt{n}}\mathbf{\Pi}_{22}'(n)\mathbf{Z}_{2.1}'\mathbf{V}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \ .
$$

By applying the CLT to the second term of (6.53) as (I), we complete the proof of (i) of *Theorem 2* for the LIML estimator of $\boldsymbol{\beta}$ when $1/2 \le \eta < 1$. **Q.E.D.**

Proof of Theorem 4 :

We use the similar arguments as the proofs of *Theorem 1* and *Theorem 3* and consider the representation

$$
\begin{aligned}
\frac{1}{n}\mathbf{G}^* &= \frac{1}{n}(\mathbf{\Pi}'(n)\mathbf{Z}' + \mathbf{V}')(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})(\mathbf{Z}\mathbf{\Pi}(n) + \mathbf{V})\\
&= \frac{1}{n}\mathbf{\Pi}'(n)\mathbf{Z}'(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{Z}\mathbf{\Pi}(n) + \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\mathbf{V}'(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})]\mathbf{Z}\mathbf{\Pi}(n)\right]\\
&\qquad +\frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\mathbf{\Pi}'(n)\mathbf{Z}'(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{V}\right] + \frac{1}{n}\mathbf{V}'(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{V} \ .
\end{aligned}
$$

As $n \longrightarrow \infty$, three terms of $(1/n)\mathbf{G}^*$ except the first term converge to zero matrices because of our conditions in our assumptions,

$$
\mathbf{E}[\frac{1}{n}\mathbf{V}'(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{V}|\mathbf{Z}] - \frac{K_{21}(n)}{n}\mathbf{\Omega} \xrightarrow{p} \mathbf{O} \ ,
$$

and

$$\mathbf{E}\{[\frac{1}{\sqrt{n}}\mathbf{\Pi}_{22}^{'}(n)\mathbf{Z}_2^{'}(n)(\mathbf{P}_{Z^*}-\mathbf{P}_{Z_1})\mathbf{V}][\frac{1}{\sqrt{n}}\mathbf{V}^{'}(\mathbf{P}_{Z^*}-\mathbf{P}_{Z_1})\mathbf{Z}_2(n)\mathbf{\Pi}_{22}(n)]|\mathbf{Z}\}$$
$$-\mathbf{tr}(\mathbf{\Omega})\frac{1}{n}\mathbf{\Pi}_{22}^{'}(n)\mathbf{Z}_2^{'}(n)(\mathbf{P}_{Z^*}-\mathbf{P}_{Z_1})\mathbf{Z}_2(n)\mathbf{\Pi}_{22}(n)\stackrel{p}{\longrightarrow}\mathbf{O}$$

and

$$\begin{aligned}\frac{1}{n}\mathbf{H}^* &= \frac{1}{n}\mathbf{Y}^{'}\mathbf{Y}-\frac{1}{n}\mathbf{G}^*\\ &= \frac{1}{n}(\mathbf{\Pi}^{'}(n)\mathbf{Z}^{'}+\mathbf{V}^{'})(\mathbf{Z}\mathbf{\Pi}(n)+\mathbf{V})-\frac{1}{n}\mathbf{G}^*\stackrel{p}{\longrightarrow}\mathbf{\Omega}\ .\end{aligned}$$

Then as $n\to\infty$, we have the convergence in probability as

$$\frac{1}{n}\mathbf{G}^*\stackrel{p}{\longrightarrow}\mathbf{G}_0^{\dagger}=\begin{bmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{bmatrix}\mathbf{\Phi}_{22.1}\left[\boldsymbol{\beta}_2,\mathbf{I}_{G_2}\right]\ . \tag{6.54}$$

For the SLIML estimation we set the smallest characteristic root (we may write $\lambda^*(n)$) and its associated vector as $|(1/n)\mathbf{G}^*-\lambda^*(n)(1/n)\mathbf{H}^*|=0$ and

$$[\frac{1}{n}\mathbf{G}^*-\lambda^*(n)\frac{1}{n}\mathbf{H}^*]\hat{\boldsymbol{\beta}}_{SLI}=\mathbf{0}\ . \tag{6.55}$$

The probability limit of the SLIML estimator $\hat{\boldsymbol{\beta}}_{SLI}=(1,-\hat{\boldsymbol{\beta}}_{2.LI}^{'})^{'}$ is $\boldsymbol{\beta}=(1,-\boldsymbol{\beta}_2^{'})^{'}$ as $n\to\infty$ and $\lambda^*(n)\stackrel{p}{\to}0$. Let $\hat{\mathbf{G}}_1=\sqrt{n}[(1/n)\mathbf{G}^*-\mathbf{G}_0^{\dagger}]$, $\lambda_1=\sqrt{n}\lambda^*(n)$, $\hat{\mathbf{b}}_1=\sqrt{n}[\hat{\boldsymbol{\beta}}_{SLI}-\boldsymbol{\beta}]$, $\hat{\mathbf{H}}_1=\sqrt{n}[(1/n)\mathbf{H}^*-\mathbf{\Omega}]$, and then we can write
$\hat{\mathbf{b}}_1=(-1)(\mathbf{0},\mathbf{I}_{G_2})^{'}\sqrt{n}[\hat{\boldsymbol{\beta}}_{SLI}-\boldsymbol{\beta}]$. By substituing the random variables $\hat{\mathbf{G}}_1$, $\hat{\mathbf{H}}_1$, and λ_1 into (6.6), the resulting relation becomes

$$\mathbf{G}_0^{\dagger}\boldsymbol{\beta}+\frac{1}{\sqrt{n}}[\hat{\mathbf{G}}_1-\lambda_1\mathbf{\Omega}]\boldsymbol{\beta}+\frac{1}{\sqrt{n}}\mathbf{G}_0^{\dagger}\hat{\mathbf{b}}_1=o_p(\frac{1}{\sqrt{n}})\ .$$

By ignoring the higher order terms, we need to consider the modified estimator $\mathbf{e}_{SLI}^*(\boldsymbol{\beta})$ which satisfies

$$\mathbf{G}_0^{\dagger}\begin{pmatrix}\mathbf{0}^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\mathbf{e}_{LI}^*(\beta)=[\hat{\mathbf{G}}_1-\lambda_1\mathbf{\Omega}]\boldsymbol{\beta}\ . \tag{6.56}$$

