CIRJE-F-321

A New Light from Old Wisdoms:
Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models

T. W. Anderson
Stanford University

Naoto Kunitomo
Yukitoshi Matsushita
University of Tokyo

February 2005

A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models

T. W. Anderson*
Naoto Kunitomo†
and
Yukitoshi Matsushita

February 26, 2005

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 a generalized method of moments (GMM) (or the estimating equation) estimator. Tables and figures 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 or many weak instruments in the terminology of recent econometric literatures. We give a new result on the asymptotic optimality of the LIML estimator when the number of instruments is large.

Key Words

Finite Sample Properties, Maximum Empirical Likelihood, Generalized Method of Moments, Microeconometric Models with Many Instruments, Limited Information Maximum Likelihood, Asymptotic Optimality

JEL Code: C13, C30

*Department of Statistics and Department of Economics, Stanford University
†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 by the help of many computer packages. 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 microeconometric 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 estimator and the TSLS estimator 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 estimator and the MEL estimator 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.

The main 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. 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. Advice can be given as to when one is preferred to the other. In this paper we 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. It is certainly possible to generalize the single linear structural equation with instrumental variables into several different directions.

Another 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 are the same, but the higher-order terms are different. For instance, Fujikoshi et. al. (1982) and their citations for the the LIML estimator and the TSLS estimator, and Kunitomo and Matsushita (2003b) for the MEL estimator and the GMM estimator 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 and the MEL estimator do not possess any moments of positive integer order under a set of reasonable assumptions. Although the analyses of bias and the mean squared errors of the MEL estimator based on Monte Carlo experiments have been reported in some studies, we suspect that many of them are not reliable. Therefore instead of moments we need to investigate the exact cumulative distributions of the LIML and MEL estimators directly in a systematic way. The problem of nonexistence of moments had been already discussed in the econometric literatures under a set of reasonable assumptions. For instance, see Mariano and Sawa (1972), Phillips (1980), and Kunitomo and Matsushita (2003a).

It may be important to notice that there had been alternative asymptotic theories when the number of instrumental variables is large in estimating structural equations. Recently Stock and Yogo (2003), and Hansen et. al. (2004) have mentioned some possibilities in the context of microeconometric applications and practices. 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 microeconometric models because there are many instrumental variables sometimes used in microeconometric applications. For this purpose we shall give a new result 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 result on the asymptotic optimality gives 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.

In Section 2 we state the formulation of models and alternative estimation methods of unknown parameters in the simultaneous equations. Then we shall explain our tables and figures of the finite sample distributions of the estimators in Section 3 and discuss the finite sample properties of alternative estimators in Section 4. Moreover, in Section 5 we present a new results on the asymptotic optimality of the LIML estimator when the number of instruments is large in the simultaneous equations models and discuss the theoretical explanations of the finite sample properties of alternative estimation methods based on the large K_2 asymptotic theory. Then some conclusions will be given in Section 6. The proof of our theorems shall be given in Appendix A, and our Tables and Figures are gathered in Appendix B.

2. Alternative Estimation Methods of a Structural Equation with Instrumental Variables

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

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

where y_{1i} and $\boldsymbol{y}_{2i}$ are 1×1 and $G_1 \times 1$ (vector of) endogenous variables, $\boldsymbol{z}_{1i}$ is a $K_1 \times 1$ vector of exogenous variables, $\boldsymbol{\theta}^{'} = (\boldsymbol{\beta}^{'}, \boldsymbol{\gamma}^{'})$ is a $1 \times p$ $(p = G_1 + K_1)$ vector of 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 $(G_1 + 1)$ structural equations in which the vector of $1 + G_1$ endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$ and the vector of K $(= K_1 + K_2)$ exogenous variables $\{\boldsymbol{z}_i\}$ including $\{\boldsymbol{z}_{1i}\}$ are related linearly with the condition $n > K$. The set of exogenous variables $\{\boldsymbol{z}_i\}$ are often called the instrumental variables and we can write the orthogonal condition

$$\mathbf{E}[u_i \; \boldsymbol{z}_i] = \mathbf{0} \;\; (i = 1, \cdots, n) \; . \tag{2.2}$$

Because we do not specify the equations except (2.1) and we only have the limited information on the set of exogenous variables or instruments, we only consider the limited information estimation methods. Furthermore, when all structural equations in the econometric model are linear, the reduced form of $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})$ is

$$\boldsymbol{y}_i = \boldsymbol{\Pi}^{'} \mathbf{z}_i + \mathbf{v}_i \;\; (i = 1, \cdots, n) \; , \tag{2.3}$$

where $\mathbf{v}_i^{'} = (v_{1i}, \mathbf{v}_{2i}^{'})$ is a $1 \times (1 + G_1)$ disturbance vector with $\mathbf{E}[\mathbf{v}_i] = \mathbf{0}$ and $\mathbf{E}[\mathbf{v}_i \mathbf{v}_i^{'}] < \infty$. Let

$$\boldsymbol{\Pi} = (\boldsymbol{\pi}_1 \; , \boldsymbol{\Pi}_2) = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \tag{2.4}$$

be a $(K_1 + K_2) \times (1 + G_1)$ $(K = K_1 + K_2)$ partitioned matrix of the reduced form coefficients. By multiplying (2.3) on the left by $(1, -\boldsymbol{\beta}^{'})$, we have the relation $u_i = v_{1i} - \boldsymbol{\beta}^{'} \mathbf{v}_{2i}$ $(i = 1, \cdots, n)$ and the restriction

$$(1, -\boldsymbol{\beta}^{'}) \boldsymbol{\Pi}^{'} = (\boldsymbol{\gamma}^{'}, \mathbf{0}^{'}) \;\; . \tag{2.5}$$

The maximum empirical likelihood (MEL) estimator for the vector of parameters θ in (2.1) is defined by maximizing the Lagrangian form

$$L_n^{*}(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log(n p_i) - \mu(\sum_{i=1}^{n} p_i - 1) - n \boldsymbol{\lambda}^{'} \sum_{i=1}^{n} p_i \; \mathbf{z}_i [y_{1i} - \boldsymbol{\beta}^{'} \mathbf{y}_{2i} - \boldsymbol{\gamma}^{'} \mathbf{z}_{1i}] \; , \tag{2.6}$$

where μ and $\boldsymbol{\lambda}$ are a scalar and a $K \times 1$ vector of Lagrangian multipliers, and p_i $(i = 1, \cdots, n)$ are the weighted probability functions to be chosen. It has been known (see Qin and Lawles (1994) or Owen (1990, 2001)) that the above maximization problem is the same as to maximize

$$L_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) = - \sum_{i=1}^{n} \log\{1 + \boldsymbol{\lambda}^{'} \mathbf{z}_i \; [y_{1i} - \boldsymbol{\beta}^{'} \mathbf{y}_{2i} - \boldsymbol{\gamma}^{'} \mathbf{z}_{1i}]\} \; , \tag{2.7}$$

where we have the conditions $\hat{\mu} = n$ and $[n \hat{p}_i]^{-1} = 1 + \boldsymbol{\lambda}^{'} \mathbf{z}_i [y_{1i} - \hat{\boldsymbol{\beta}}^{'} \mathbf{y}_{2i} - \hat{\boldsymbol{\gamma}}^{'} \mathbf{z}_{1i}]$. By differentiating (2.7) with respect to $\boldsymbol{\lambda}$ and combining the resulting equation for $\hat{p}_i$ $(i = 1, \cdots, n)$, we have the relation

$$\sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i \; [y_{1i} - \boldsymbol{\beta}^{'} \mathbf{y}_{2i} - \boldsymbol{\gamma}^{'} \mathbf{z}_{1i}] = 0 \tag{2.8}$$