By defining the normalized (SLIML) random vector $\hat{\mathbf{e}}_{SLI}(\boldsymbol{\beta})=\sqrt{n}[\hat{\boldsymbol{\beta}}_{2.SLI}-\boldsymbol{\beta}_2]$, we can show that $\mathbf{e}_{SLI}^*(\beta)=\hat{\mathbf{e}}_{SLI}(\beta)+o_p(1)$. Then by multiplying $(\mathbf{0},\mathbf{I}_{G_2})$ and $\boldsymbol{\beta}^{'}$ from the left-hand-side of (6.56), we have the relation

$$(\mathbf{0},\mathbf{I}_{G_2})\mathbf{G}_0^{\dagger}\begin{pmatrix}\mathbf{0}^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\mathbf{e}_{SLI}^*(\boldsymbol{\beta})=(\mathbf{0},\mathbf{I}_{G_2})(\hat{\mathbf{G}}_1-\lambda_1\mathbf{\Omega})\boldsymbol{\beta}\ , \tag{6.57}$$

and

$$\boldsymbol{\beta}^{'}\mathbf{G}_0^{\dagger}\begin{pmatrix}\mathbf{0}^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\mathbf{e}_{SLI}^{*}(\boldsymbol{\beta}) = \boldsymbol{\beta}^{'}(\hat{\mathbf{G}}_1 - \lambda_1\boldsymbol{\Omega})\boldsymbol{\beta}\ . \tag{6.58}$$

Since $\mathbf{G}_0^{\dagger}\boldsymbol{\beta} = \mathbf{0}$ and $(\mathbf{0}, \mathbf{I}_{G_2})\mathbf{G}_0^{\dagger}(\mathbf{0}, \mathbf{I}_{G_2})^{'} = \boldsymbol{\Phi}_{22.1}$, we find $\lambda_1 = \boldsymbol{\beta}^{'}\hat{\mathbf{G}}_1\boldsymbol{\beta}/\sigma^2$ and

$$\mathbf{e}_{SLI}^{*}(\boldsymbol{\beta}) = \Phi_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}]\hat{\mathbf{G}}_1\boldsymbol{\beta}\ .$$

Since $\mathbf{V}\boldsymbol{\beta} = \mathbf{u}$, we have

$$\hat{\mathbf{G}}_1\boldsymbol{\beta} = \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_2^{'}(n)\mathbf{Z}^{'}(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{u} + \frac{1}{\sqrt{n}}\mathbf{V}^{'}(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{V}\boldsymbol{\beta}\ . \tag{6.59}$$

By using Conditon (I), we have $(1/\sqrt{n})\mathbf{E}[\mathbf{V}^{'}(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{V}|\mathbf{Z}] - (1/\sqrt{n})[K_{21}(n) - K_1]\boldsymbol{\Omega} \xrightarrow{p} \mathbf{O}$ and $\lambda_1 \xrightarrow{p} 0$ as $n \to \infty$. Then we have the asymptotic distribution of (6.59) by applying the central limit theorem with the Lindeberg conditions. (See *Theorem 1* of Anderson and Kunitomo (1992), for instance.) In order to obtain the asymptotic covariance matrix, we need to find the probability limit of the conditional expectation given $\mathbf{Z}$ as $\mathbf{E}\left[\boldsymbol{\Pi}_2^{'}(n)\mathbf{Z}^{'}(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\mathbf{u}\mathbf{u}^{'}(\mathbf{P}_{Z^*} - \mathbf{P}_{Z_1})\boldsymbol{\Pi}_2(n)|\mathbf{Z}\right]$. Then by using Conditions of *Theorem 4*, we have the asymptotic normality and its asymptotic covariance matrix is given by the lower-right block matrix of

$$\sigma^2\begin{pmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{pmatrix}\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\ .$$

Finally, by setting $\lambda^*(n) = 0$, we have the asymptotic distribution of the STSLS estimator, which is the same as the LIML estimator. It is because $\sqrt{n}\lambda^*(n) \xrightarrow{p} 0$ in the present situation. **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. 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.

[5] 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.

[6] Anderson, T.W. and T. Sawa (1979), "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," *Econometrica*, Vol. 47, 163-182.

[7] 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.

[8] 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.

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

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

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

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

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

[14] Chamberlin, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restriction," *Journal of Econometrics*, Vol. 34-3, 305-334.

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

[16] Fujikoshi, Y., K. Morimune, N. Kunitomo, and M. Taniguchi (1982), "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System," *Journal of Econometrics*, Vol. 18, 2, 191-205.

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

[18] Hansen, C. J. Hausman, and Newey, W. K. (2004), "Many Instruments, Weak Instruments, and Microeconometric Practice," Unpublished Manuscript.

[19] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, 1029-1054.

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

[21] Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Equation," *Annals of Mathematical Statistics*, Vol. 31, 1208-1211.

[22] Guggenberger, P. (2004), "Finite-Sample Evidence Suggesting a heavy Tail Problem of the Generalized Empirical Likelihhood Estimator," Unpublished Manuscript, Department of Economics, UCLA.

[23] Kitamura, Y. (1997), "Empirical Likelihood Methods with Weakly Dependent Processes," *The Annals of Statistics*, Vol.19-2, 1053-61.

[24] 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.

[25] 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 Mathmeatical Studies in the Social Sciences, Stanford University.

[26] Kunitomo, N. (1987), "A Third Order Optimum Properties 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.

[27] Kunitomo,N. and Matsushita, Y. (2003a), "Finite Sample Distributions of the Empirical Likelihood Estimator and GMM Estimator," Discussion Paper CIRJE-F-200, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2003/2003cf200.pdf).

[28] Kunitomo,N. and Matsushita, Y. (2003b) , "Asymptotic Expansions of the Distributions of Semi-Parametric Estimators in a Linear Simultaneous Equations System," Discussion Paper CIRJE-F-237, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2003/2003cf237.pdf).

[29] 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.

[30] Mariano, R.S. and Sawa, T. (1972), "The Exact Finite Sample Distribution of the Limited Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, Vol. 67, 159-163.

[31] Mittelhammer, R., G. Judge, and R. Schoenberg (2004), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods," Unpublished Manuscript,

[32] Newey, W. K. (2004), "Many Instrument Asymptotics," Unpublished Manuscript, Department of Economics, MIT.

[33] Newey, W. K. and R. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimator," *Econometrica*, Vol. 72, 219-255.

[34] Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, Vol. 22, 300-325.

[35] Owen, A. B. (2001), *Empirical Likelihood*, Chapman and Hall.

[36] Phillips, P.C.B. (1980), "The Exact Finite Sample Density of Instrumental Variables Estimators in an Equation with n+1 Endogenous Variables," *Econometrica*, Vol. 48, 861-878.

[37] Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, 300-325.

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

[39] Stock, J. and Yogo, M. (2003), "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," Unpublished Manuscript.

APPENDIX : TABLES AND FIGURES

Notes on Tables

In Tables the distributions are tabulated in the standardized terms, that is, of (4.1). The tables include three quartiles, the 5 and 95 percentiles and the interquartile range of the distribution for each case. Since the limiting distributions of (4.1) for the MEL and GMM estimators in the standard large sample asymptotic theory are $N(0,1)$ as $n \to \infty$, we add the standard normal case as the bench mark.

Notes on Figures

In Figures the cdf's of the LIML, MEL and GMM estimators are shown in the standardized terms, that is, of (4.1). (The cdf of the TSLS estimator is quite similar to that of the GMM estimator in all cases and it was omitted in many cases.) The dotted lines were used for the distributions of the GMM estimator. For the comparative purpose we give the standard normal distribution as the bench mark for each case.

Table 1: $n - K = 30, K_2 = 3, \alpha = 1$

		$\delta^2 = 30$				$\delta^2 = 100$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.40	-1.52	-1.55	-1.64	-1.47	-1.54	-1.59	-1.63
L.QT	-0.67	-0.64	-0.66	-0.83	-0.85	-0.65	-0.67	-0.77	-0.79
MEDN	0	0.00	-0.01	-0.24	-0.26	0.00	0.01	-0.14	-0.14
U.QT	0.67	0.76	0.80	0.44	0.47	0.71	0.75	0.55	0.57
X95	1.65	2.14	2.37	1.64	1.66	1.90	1.98	1.71	1.74
IQR	1.35	1.40	1.46	1.27	1.31	1.36	1.42	1.32	1.36

Table 2: $n - K = 100, K_2 = 10, \alpha = 1$

		$\delta^2 = 50$				$\delta^2 = 100$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.49	-1.68	-1.98	-2.09	-1.54	-1.61	-1.97	-2.04
L.QT	-0.67	-0.66	-0.74	-1.31	-1.33	-0.66	-0.72	-1.17	-1.22
MEDN	0	0.00	0.01	-0.77	-0.77	0.00	-0.01	-0.59	-0.61
U.QT	0.67	0.76	0.83	-0.18	-0.15	0.73	0.81	0.05	0.08
X95	1.65	2.11	2.35	0.76	0.89	1.90	2.11	1.06	1.18
IQR	1.35	1.42	1.57	1.12	1.19	1.39	1.53	1.22	1.30

Table 3: $n - K = 300, K_2 = 30, \alpha = 1$

		$\delta^2 = 50$				$\delta^2 = 100$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.63	-1.82	-2.88	-2.95	-1.56	-1.77	-2.76	-2.87
L.QT	-0.67	-0.75	-0.79	-2.28	-2.30	-0.69	-0.75	-2.10	-2.14
MEDN	0	0.00	0.02	-1.85	-1.85	0.00	0.02	-1.60	-1.59
U.QT	0.67	0.85	0.97	-1.40	-1.37	0.77	0.86	-1.07	-1.02
X95	1.65	2.48	2.94	-0.67	-0.60	2.08	2.38	-0.21	-0.12
IQR	1.35	1.60	1.76	0.88	0.94	1.46	1.61	1.03	1.11

Table 4: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50$

		$u_i = (\chi^2(3) - 3)/\sqrt{6}$				$u_i = t(5)$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.52	-1.53	-2.06	-1.96	-1.51	-1.55	-2.02	-1.97
L.QT	-0.67	-0.67	-0.67	-1.32	-1.24	-0.62	-0.67	-1.28	-1.22
MEDN	0	-0.01	-0.01	-0.77	-0.69	0.02	0.01	-0.75	-0.69
U.QT	0.67	0.75	0.76	-0.17	-0.09	0.77	0.83	-0.18	-0.12
X95	1.65	2.17	2.24	0.78	0.82	2.12	2.33	0.78	0.86
IQR	1.35	1.42	1.43	1.14	1.15	1.39	1.50	1.10	1.10

Table 5: $\alpha = 1, \delta^2 = 100, u_i = \|Z_i\|\epsilon_i$

		$n-K=30, K_2=3$				$n-K=100, K_2=10$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.39	-1.51	-1.52	-1.57	-1.52	-1.64	-1.96	-2.03
L.QT	-0.67	-0.60	-0.66	-0.73	-0.78	-0.67	-0.70	-1.20	-1.22
MEDN	0	0.02	-0.02	-0.14	-0.17	-0.04	0.03	-0.65	-0.60
U.QT	0.67	0.70	0.71	0.52	0.51	0.70	0.83	-0.03	0.07
X95	1.65	1.93	2.05	1.62	1.70	1.97	2.20	1.03	1.09
IQR	1.35	1.29	1.36	1.25	1.29	1.37	1.53	1.18	1.29