and

$$\hat{\boldsymbol{\lambda}} = [\sum_{i=1}^{n} \hat{p}_i u_i^2(\hat{\boldsymbol{\theta}}) \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} u_i(\hat{\boldsymbol{\theta}}) \mathbf{z}_i] \ , \tag{2.9}$$

where $u_i(\hat{\boldsymbol{\theta}}) = y_{1i} - \hat{\boldsymbol{\beta}}^{'} \mathbf{y}_{2i} - \hat{\boldsymbol{\gamma}}^{'} \mathbf{z}_{1i}$ and $\hat{\boldsymbol{\theta}}^{'} = (\hat{\boldsymbol{\beta}}^{'}, \hat{\boldsymbol{\gamma}}^{'})$ is the maximum empirical likelihood (MEL) estimator for the vector of unknown parameters $\boldsymbol{\theta}$.
In the actual computation we first minimize (2.7) with respect to λ and then the MEL estimator can be defined as the solution of constrained maximization of the criterion function with respect to $\boldsymbol{\theta}$ under the restrictions $0 < \epsilon \leq p_i < 1$ $(i = 1, \cdots, n)$, where we take a sufficiently small (positive) ϵ . Alternatively, from (2.7) the MEL estimator of $\boldsymbol{\theta}$ can be written as the solution of the set of p equations

$$\hat{\boldsymbol{\lambda}}^{'} \sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i [-(\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})] = 0 \ , \tag{2.10}$$

which implies

$$\begin{aligned} & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \ y_{1i}] \\ = \ & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})] \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} . \end{aligned} \tag{2.11}$$

On the other hand, a GMM estimator of $\boldsymbol{\theta}^{'} = (\boldsymbol{\beta}^{'}, \boldsymbol{\gamma}^{'})$ can be given by the solution of the equation [1]

$$\begin{aligned} & [\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\frac{1}{n} \sum_{i=1}^{n} u_i(\tilde{\boldsymbol{\theta}})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \ y_{1i}] \\ = \ & [\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}] [\frac{1}{n} \sum_{i=1}^{n} u_i(\tilde{\boldsymbol{\theta}})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1} [\frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})] \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} , \end{aligned} \tag{2.12}$$

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 $p_i = \frac{1}{n}$ $(i = 1, \cdots, n)$. In the actual computation we use the two-step efficient GMM procedure explained by Page 213 of Hayashi (2000), which seems to be standard in many empirical analyses.

By using the fact that $\log(1 + x) \sim x - x^2/2$ for small x and the expression of the Lagrangean multiplier vector in (2.9), it is possible to approximate the criterion function as

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

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

[1] This formulation is different from the original one. See Hayashi (2000) on the details of the GMM estimation method.

we have $(-1/2)(n-p)$ times the variance ratio

$$(2.13)\quad L_{2n}(\boldsymbol{\theta}) = \frac{[\sum_{i=1}^{n}\mathbf{z}_i^{'}(y_{1i}-\boldsymbol{\beta}^{'}\mathbf{y}_{2i}-\boldsymbol{\gamma}^{'}\mathbf{z}_{1i})][\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}]^{-1}[\sum_{i=1}^{n}\mathbf{z}_i(y_{1i}-\boldsymbol{\beta}^{'}\mathbf{y}_{2i}-\boldsymbol{\gamma}^{'}\mathbf{z}_{1i})]}{\sum_{i=1}^{n}(y_{1i}-\boldsymbol{\beta}^{'}\mathbf{y}_{2i}-\boldsymbol{\gamma}^{'}\mathbf{z}_{1i})^2}\ ,$$

where we can interpret the estimator of the homoscedastic variance in the form of $(n-p)\hat{\sigma}^2 = \sum_{i=1}^{n}(y_{1i}-\boldsymbol{\beta}^{'}\mathbf{y}_{2i}-\boldsymbol{\gamma}^{'}\mathbf{z}_{1i})^2$.
It has been known in the traditional econometrics that the LIML estimator is the minimum variance ratio estimation based on $L_{2n}(\boldsymbol{\theta})$ and is the maximum likelihood estimator with limited information under the normal disturbances on the disturbance terms of $\{\mathbf{v}_i\}$ (Anderson and Rubin (1949)). For this purpose, define the random matrices ($n \times G_1, n\times K, n\times K_1$, respectively) as $\mathbf{Y} = (\mathbf{y}_i^{'})$, $\mathbf{Z} = (\mathbf{z}_i^{'})$, and $\mathbf{Z}_1 = (\mathbf{z}_{1i}^{'})$. Also let two $(1+G_1)\times(1+G_1)$ random matrices be

$$(2.14)\qquad \mathbf{G} = \mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}$$

and

$$(2.15)\qquad \mathbf{H} = \mathbf{Y}^{'}(\mathbf{I}_n - \mathbf{P}_Z)\mathbf{Y}\ ,$$

where $\mathbf{P}_Z$ and $\mathbf{P}_{Z_1}$ and are the projection operators onto the space of $\mathbf{Z}$ and $\mathbf{Z}_1$, respectively. Then the LIML estimator $\hat{\mathbf{b}}_{LI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{LI}^{'})^{'})$ for the vector of coefficients $\mathbf{b}_0 = (1, -\boldsymbol{\beta}^{'})^{'}$ is given by

$$(2.16)\qquad (\mathbf{G} - \lambda\mathbf{H})\hat{\mathbf{b}}_{LI} = \mathbf{0}\ ,$$

where λ is the smallest root of

$$(2.17)\qquad |\mathbf{G} - l\mathbf{H}| = 0\ .$$

If we set $\lambda = 0$ in (2.16) and omit the second component, we have the TSLS estimator $\hat{\mathbf{b}}_{TS}$ $(= (1, -\beta_{TS}^{'})^{'})$ for the vector $\mathbf{b}_0 = (1, -\boldsymbol{\beta}^{'})^{'}$, which corresponds to minimizing the numerator of the variance ratio in (2.13). For the LIML estimator and TSLS estimator the coefficients of $\boldsymbol{\gamma}$ can be estimated by solving

$$(2.18)\qquad \hat{\gamma} = (\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Y}\hat{\mathbf{b}}\ ,$$

where $\hat{\mathbf{b}}$ is either $\hat{\mathbf{b}}_{LI}$ or $\hat{\mathbf{b}}_{TS}$.

Let the normalized error of estimators be in the form of

$$(2.19)\qquad \hat{\mathbf{e}} = \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix}\ ,$$

where $\hat{\boldsymbol{\theta}}^{'} = (\hat{\boldsymbol{\beta}}^{'}, \hat{\gamma}^{'})$ and $\boldsymbol{\theta}$ is the vector of unknown coefficient parameters. Under a set of regularity conditions in the standard large sample asymptotic theory including the assumption that both n and the noncentrality increase while K_2 is fixed [2] , the inverse of the asymptotic variance-covariance matrix of the asymptotically efficient estimators is

$$(2.20)\qquad \mathbf{Q}^{-1} = \mathbf{D}^{'}\mathbf{M}\mathbf{C}^{-1}\mathbf{M}\mathbf{D}\ ,$$

[2] See Qin and Lawles (1994) for the asymptotic covariance matrix in the i.i.d. case, which can be extended to more general situations.

where

$$(2.21) \qquad \mathbf{D} = [\mathbf{\Pi}_2 , \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}] ,$$

$$(2.22) \qquad \mathbf{M} = \mathbf{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^{'} ,$$

$$(2.23) \qquad \mathbf{C} = \mathbf{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_i^2 \mathbf{z}_i \mathbf{z}_i^{'} .$$

provided that the constant matrices $\mathbf{M}$ and $\mathbf{C}$ are positive definite, and the rank condition

$$(2.24) \qquad \mathbf{rank}(\mathbf{D}) = p \ (= G_1 + K_1) .$$

From (2.20) the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be written as the inverse of

$$(2.25) \qquad \mathbf{\Theta}^*(\beta) = \mathbf{\Pi}_{22}^{'} (\mathbf{M}\mathbf{C}^{-1}\mathbf{M})_{22.1} \mathbf{\Pi}_{22} ,$$

where we use the notation of a $K_2 \times K_2$ matrix $\mathbf{J}_{22.1} = \mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12}$ for any $(K_1 + K_2) \times (K_1 + K_2)$ partitioned matrix

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix} .$$

The above conditions assure that the limiting variance-covariance matrix $\mathbf{Q}$ is non-degenerate. The rank condition implies the order condition

$$(2.26) \qquad K - p = K_2 - G_1 \geq 0 ,$$

which has been called the degree of over-identification.

3. Evaluation of Distributions and Tables

3.1 Parameterizations

The estimation method of the cdf's of estimators we have used in this study is based on the simulation method developed by Kunitomo and Matsushita (2003a) except the TSLS method since the finite sample properties of alternative estimators are difficult to be investigated analytically. The exact distribution of the TSLS estimator was investigated by Anderson and Sawa (1979) systematically. In order to describe our estimation methods, we need to introduce some notations which are similar to the ones used by Anderson et. al. (1982) for the ease of comparison except the notation of sample size n for T. We shall concentrate on the comparison of the estimators of the coefficient parameter on the endogenous variable when $G_1 = 1$ in Sections 3 and 4 for presenting Figures and Tables.

Let $\mathbf{A} = \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^{'}$ and $\mathbf{M}$ be $K \times K$ matrices associated with (2.22), and we partition the nonsingular matrices $\mathbf{A}$ and $\mathbf{M}$ into $(K_1 + K_2) \times (K_1 + K_2)$ sub-matrices $\mathbf{A} = (\mathbf{A}_{ij})$ and $\mathbf{M} = (\mathbf{M}_{ij})$ $(i, j = 1, 2)$. When $\mathbf{C} = \sigma^2\mathbf{M}$, that is, the disturbance terms are homoscedastic ($\sigma^2 = \mathbf{E}[u_i^2]$ $(i = 1, \cdots, n)$), the (1,1) element of the inverse of the asymptotic variance-covariance matrix $\mathbf{Q}^{-1}$ can be expressed as $\boldsymbol{\Theta}^*(\beta) = \boldsymbol{\Pi}_{22}'\mathbf{M}_{22.1}\boldsymbol{\Pi}_{22}/\sigma^2$. Let

$$\boldsymbol{\Theta}_n(\beta) = \frac{1}{\sigma^2}\boldsymbol{\Pi}_{22}'\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22} \tag{3.1}$$

be the noncentrality parameter; the limit of $(1/n)\mathbf{A}_{22.1}$ is $\mathbf{M}_{22.1}$. There are some notations which lead to the *key parameters* used by Anderson et. al. (1982) on the study of the finite sample properties of the LIML and TSLS estimators in the classical parametric framework, and Kunitomo and Matsushita (2003a) on the MEL and GMM estimators in the semi-parametric case. In the rest of our study we shall consider the finite sample distribution for the coefficient of the endogenous variable β because of the simplification. We expect that we have similar results on other coefficients parameters in the more general cases.

We shall investigate the exact finite sample distributions of the normalized estimator as

$$[\boldsymbol{\Theta}_n(\beta)]^{1/2}(\hat{\beta} - \beta) , \tag{3.2}$$

where $\boldsymbol{\Theta}_n(\beta)$ is the (1,1)element of $n^{-1} \times \mathbf{Q}^{-1}$. The distribution of (3.2) for the LIML estimator and TSLS estimator depends only on the key parameters used by Anderson et. al. (1982) which are K_2,$n - K$,

$$\delta^2 = \frac{\boldsymbol{\Pi}_{22}'\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}}{\omega_{22}} , \tag{3.3}$$

and

$$\alpha = \frac{\omega_{22}\beta - \omega_{12}}{|\boldsymbol{\Omega}|^{1/2}} = \frac{\sqrt{\omega_{22}}}{\sqrt{\omega_{11.2}}}(\beta - \frac{\omega_{12}}{\omega_{22}}) . \tag{3.4}$$

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

$$-\frac{\alpha}{\sqrt{1 + \alpha^2}} = \frac{\omega_{12} - \omega_{22}\beta}{\sigma\sqrt{\omega_{22}}} ,$$

which is the correlation coefficient between the two random variables u_i and v_{2i} (or y_{2i}). It has been called 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. $n - K$ is the number of degree of freedom of $\mathbf{H}$ which estimates $\boldsymbol{\Omega}$ in the LIML method; it is not relevant to the TSLS method.

3.2 Simulation Procedures

By using a set of Monte Carlo simulations we can obtain the empirical cdf's of the MEL and GMM estimators for the coefficient of the endogenous variable in the structural equation of our interest. First, we consider the case when both the disturbances and the exogenous variables are normally distributed. We generate a set of random numbers by using the two equations system

$$y_{1i} = y_{2i}\beta^{(0)} + z_{1i}\gamma^{(0)} + u_i \ , \tag{3.5}$$

and

$$y_{2i} = \mathbf{z_i}'\boldsymbol{\pi}_2^{(0)} + v_{2i} \ , \tag{3.6}$$

where $\mathbf{z}_i \sim N(0, \mathbf{I}_K)$, $u_i \sim N(0,1)$, $v_{2i} \sim N(0,1)$ $(i = 1, \cdots, n)$, and we set the true values [3] of parameters $\beta^{(0)} = \gamma^{(0)} = 0$. We have controlled the values of δ^2 by choosing a real value of c and setting $(1 + K_2) \times 1$ vector $\boldsymbol{\pi}_2^{(0)} = c(1, \cdots, 1)'$. The model we have used has been restricted to the special case when $K_1 = 1$ because in general it takes prohibitively long computational time to estimate the empirical cdf of the MEL estimator when the number of parameters included the structural equation is large. For each simulation we have 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 eariler studies.

In order to investigate the effects of non-normal disturbances on the distributions of estimators, we took 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 (3.5), (3.6), and

$$u_i = -\frac{\chi_i^2(3) - 3}{\sqrt{6}} \ , \tag{3.7}$$

where $\chi_i^2(3)$ 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.

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 that

$$u_i = \|\mathbf{z}_i\| u_i^* \ (i = 1, \cdots, n) \tag{3.8}$$

where u_i^* $(i = 1, \cdots, n)$ are homoscedastic disturbance terms. In this case the matrix $\mathbf{C}$ of (2.20) 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 asymptotic theory.

3.3 Tables and Figures

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, that is, of (3.2); 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. The estimators with which we wish to compare (the LIML estimator with the TSLS estimator, or the MEL estimator with the GMM estimator), have the same asymptotic distribution. Therefore, the

[3] In order to examine whether our results strongly depend on the specific values of parameters $\beta^{(0)} = \gamma^{(0)} = 0$, we have done the several simulations for the values of $\beta^{(0)} \neq 0$ and $\gamma^{(0)} \neq 0$. These experiments suggest that our results holds in the general situation.

limiting distributions of (3.2) for the MEL and GMM estimators are $N(0,1)$ as $n \to \infty$ in the standard large sample asymptotic theory. We have summarized our results on the cdf's of four estimators in Tables of Appendix B.