Table 6: $n-K = 300, K_2 = 30, \alpha = 1, u_i = \|Z_i\|\epsilon_i$

		$\delta^2 = 50$				$\delta^2 = 100$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.62	-1.77	-2.90	-2.97	-1.56	-1.70	-2.76	-2.83
L.QT	-0.67	-0.72	-0.77	-2.30	-2.31	-0.70	-0.74	-2.14	-2.14
MEDN	0	0.02	0.03	-1.87	-1.86	0.00	0.01	-1.63	-1.60
U.QT	0.67	0.89	0.97	-1.43	-1.39	0.79	0.88	-1.10	-1.05
X95	1.65	2.55	2.97	-0.76	-0.68	2.13	2.34	-0.25	-0.14
IQR	1.35	1.61	1.73	0.87	0.92	1.49	1.61	1.04	1.09

Table 7: $n-K = 1000, K_2 = 100, \alpha = 1, \delta^2 = 100$

		$u_i = N(0,1)$			$u_i = \|Z_i\|\epsilon_i$		
	normal	LIML	TSLS	GMM	LIML	TSLS	GMM
X05	-1.65	-1.82	-4.46	-4.51	-1.84	-4.44	-4.49
L.QT	-0.67	-0.78	-3.89	-3.92	-0.81	-3.91	-3.93
MEDN	0	0.00	-3.53	-3.53	0.01	-3.54	-3.53
U.QT	0.67	0.89	-3.14	-3.12	0.93	-3.17	-3.12
X95	1.65	2.39	-2.57	-2.49	2.51	-2.59	-2.51
IQR	1.35	1.67	0.75	0.80	1.74	0.75	0.81

Table 8: $n-K = 300, K_2 = 30, \delta^2 = 100$

		$\alpha = 0$				$\alpha = 5$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.90	-2.16	-1.44	-1.53	-1.43	-1.52	-3.14	-3.24
L.QT	-0.67	-0.78	-0.90	-0.60	-0.66	-0.64	-0.69	-2.63	-2.65
MEDN	0	0.00	-0.02	0.00	-0.01	0.00	-0.02	-2.22	-2.22
U.QT	0.67	0.78	0.86	0.60	0.64	0.73	0.76	-1.77	-1.73
X95	1.65	1.93	2.14	1.46	1.56	1.98	2.14	-1.02	-0.96
IQR	1.35	1.56	1.76	1.19	1.30	1.37	1.45	0.86	0.92

Table 9: $\alpha = 1$

		$n-K=100, K_2=3, \delta^2=5$				$n-K=100, K_2=10, \delta^2=10$			
	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
X05	-1.65	-1.78	-1.84	-1.68	-1.66	-1.72	-2.16	-2.09	-2.04
L.QT	-0.67	-0.70	-0.73	-0.97	-0.95	-0.77	-0.90	-1.59	-1.47
MEDN	0	-0.08	-0.10	-0.52	-0.51	-0.06	-0.14	-1.08	-1.09
U.QT	0.67	0.81	0.80	0.02	0.02	1.00	0.94	-0.64	-0.68
X95	1.65	4.37	4.71	1.22	1.16	4.45	4.40	0.11	0.02
IQR	1.35	1.51	1.53	0.99	0.97	1.77	1.84	0.85	0.79

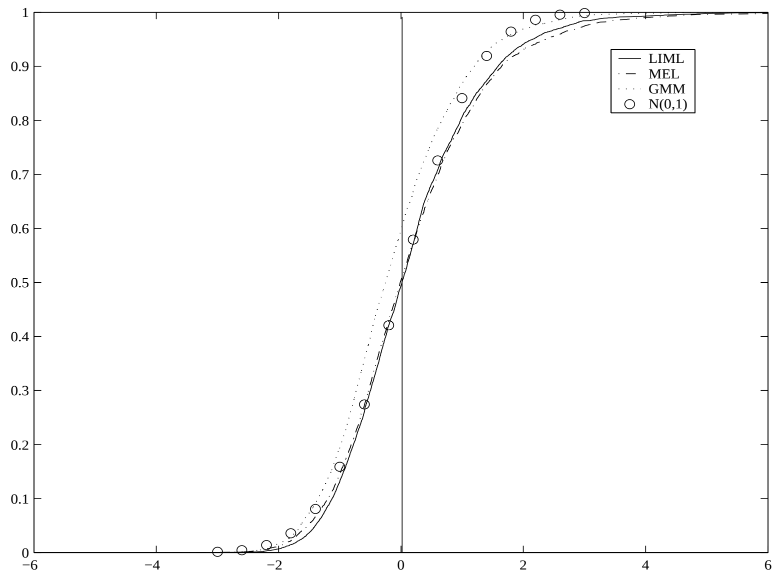

Figure 1: $n-K=30, K_2=3, \alpha=1, \delta^2=30$

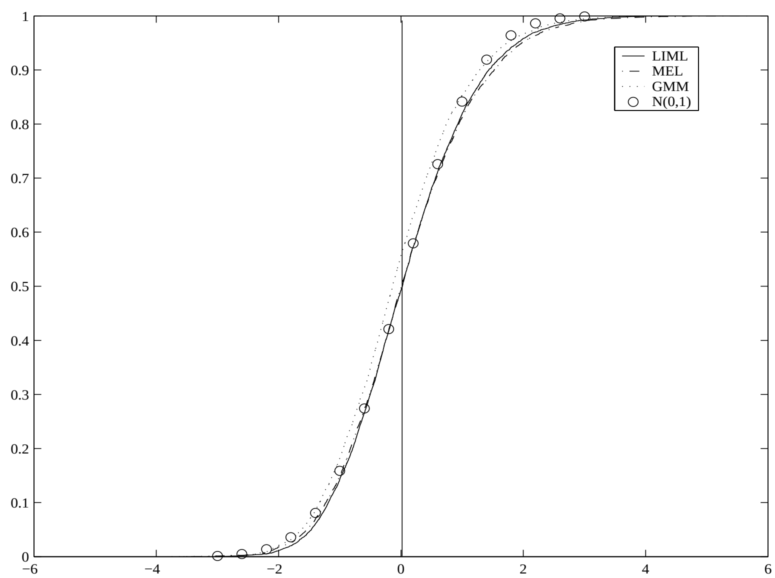

Figure 2: $n-K=30, K_2=3, \alpha=1, \delta^2=100$

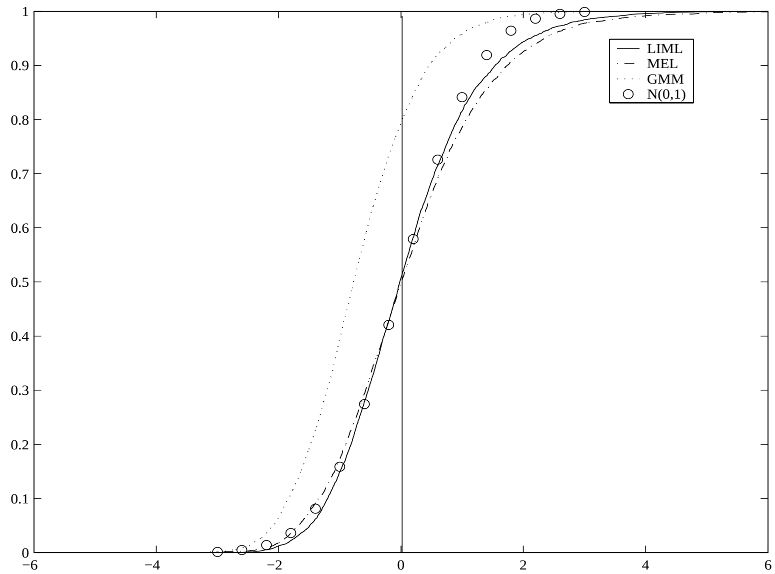

Figure 3: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50$

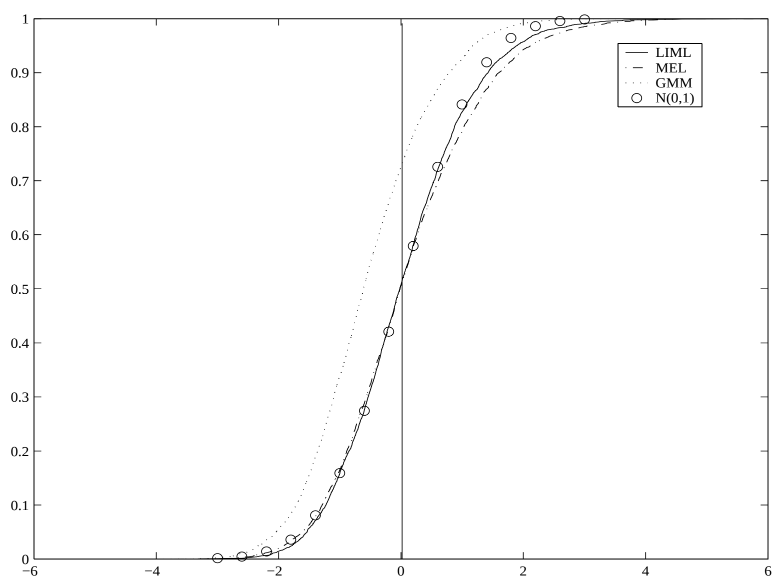

Figure 4: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 100$

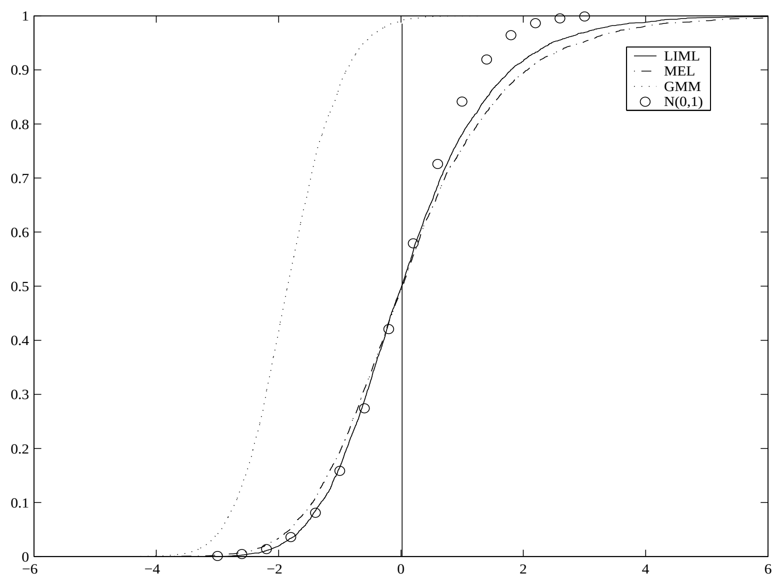

Figure 5: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 50$

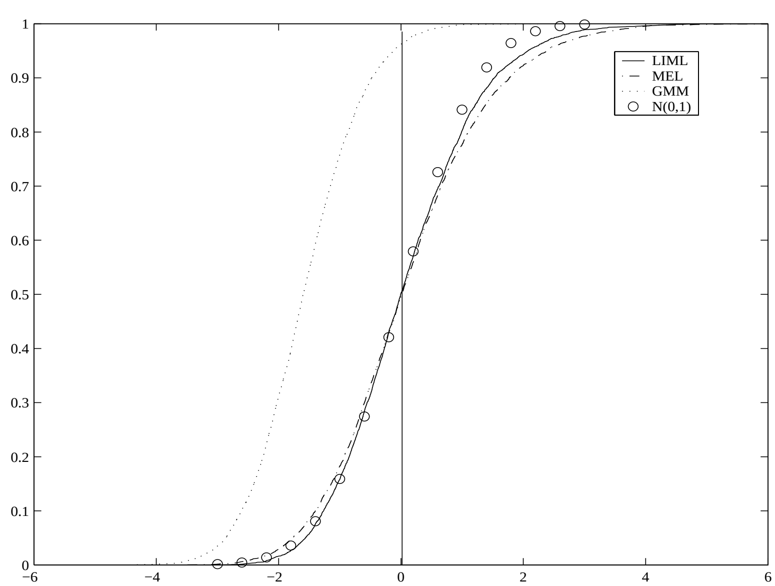

Figure 6: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 100$

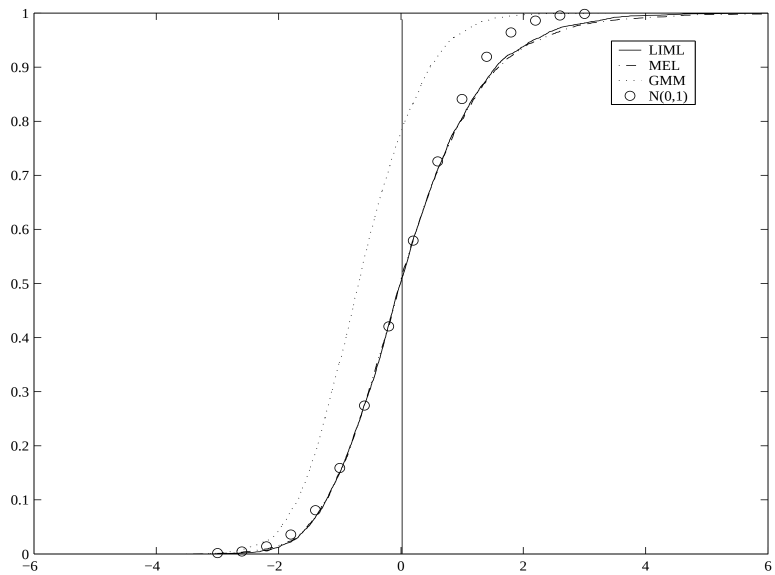

Figure 7: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = \frac{\chi^2(3)-3}{\sqrt{6}}$