3.4 Accuracy of the Procedures

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 (2.12) when $\hat{u}_i^2$ is replaced by a constant (namely σ^2), that is, the variance-covariance matrix is homoscedastic and known. (See (2.16) with $\lambda = 0$ for the TSLS estimator.) 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 estimates is about 0.008 (see Kunitomo and Matsushita (2003a) for the details). Hence our estimates of the cdf's are quite accurate and we have enough accuracy with two digits at least. 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.

4. Discussions on Distributions

4.1 Distributions of the MEL and LIML Estimators

The distributions are tabulated in standardized terms, that is, of (3.2). When we have the homoscedastic disturbance terms, the asymptotic standard deviation (ASD) of $\hat{\beta}$ is given by

$$\frac{\sigma}{\delta\sqrt{\omega_{22}}} = \frac{\sqrt{1+\alpha^2}\sqrt{|\mathbf{\Omega}|}}{\delta\omega_{22}} . \tag{4.1}$$

The spread of the distribution of the unstandardized estimator increases with $|\alpha|$ and decreases with δ . The other estimators with which we wish to compare the MEL estimator have the same asymptotic standard deviation and in the remainder of the discussion we consider the normalized distributions. For $\alpha = 0$, the densities 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 .

A 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 faily large. On the other hand, the distributions of the MEL and LIML estimators 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's. As δ^2 increases, the spread of the normalized distribution decreases. The distribution of the LIML estimator has slightly tigher 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 .

4.2 Distributions of the GMM and TSLS Estimators

We have given tables of the distributions of the GMM and TSLS estimators. Since they are quite similar in most cases, however, we often have given the distribution of the GMM estimator only in the 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.

4.3 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. We calculated a large number of distributions for disturbance terms including the χ^2 distribution and $t(5)$ distribution. The former represents the skewed distribution while the latter represents the distributions with longer tails. From these 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. We have found that the effects of heteroscedastic disturbances on the exact distributions of alternative estimators are not large in our setting, but we need some further investigations on this issue.

5 Many Instruments in Simultaneous Equations Models and an Asymptotic Optimality of the LIML Estimator

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 notation of weak instruments. One interpretation for weak instruments may be the case that we have a structural equation but the noncentrality parameter is not large in comparison with the sample size. Second, Bekker (1994) and others have pointed out that the standard asymptotic theory in econometrics may not be appropriate for practice when the number of instruments is large and the large K_2 theory would be suited better to applications by referring to the eariler studies of Kunitomo (1980) and Morimune (1983). The point was that there have been some microeconometric applications when many instruments have been used, but the applications of the GMM method give large biases in important estimates empirically. 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 or L is large in our notation) and it could be comparable to the size of noncentrality parameter. We should note that it is exactly the situation which Kunitomo (1982, 1987) investigated under a set of limited assumptions (which could have been removed).

In this section we consider a sequence of integers K_2 as $K_2(n)$ which can be dependent on the sample size n and $\mathbf{z}_{2i}(n)$ $(i = 1, \cdots, n)$ are a sequence of $K_2(n) \times 1$ vectors. We take the case when both K_1 and G_1 are fixed integers and $G_1 \geq 1$. Then we need to use the notations of the number of instruments $K(n)$ $(= K_1 + K_2(n))$ for K, the coefficients $\boldsymbol{\pi}_{21}(n)$ for $\boldsymbol{\pi}_{21}$ and $\boldsymbol{\Pi}_{22}(n)$ for $\boldsymbol{\Pi}_{22}$. We rewrite (2.3) as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi}(n) + \mathbf{V} \ , \tag{5.1}$$

where $\mathbf{Z} = (\mathbf{z}_i^{'}(n))$ 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 $\boldsymbol{\Pi}(n) = (\pi_1^{'}, \boldsymbol{\Pi}_2^{'}(n))^{'}$ is the $(K_1 + K_2(n)) \times (1 + G_1)$ matrix of coefficients. The restrictions on the coefficients can be expressed as $(1, -\boldsymbol{\beta}^{'})\boldsymbol{\Pi}^{'}(n) = (\boldsymbol{\gamma}^{'}, \mathbf{0}^{'})$ and $\boldsymbol{\pi}_{21}(n) = \boldsymbol{\Pi}_{22}(n)\boldsymbol{\beta}$.

We first state the asymptotic distribution of the LIML estimator under a set of simplified assumptions when $K_2(n)$ can be dependent on n and $n \to +\infty$. The proof will be given in Appendix A.

Theorem 5.1 : Assume that (2.1) and (2.3) hold with $\mathbf{v}_1, \cdots, \mathbf{v}_n$ independently distributed each according to $N(\mathbf{0}, \boldsymbol{\Omega})$. Suppose further 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 $a(n) \longrightarrow \infty$. Suppose

$$\textbf{(I)} \qquad \frac{K_2(n)}{a(n)} \longrightarrow c_1 \ \ (0 \leq c_1 < \infty),$$

$$\textbf{(II)} \qquad \frac{1}{n}\mathbf{A}_{11} \stackrel{p}{\longrightarrow} \boldsymbol{\Phi}_{11} \ ,$$

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

$$\textbf{(IV)} \qquad \frac{K_2(n)}{q(n)} \longrightarrow c_2 \ \ (0 \leq c_2 < \infty),$$

where $\boldsymbol{\Phi}_{11}$ and $\boldsymbol{\Phi}_{22.1}$ are nonsingular constant matrices.

Then

$$\left[\boldsymbol{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \boldsymbol{\Psi}^*) \ , \tag{5.2}$$

where

$$\boldsymbol{\Psi}^* = \sigma^2\mathbf{I}_{G_1} + c_1(1 + c_2)\boldsymbol{\Phi}_{22.1}^{-1/2}\left[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\mathbf{b}_0\mathbf{b}_0^{'}\boldsymbol{\Omega}\right]_{22}\boldsymbol{\Phi}_{22.1}^{-1/2} \tag{5.3}$$

and $\sigma^2 = \mathbf{b}_0' \mathbf{\Omega} \mathbf{b}_0$. Alternatively

$$\left[\frac{\mathbf{\Pi}_{22}'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)}{a(n)}\right]\sqrt{a(n)}(\hat{\boldsymbol{\beta}}_{LI}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^{**}) \ , \tag{5.4}$$

where

$$\mathbf{\Psi}^{**} = \sigma^2 \mathbf{\Phi}_{22.1} + c_1(1+c_2)\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\mathbf{b}_0\mathbf{b}_0'\mathbf{\Omega}\right]_{22} \ . \tag{5.5}$$

Alternatively

$$\sqrt{a(n)}(\hat{\boldsymbol{\beta}}_{LI}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^{***}) \ , \tag{5.6}$$

where

$$\mathbf{\Psi}^{***} = \sigma^2 \mathbf{\Phi}_{22.1}^{-1} + c_1(1+c_2)\mathbf{\Phi}_{22.1}^{-1}\left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\mathbf{b}_0\mathbf{b}_0'\mathbf{\Omega}\right]_{22}\mathbf{\Phi}_{22.1}^{-1} \ . \tag{5.7}$$

If $G_1 = 1$, we have $[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\mathbf{b}_0\mathbf{b}_0'\mathbf{\Omega}]_{22} = \omega_{11}\omega_{22} - \omega_{12}^2 = |\mathbf{\Omega}|$.

In order to compare our results in *Theorem 5.1* with the standard asymptotic theory, suppose $K_2(n)$ is fixed and $(1/n)\mathbf{A}_{22.1} \xrightarrow{p} \mathbf{M}_{22.1}$ (nonsingular) as $n \to \infty$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{LI}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2(\mathbf{\Pi}_{22}'\mathbf{M}_{22.1}\mathbf{\Pi}_{22})^{-1}) \ . \tag{5.8}$$

We note that in this case $q(n)/n \to 1$ and $K_2(n)/n \to 0$.