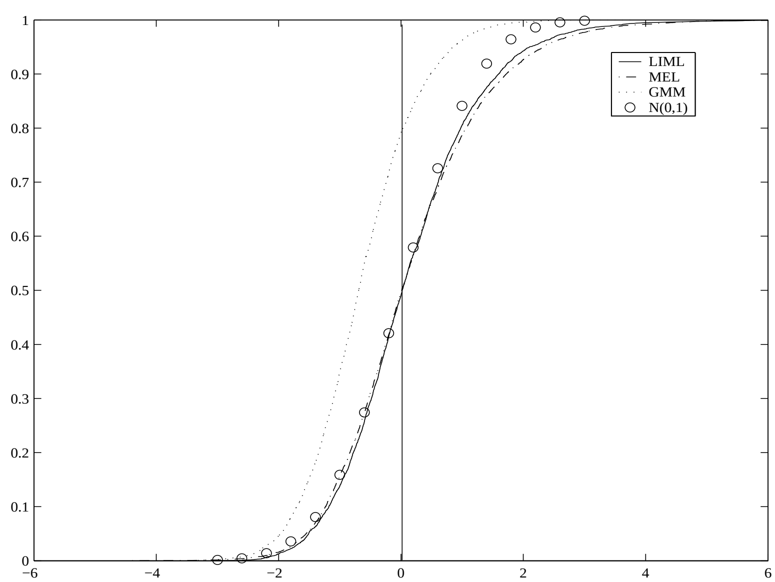

Figure 8: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = t(5)$

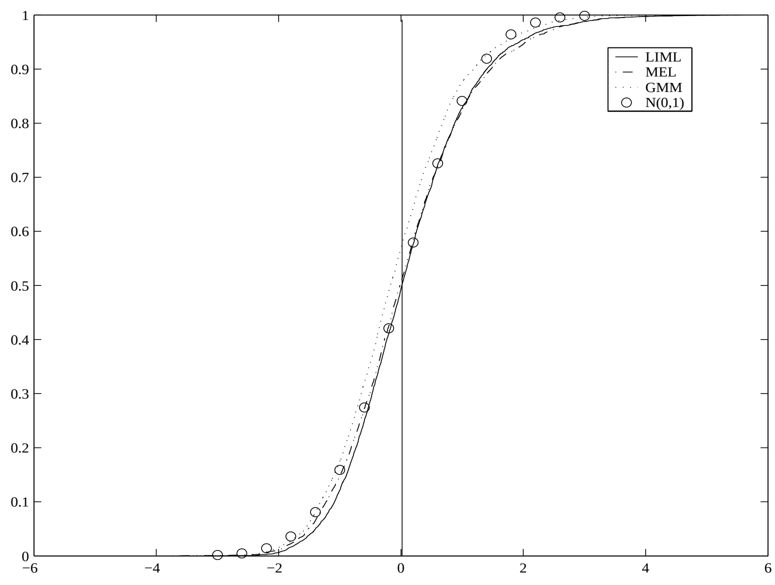

Figure 9: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

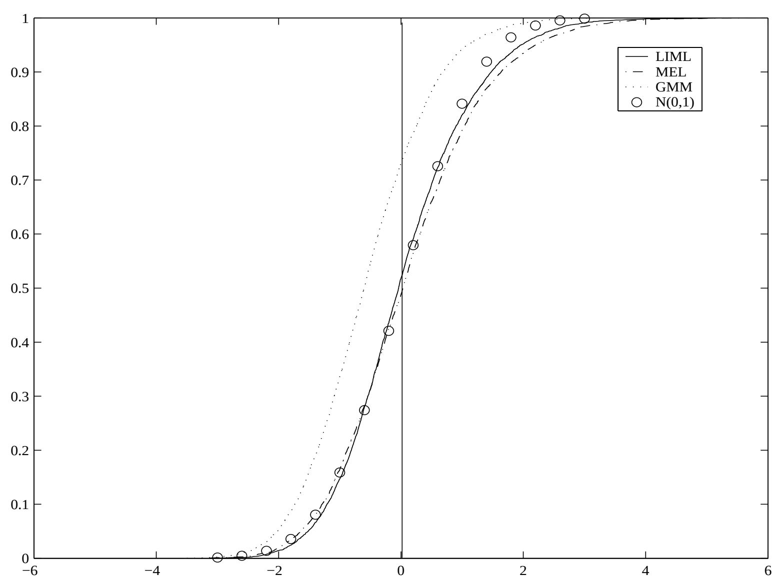

Figure 10: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

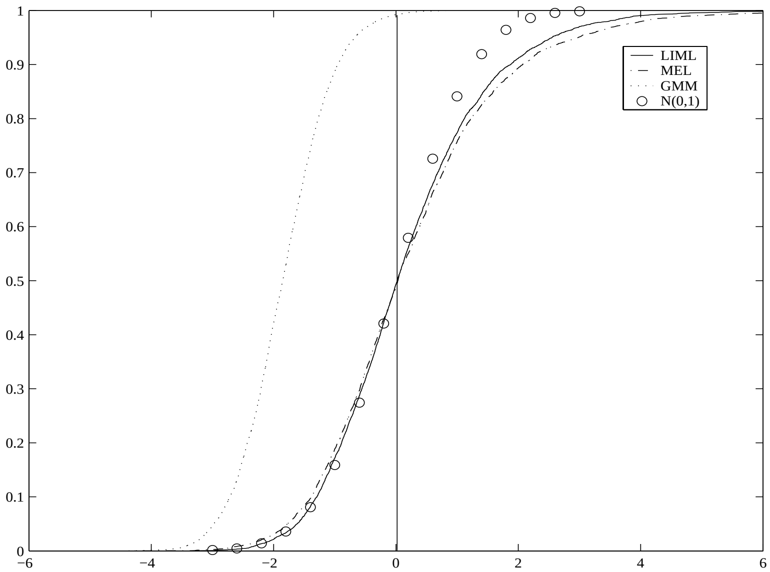

Figure 11: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 50, u_i = \|z_i\|\epsilon_i$

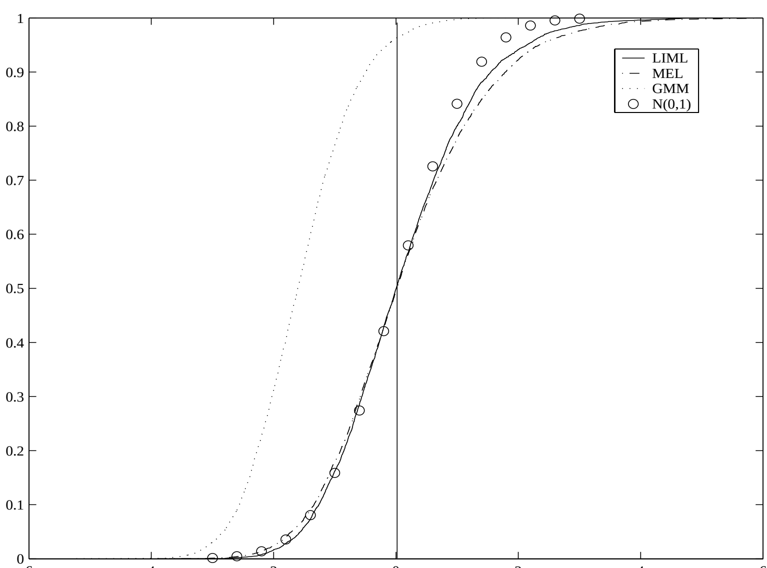

Figure 12: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

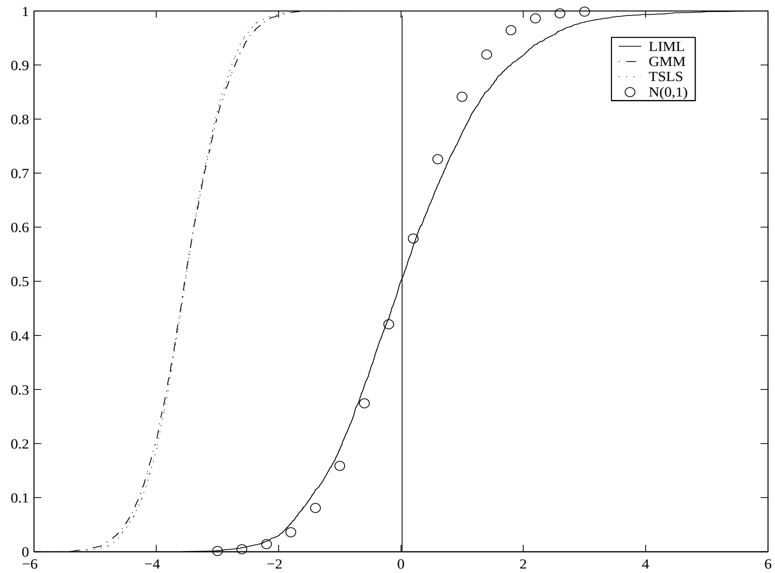

Figure 13: $n-K=1000, K_2=100, \alpha=1, \delta^2=100$

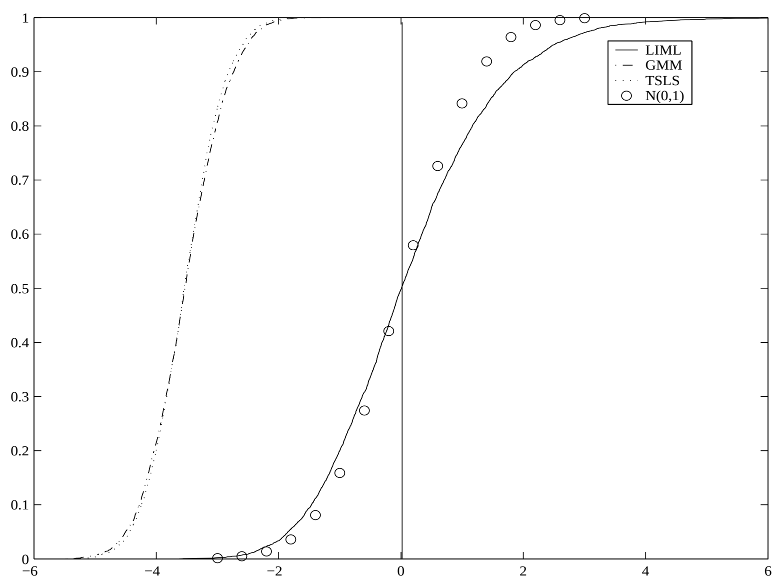

Figure 14: $n-K=1000, K_2=100, \alpha=1, \delta^2=100, u_i=\|z_i\|\epsilon_i$

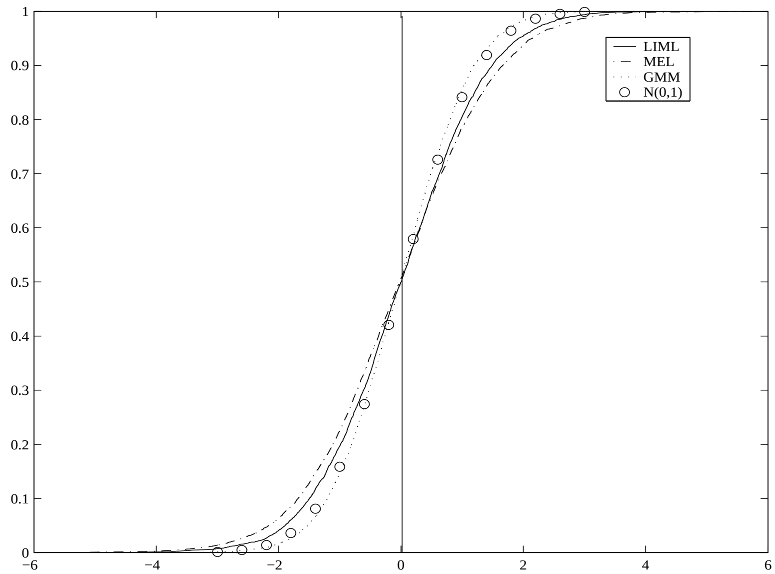