For the estimation problem of the vector of structural parameters $\boldsymbol{\beta}$, it may be natural to consider a set of statistics of two $(1+G_1)\times(1+G_1)$ 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 will be also given in Appendix A.

Theorem 5.2 : Assume that (2.1) and (2.3) hold and define the class of consistent estimators for $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}} = \phi(\frac{1}{a(n)}\mathbf{G}, \frac{1}{q(n)}\mathbf{H}) \ , \tag{5.9}$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices in (5.9) as $a(n) \to \infty, q(n) \to \infty$ $(n \to \infty)$. We also define the normalized form of consistent estimators for $\boldsymbol{\beta}$ by $\hat{\mathbf{e}}(\boldsymbol{\beta}) = \left[\mathbf{\Pi}_{22}'(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n)\right]^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$. Then under the assumptions of *Theorem 5.1*

$$\hat{\mathbf{e}}(\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}) \ , \tag{5.10}$$

where

$$\mathbf{\Psi} \geq \mathbf{\Psi}^* \tag{5.11}$$

in the sense of positive definiteness and $\mathbf{\Psi}^*$ is given by (5.3).

The above theorems are the generalized versions of the results given by Kunitomo (1982) or Theorem 3.1 of Kunitomo (1987). Although we have assumed that the disturbances are normally distributed and they are homoscedastic, it is certainly possible to replace the normality assumption by some moments conditions on disturbances and certain type of heteroscedasticity assumptions. Then we need lengthy discussions on the technical details of derivations for the asymptotic normality of a sequence of $G_1 \times (1+G_1)$ random matrix

$$\frac{1}{\sqrt{a(n)}} \mathbf{\Pi}_{22}^{'}(n) \mathbf{Z}_2^{'} (\mathbf{I}_n - \mathbf{P}_{Z_1}) \mathbf{V} \tag{5.12}$$

and $G_1 \times (1+G_1)$ random matrix

$$\sqrt{q(n)} [\frac{1}{q(n)} \mathbf{Y}^{'} (\mathbf{I}_n - \mathbf{P}_Z) \mathbf{Y} - \mathbf{\Omega}] \tag{5.13}$$

for a sequence of normalized constants $a(n)$ and $q(n)$ ($a(n), q(n) \to \infty$ as $n \to \infty$). Also the assumption of independence between $\mathbf{v}_i$ and $\mathbf{z}_i(n)$ ($i = 1, \cdots, n$) can be relaxed. By using different notations, Stock and Yogo (2003), and Hansen et. al. (2004) have discussed a set of assumptions for the asymptotic normality of similar quantities.

The results for the simplest case when $K_2(n)$ is fixed and we take $a(n) = n$ as the normalization factor 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 the second case, called the large K_2-asymptotic theory, has not been treated in formal ways. 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, while both the TSLS estimator and the GMM estimator are inconsistent when $c_1 > 0$ and $c_2 \geq 0$. This is a non-regular situation because the number of incidental parameters increases as $K_2(n)$ increases in the simultaneous equation models. The statsitical reasons will be clearer if we formulate the simultaneous equation models as the linear functional relationship model in the statistical literature or the errors-in-variables model which we shall explore at the end of this section. 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 $K_2(n)/\mathbf{tr}\mathbf{\Theta}_n(\beta) \to 0$ in probability. In this case we have $c_1 = 0$ and the asymptotic lower bound of the covariance matrix is the same as the case of the large sample asymptotic theory. It may be also possible to show that the MEL estimator can attain the lower bound with some stronger conditions.

Furthermore, Kunitomo (1982) has already investigated the higher order efficiency property of the LIML estimation method under a set of restrictive assumptions when $G_1 = 1, c_1 > 0$ and $c_2 = 0$. These considerations of this section shall give some new light on the practical use of estimation methods in microeconometric models with weak instruments or many instruments. Since the LIML estimator has 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 we have discussed in the previous sections.

Let $\mathbf{p}_{21}$ and $\mathbf{P}_{22}$ be $K_2(n) \times 1$ and $K_2(n) \times G_1$ matrices which are the least squares estimators of the corresponding parameters in $\boldsymbol{\pi}_{21}(n)$ and $\mathbf{\Pi}_{22}(n)$, respectively. Then we define $K_2(n) \times 1$ vector $\mathbf{x}_1 = (x_{1i})$ and $K_2(n) \times G_1$ matrix $\mathbf{X}_2 = (\mathbf{x}_{2i}^{'})$ by

$$\mathbf{x}_1 = \mathbf{A}_{22.1}^{1/2} \mathbf{p}_{21} \ \ , \mathbf{X}_2 = \mathbf{A}_{22.1}^{1/2} \mathbf{P}_{22} \ , \tag{5.14}$$

where $\mathbf{x}_{2i}^{'}$ $(i = 1, \cdots, K_2(n))$ are $1 \times G_1$ vectors, and we have $1 \times K(n)$ $(K(n) = K_1 + K_2(n))$ partitioned vectors $\mathbf{z}_i^{'} = (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{'})$ $(i = 1, \cdots, n)$. Also we define $K_2(n) \times 1$ vector $\boldsymbol{\eta} = (\eta_i)$ and $K_2(n) \times G_1$ matrix $\boldsymbol{\Xi} = (\xi_{ij})$ by

$$\boldsymbol{\eta} = \mathbf{A}_{22.1}^{1/2}\boldsymbol{\pi}_{21}(n) \ , \boldsymbol{\Xi} = \mathbf{A}_{22.1}^{1/2}\boldsymbol{\Pi}_{22}(n) \ . \tag{5.15}$$

The information matrix for β (or the noncentrality parameter in the structural equation estimation) under the assumption of the homoscedasticity and normality for the disturbance terms can be rewritten [4] as

$$\boldsymbol{\Theta}_n(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^{K_2(n)} \boldsymbol{\xi}_i\boldsymbol{\xi}_i^{'} \ , \tag{5.16}$$

where $\boldsymbol{\xi}_i^{'}$ $(i = 1, \cdots, K_2(n))$ is the i-th row vector of $\boldsymbol{\Xi}$, $\mathbf{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \boldsymbol{\Omega}$ $(i = 1, \cdots, n)$, and $\sigma^2 = \mathbf{b}_0^{'}\boldsymbol{\Omega}\mathbf{b}_0$. Then we have the representation of the linear functional relationships model in the statistical literatures as

$$(\mathbf{x}_1, \mathbf{X}_2) = (\boldsymbol{\eta}, \boldsymbol{\Xi}) + (\mathbf{w}_1, \mathbf{W}_2) \ , \tag{5.17}$$

where we have defined a $K_2(n) \times (1 + G_1)$ random matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{W}_2)$ and the variance-covariance matrix of the i-th row vector $\mathbf{w}_i^{'}$ $(i = 1, \cdots, K_2(n))$ of $\mathbf{W}$ is given by $\boldsymbol{\Omega}$. This model has been called the errors-in-variables model in econometric literatures and the linear functional relationships model in the statistical literatures because we have the statistical relation

$$\boldsymbol{\eta} = \boldsymbol{\Xi}\boldsymbol{\beta} \tag{5.18}$$

and the number of incidental parameters of $\boldsymbol{\Xi} = (\xi_{ij})$ can be large when $K_2(n)$, which is the sample size in a sense, is large. 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 literatures there have been several eariler studies including Kunitomo (1980, 1982), Morimune (1983), and Bekker (1994).

Anderson (1976, 1984) 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 (5.17) and (5.18). This observation gives the persuasive reason why we have finite sample properties of the TSLS and GMM estimators discussed in the previous sections.

6. 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 estimator if K_2 is large. In some microeconometric models and models

[4] It is another expression of (3.1) when $G_1 \geq 1$ in the form of the linear functional relationship model or the errors-in-variables model.

on panel data, it is often a common feature that K_2 is fairly large. In such situations the LIML estimator has asymptotically optimal property in the large K_2-asymptotics and the finite sample properties of the MEL estimator is very similar.

Second, the large-sample normal approximation is relatively accurate for the MEL and LIML estimators. Hence the usual methods with asymptotic standard deviations give reasonable inferences. 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 MEL and LIML estimators do not posses 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, GMM, 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.

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

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

[10] 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.

[11] 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.

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

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

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

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

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

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

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

[19] 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.

[20] 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)

[21] 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)

[22] 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.

[23] 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.

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

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

[26] 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.

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

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

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

In this Appendix A we give the proofs of *Theorem 5.1* and *Theorem 5.2*. We freely use the notations of $\mathbf{vec}(A)$ for stacking column vectors of any matrix $\mathbf{A}$ and the transpose of $\mathbf{A}$ as $\mathbf{A}^{'}$.

Proof of Theorem 5.1 :
By substituting (5.1) into (2.14), we have

$$\begin{aligned}\mathbf{G} &= (\mathbf{\Pi}^{'}(n)\mathbf{Z}^{'}+\mathbf{V}^{'})(\mathbf{P}_Z-\mathbf{P}_{Z_1})(\mathbf{Z}\mathbf{\Pi}(n)+\mathbf{V}) \\ &= \mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2\mathbf{\Pi}_2(n)+\mathbf{V}^{'}(\mathbf{P}_Z-\mathbf{P}_{Z_1})\mathbf{V}+\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}+\mathbf{V}^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2\mathbf{\Pi}_2(n)\ .\end{aligned}$$

Then by using the similar arguments for partitioned matrices as *Theorem A.3.3* of Anderson (2003), we have

$$(6.1)\qquad \begin{aligned}\mathbf{V}^{'}(\mathbf{P}_Z-\mathbf{P}_{Z_1})\mathbf{V} &= \mathbf{V}^{'}\mathbf{Z}\begin{pmatrix}-\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\\ \mathbf{I}_{K_2(n)}\end{pmatrix}\mathbf{A}_{22.1}^{-1}(-\mathbf{A}_{21}\mathbf{A}_{11}^{-1},\mathbf{I}_{K_2(n)})\mathbf{Z}^{'}\mathbf{V} \\ &= \mathbf{V}^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2\mathbf{A}_{22.1}^{-1}\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}\ ,\end{aligned}$$

where $\mathbf{A}_{11}=\sum_{i=1}^{n}\mathbf{z}_{1i}\mathbf{z}_{1i}^{'}$, $\mathbf{A}_{12}=\sum_{i=1}^{n}\mathbf{z}_{1i}\mathbf{z}_{2i}^{'}(n)$ and $\mathbf{A}_{22.1}=\sum_{i=1}^{n}\mathbf{z}_{2i}(n)\mathbf{z}_{2i}^{'}(n)-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. Then we rewrite

$$(6.2)\qquad \begin{aligned}&\mathbf{G}-[\mathbf{\Pi}_2^{'}(n)\mathbf{A}_{22.1}^{'}\mathbf{\Pi}_2(n)+K_2(n)\mathbf{\Omega}] \\ &= \mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}+\mathbf{V}^{'}\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{\Pi}_2+(\mathbf{V}^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2\mathbf{A}_{22.1}^{-1}\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}-K_2(n)\mathbf{\Omega})\ .\end{aligned}$$

We note that conditional on the instrumental variables $\mathbf{Z}$ two random matrices $\mathbf{V}^{'}\bar{P}_{Z_1}\mathbf{Z}_2\mathbf{A}_{22.1}^{-1}\mathbf{Z}_2^{'}\bar{P}_{Z_1}\mathbf{V}$ and $\mathbf{V}^{'}\bar{\mathbf{P}}_Z\mathbf{V}$ $(=\mathbf{H})$ have the distributions of $\sum_{i=K_1+1}^{K(n)}\mathbf{w}_i\mathbf{w}_i^{'}$ and $\sum_{i=K(n)+1}^{n}\mathbf{w}_i\mathbf{w}_i^{'}$, respectively, when $\mathbf{w}_i$ $(i=K_1+1,\cdots,n)$ have the distribution of $N(\mathbf{0},\mathbf{\Omega})$ with $K_2(n)=K(n)-K_1$ and $q(n)=n-K(n)$. Since they do not depend on the conditions $\mathbf{Z}$, they are unconditionally the central Whishart distributions $\mathcal{W}_{1+G_1}(\mathbf{\Omega},K_2(n))$ and $\mathcal{W}_{1+G_1}(\mathbf{\Omega},q(n))$, respectively. As $a(n)\to\infty$, we have the convergence in probability as

$$(6.3)\qquad \frac{1}{a(n)}\mathbf{G}\xrightarrow{p}\mathbf{G}_0=\begin{pmatrix}\boldsymbol{\beta}^{'}\\ \mathbf{I}_{G_1}\end{pmatrix}\mathbf{\Phi}_{22.1}(\boldsymbol{\beta},\mathbf{I}_{G_1})+c_1\mathbf{\Omega}$$

and

$$(6.4)\qquad \frac{1}{q(n)}\mathbf{H}\xrightarrow{p}\mathbf{\Omega}\ .$$

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

$$(6.5)\qquad [\frac{1}{a(n)}\mathbf{G}-\lambda(n)\frac{1}{q(n)}\mathbf{H}]\hat{\mathbf{b}}_{LI}=\mathbf{0}\ .$$

Then it is straightforward to show that the probability limit of the LIML estimator $\hat{\mathbf{b}}_{LI}=(1,\hat{\boldsymbol{\beta}}_{LI}^{'})^{'}$ is $\mathbf{b}_0=(1,-\boldsymbol{\beta}^{'})^{'}$ as $n\to+\infty$ and $\lambda(n)\xrightarrow{p}\lambda_0$, where

$$(6.6)\qquad \lambda_0\mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{b}_0=\mathbf{b}_0^{'}\mathbf{G}_0\mathbf{b}_0\ .$$

Let $\hat{\mathbf{G}}_1 = \sqrt{a(n)}[(1/a(n))\mathbf{G} - \mathbf{G}_0]$, $\lambda_1 = \sqrt{a(n)}[\lambda - \lambda_0]$, $\hat{\mathbf{b}}_1 = \sqrt{a(n)}[\hat{\mathbf{b}}_{LI} - \mathbf{b}_0]$, $\hat{\mathbf{H}}_1 = \sqrt{q(n)}[(1/q(n))\mathbf{H} - \mathbf{\Omega}]$. Then by using the $(1+G_1)\times G_1$ choice matrix $\mathbf{J}_2 = (\mathbf{0}, \mathbf{I}_{G_1})^{'}$, we can write $\hat{\mathbf{b}}_1 = (-1)\mathbf{J}_2\sqrt{a(n)}[\hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta}]$. By substituing the random variables $\hat{\mathbf{G}}_1$, $\hat{\mathbf{H}}_1$, and λ_1 into (6.4), the resulting relation becomes

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

Then by ignoring the higher order terms and using the fact $\lambda_0 = c_1$, we shall consider the modified estimator $\mathbf{e}^*_{LI}(\beta)$ which satisfies

$$(6.7)\qquad [\mathbf{G}_0 - \lambda_0\mathbf{\Omega}]\mathbf{J}_2\mathbf{e}^*_{LI}(\beta) = [\hat{\mathbf{G}}_1 - \lambda_1\mathbf{\Omega}]\mathbf{b}_0 - \sqrt{c_1c_2}\hat{\mathbf{H}}_1\mathbf{b}_0 \;.$$