Figure 15: $n - K = 300, K_2 = 30, \alpha = 0, \delta^2 = 100$

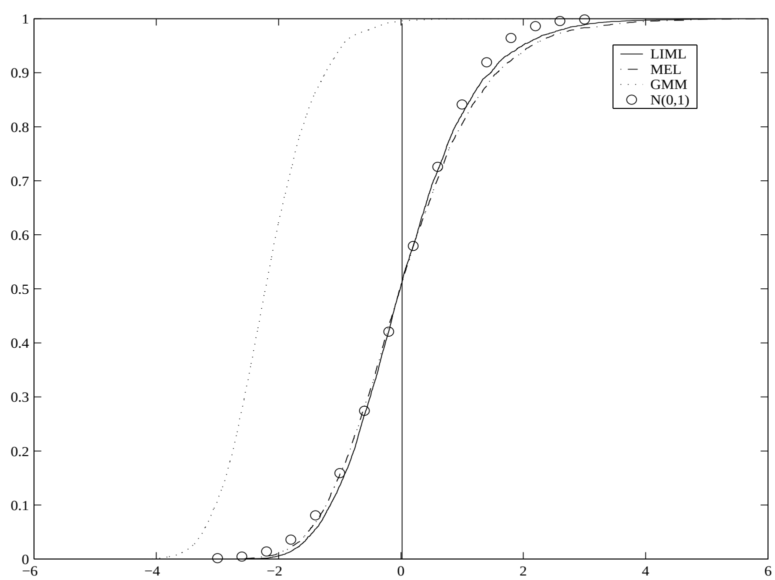

Figure 16: $n - K = 300, K_2 = 30, \alpha = 5, \delta^2 = 100$

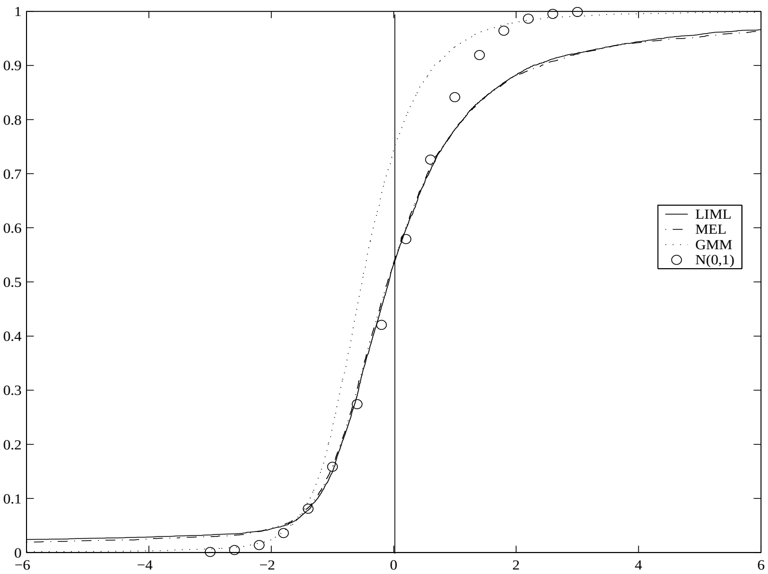

Figure 17: $n - K = 100, K_2 = 3, \alpha = 1, \delta^2 = 5$

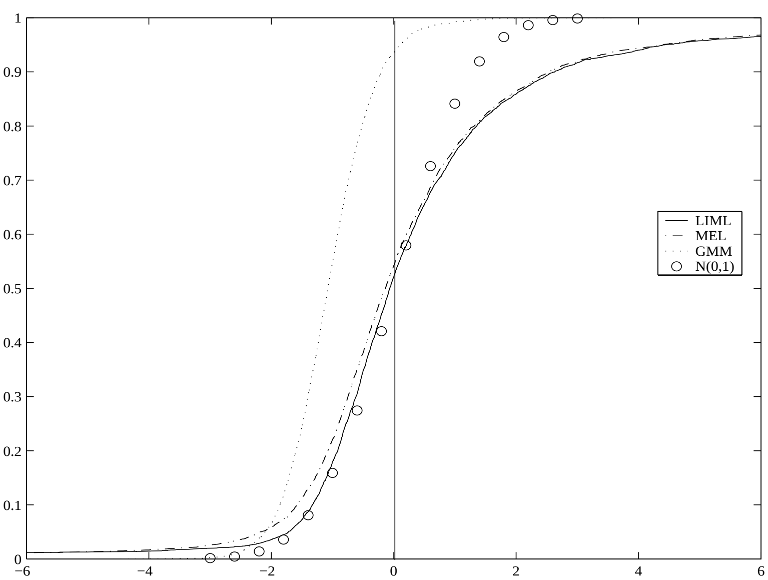

Figure 18: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 10$