By defining the normarized (LIML) random vector $\hat{\mathbf{e}}_{LI}(\beta) = \sqrt{a(n)}[\hat{\beta}_{LI} - \beta]$, we can show that $\mathbf{e}^*_{LI}(\beta) = \hat{\mathbf{e}}_{LI}(\beta) + o_p(1)$. By multiplying $\mathbf{J}_2^{'}$ and $\mathbf{b}_0^{'}$ from the left-hand-side of (6.6), we have the relations

$$(6.8)\qquad \mathbf{J}_2^{'}(\mathbf{G}_0 - \lambda_0\mathbf{\Omega})\mathbf{J}_2\mathbf{e}^*_{LI}(\beta) = \mathbf{J}_2^{'}(\hat{\mathbf{G}}_1 - \lambda_1\mathbf{\Omega} - \sqrt{c_1c_2}\hat{\mathbf{H}}_1)\mathbf{b}_0 \;,$$

and

$$(6.9)\qquad \mathbf{b}_0^{'}(\mathbf{G}_0 - \lambda_0\mathbf{\Omega})\mathbf{J}_2\mathbf{e}^*_{LI}(\beta) = \mathbf{b}_0^{'}(\hat{\mathbf{G}}_1 - \lambda_1\mathbf{\Omega} - \sqrt{c_1c_2}\lambda_0\hat{\mathbf{H}}_1)\mathbf{b}_0 \;.$$

Since $(\mathbf{G}_0 - \lambda_0\mathbf{\Omega})\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{J}_2^{'}(\mathbf{G}_0 - \lambda_0\mathbf{\Omega})\mathbf{J}_2 = \mathbf{\Phi}_{22.1}$, we can simplify these relations as

$$\lambda_1 = \frac{\mathbf{b}_0^{'}(\hat{\mathbf{G}}_1 - \sqrt{c_1c_2}\hat{\mathbf{H}}_1)\mathbf{b}_0}{\mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{b}_0} \;,$$

and then

$$\begin{aligned}
(6.10)\qquad \mathbf{e}^*_{LI}(\beta) &= [\mathbf{J}_2^{'}(\mathbf{G}_0 - \lambda_0\mathbf{\Omega})\mathbf{J}_2]^{-1}[\mathbf{J}_2^{'}(\hat{\mathbf{G}}_1 - \lambda_1\mathbf{\Omega} - \sqrt{c_1c_2}\hat{\mathbf{H}}_1)\mathbf{b}_0] \\
&= \Phi_{22.1}^{-1}\mathbf{J}_2^{'}[\mathbf{I}_{G_1+1} - \frac{\mathbf{\Omega}\mathbf{b}_0\mathbf{b}_0^{'}}{\mathbf{b}^{'}\mathbf{\Omega}\mathbf{b}_0}](\hat{\mathbf{G}}_1 - \sqrt{c_1c_2}\hat{\mathbf{H}}_1)\mathbf{b}_0 \;.
\end{aligned}$$

We notice that

$$\begin{aligned}
(6.11)(\hat{\mathbf{G}}_1 - \sqrt{c_1c_2}\hat{\mathbf{H}}_1)\mathbf{b}_0 &= \frac{1}{\sqrt{a(n)}}\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}\mathbf{b}_0 \\
&+\sqrt{\frac{K_2(n)}{a(n)}}\frac{1}{\sqrt{K_2(n)}}\sum_{i=K_1+1}^{K(n)}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})\mathbf{b}_0 - \sqrt{c_1c_2}\frac{1}{\sqrt{q(n)}}\sum_{i=K(n)+1}^{n}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})\mathbf{b}_0,
\end{aligned}$$

where $K(n) + q(n) = n$. Then the asymptotic distributions of each terms on the right-hand side are normal. In order to obtain the asymptotic covariance matrix of (6.10), we use the conditional expectation given $\mathbf{Z}$ as

$$\mathbf{E}\left[\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V}\mathbf{b}_0\mathbf{b}_0^{'}\mathbf{V}^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{Z}_2\mathbf{\Pi}_2(n)|\mathbf{Z}\right] = \mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{b}_0\mathbf{\Pi}_{22}^{'}(n)\mathbf{A}_{22.1}\mathbf{\Pi}_{22}(n) \;.$$

Then by using Condition (III) in *Theorem 5.1*, we find that the covariance matrix of the first term of (6.11) is given by $\sigma^2\mathbf{\Phi}_{22.1}$. For the second and third terms of (6.11) we can use the relation

$$\mathbf{E}[\frac{1}{K_2(n)}\sum_{i=K_1+1}^{K(n)}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})\mathbf{b}_0\mathbf{b}_0^{'}\sum_{i=K_1+1}^{K(n)}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})] = \mathbf{\Omega}\mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{b}_0 + \mathbf{\Omega}\mathbf{b}\mathbf{b}_0^{'}\mathbf{\Omega}$$

because $\mathbf{w}_i$ $(i = K_1 + 1, \cdots, n)$ are normally distributed as $N(\mathbf{0}, \mathbf{\Omega})$. Then by using direct calculations as

$$\mathbf{J}_2^{'}[\mathbf{I}_{G_1+1} - \frac{\mathbf{\Omega b}_0\mathbf{b}_0^{'}}{\mathbf{b}^{'}\mathbf{\Omega b}_0}]\mathbf{\Pi}_2^{'}(n) = \mathbf{\Pi}_{22}^{'}(n)$$

and

$$\mathbf{J}_2^{'}[\mathbf{I}_{G_1+1} - \frac{\mathbf{\Omega b}_0\mathbf{b}_0^{'}}{\mathbf{b}^{'}\mathbf{\Omega b}_0}][\sigma^2\mathbf{\Omega} - \mathbf{\Omega b}_0\mathbf{b}_0^{'}\mathbf{\Omega}][\mathbf{I}_{G_1+1} - \frac{\mathbf{\Omega b}_0\mathbf{b}_0^{'}}{\mathbf{b}^{'}\mathbf{\Omega b}_0}]\mathbf{J}_2 = \mathbf{J}_2^{'}[\sigma^2\mathbf{\Omega} - \mathbf{\Omega b}_0\mathbf{b}_0^{'}\mathbf{\Omega}]\mathbf{J}_2 \ ,$$

we find that the asymptotic covariance matrix of the normalized LIML estimator is given by $\mathbf{\Psi}^*$.
When $G_1 = 1$, we can use the relation $\sigma^2 = \omega_{11} - 2\beta\omega_{12} + \beta^2\omega_{22}$ for $\mathbf{\Omega} = (\omega_{ij})$ to obtain $\sigma^2\omega_{22} - (\omega_{12} - \beta\omega_{22})^2 = |\mathbf{\Omega}|$.
(**Q.E.D**)

Proof of Theorem 5.2 :
We use a $(1 + G_1) \times (1 + G_1)$ nonsingular matrix $\mathbf{B} = (\mathbf{b}_0, \mathbf{B}_2)$ with $\mathbf{b}_0 = (1, -\beta^{'})^{'}$ for the transformation such that $\mathbf{B}^{'}\mathbf{\Omega B} = \mathbf{\Sigma}$, which is the block-diagonal matrix given by

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & \mathbf{0}^{'} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix} \tag{6.12}$$

and $\mathbf{\Sigma}_{22} = \mathbf{B}_2^{'}\mathbf{\Omega B}_2$. (Actually by choosing some $\mathbf{B}_2$, we can have $\mathbf{\Sigma}_{22} = \mathbf{I}_{G_1}$. See *Theorem A.2.2* of Anderson (2003).) Then we define $(1 + G_1) \times (1 + G_1)$ random matrices $\mathbf{G}^* = \mathbf{\Sigma}^{-1/2}\mathbf{B}^{'}\mathbf{GB\Sigma}^{-1/2}$ and $\mathbf{H}^* = \mathbf{\Sigma}^{-1/2}\mathbf{B}^{'}\mathbf{HB\Sigma}^{-1/2}$.

By using the assumptions of *Theorem 5.1*, we have the convergence in probability as $a(n)$ goes to $+\infty$ that

$$\frac{1}{a(n)}\mathbf{G}^* \xrightarrow{p} \mathbf{G}_0^* = \begin{pmatrix} 0 & \mathbf{0}^{'} \\ \mathbf{0} & \mathbf{E}^{'}\mathbf{\Phi}_{22.1}\mathbf{E} \end{pmatrix} + c_1\mathbf{I}_{1+G_1} \ , \tag{6.13}$$

where $G_1 \times G_1$ (positive definite) matrix $\mathbf{\Phi}_{22.1}$ is given by Condition (III) and $\mathbf{E}^{'} = \mathbf{B}_2^{'}(\boldsymbol{\beta}, \mathbf{I}_{G_1})^{'}$. We note that by solving $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_{1+G_1}$ with $\mathbf{B} = (\mathbf{b}_0, \mathbf{B}_2)$, we have the relation

$$\mathbf{E}^{-1} = [(\boldsymbol{\beta}, \mathbf{I}_{G_1})\mathbf{B}_2]^{-1} = \mathbf{J}_2^{'}\mathbf{B}^{-1}\mathbf{J}_2.$$

We notice that the first $(1+G_1)$ elements of the random vector $\mathbf{vec}[\sqrt{a(n)}(\frac{1}{a(n)}\mathbf{G} - \mathbf{G}_0)]$ can be written as

$$\begin{aligned} \mathbf{g}_1(n) &= \mathbf{\Sigma}^{-1/2}\mathbf{B}^{'}\frac{1}{\sqrt{a(n)}}[\frac{1}{a(n)}\mathbf{G} - \mathbf{G}_0]\mathbf{b}_0\sigma^{-1} \\ &= \mathbf{\Sigma}^{-1/2}\mathbf{B}^{'}[\frac{1}{\sqrt{a(n)}}\mathbf{\Pi}_2^{'}(n)\mathbf{Z}_2^{'}\bar{\mathbf{P}}_{Z_1}\mathbf{V} + \sqrt{\frac{K_2(n)}{a(n)}}\frac{1}{\sqrt{K_2(n)}}\sum_{i=K_1+1}^{K(n)}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})]\mathbf{b}_0\sigma^{-1} \ . \end{aligned}$$

Then we rewrite

$$\mathbf{g}_1(n) = \begin{bmatrix} \mathbf{0} \\ \mathbf{E}^{'}\mathbf{\Phi}_{22.1}^{1/2}\mathbf{w}_{21}(n) \end{bmatrix} + \sqrt{c_1}\begin{bmatrix} x_{11}(n) \\ \mathbf{x}_{21}(n) \end{bmatrix} \ , \tag{6.14}$$

where the $G_1 \times 1$ vector $\mathbf{w}_{21}(n)$ follows $N(\mathbf{0}, \mathbf{I}_{G_1})$ and $\mathbf{x}_{21}(n)$ is an asymptotically normal random vector as $N(\mathbf{0}, \mathbf{I}_{G_1})$ and $\mathbf{x}_{11}(n)$ is an asymptotically normal random variable as $N(0,2)$.

Similarly, we consider the first $(1+G_1)$ elements of the random vector $\mathbf{vec}[\sqrt{q(n)}(\frac{1}{q(n)}\mathbf{H}^* - \mathbf{I}_{1+G_1})]$ can represented as

$$\text{(6.15)} \qquad \mathbf{h}_1(n) = \mathbf{\Sigma}^{-1/2}\mathbf{B}^{'}\sqrt{q(n)}[\frac{1}{q(n)}\sum_{i=K(n)+1}^{n}(\mathbf{w}_i\mathbf{w}_i^{'} - \mathbf{\Omega})]\mathbf{b}_0 = \begin{bmatrix} h_{11}(n) \\ \mathbf{h}_{21}(n) \end{bmatrix} ,$$

where the $G_1 \times 1$ vector $\mathbf{h}_{21}(n)$ is an asymptotically normal random vector as $N(\mathbf{0}, \mathbf{I}_{G_1})$ and $h_{11}(n)$ is an asymptotically normal random variable as $N(0,2)$.

As the second step we shall derive the lower bound of the asymptotic variance-covariance matrices. For this purpose, we write a $G_1 \times 1$ random vector $\boldsymbol{\phi} = (\phi_i)$ and $\phi_i = \hat{\beta}_i$ $(i = 1, \cdots, G_1)$. Then the conditions for consistency of the class of estimators for $\boldsymbol{\beta} = (\beta_i)$ can be summarized by

$$\text{(6.16)} \qquad \phi_i((\begin{matrix} \boldsymbol{\beta}^{'} \\ \mathbf{I}_{G_1} \end{matrix})\mathbf{\Phi}_{22.1}(\boldsymbol{\beta}, \mathbf{I}_{G_1}) + c_1\mathbf{\Omega}, \mathbf{\Omega}) = \beta_i \ (i = 1, \cdots, G_1) \ .$$

For the $(1+G_1) \times (1+G_1)$ random matrix $\mathbf{G} = (g_{ij})$ we define $(1+G_1)^2 \times 1$ vectors of derivatives

$$\text{(6.17)} \qquad \boldsymbol{\phi}^{(k)} = \mathbf{vec}[(\frac{\partial \phi_k}{\partial g_{ij}})] \ (k = 1, \cdots, G_1)$$

which are evaluated at the probability limits of $(1/a(n))\mathbf{G}$ and $(1/q(n))\mathbf{H}$ as $\mathbf{G}_0 = \mathbf{B}^{'-1}\mathbf{\Sigma}^{1/2}\mathbf{G}_0^*\mathbf{\Sigma}^{1/2}\mathbf{B}^{-1}$ and $\mathbf{\Omega}$, respectively, when $a(n) \to \infty$ as $n \to \infty$.
By differentiating (6.16) with respect to each elements of $\boldsymbol{\beta} = (\beta_i)$ and $\mathbf{\Phi}_{22.1}$ $(= (\Phi_{22.1ij}))$ $(i, j = 1, \cdots, G_1)$, the resulting relations can be written as the form of

$$\text{(6.18)} \ \mathbf{A}^{'}(\boldsymbol{\phi}^{(1)}, \cdots, \boldsymbol{\phi}^{(G_1)}) = \mathbf{e}_1 \otimes \mathbf{e}_1(\phi_{11}^{(1)}, \cdots, \phi_{11}^{(G_1)}) + (\mathbf{e}_1 \otimes \mathbf{e}_2, \cdots, \mathbf{e}_1 \otimes \mathbf{e}_{1+G_1}) \ ,$$

where $\phi_{11}^{(k)}$ are the first element of vectors $\boldsymbol{\phi}^{(k)}$ $(k = 1, \cdots, G_1)$ and $\mathbf{A}$ is a $(1+G_1)^2 \times (1+G_1)^2$ matrix consisting of vectors :

$\mathbf{e}_1 \otimes \mathbf{e}_1$,

$\frac{\partial \mathbf{vec}}{\partial \beta_i}[(\boldsymbol{\beta}, \mathbf{I}_{G_1})^{'}\mathbf{\Phi}_{22.1}(\boldsymbol{\beta}, \mathbf{I}_{G_1})]$ $(i = 1, \cdots, G_1)$,

$\mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i$ $(i < j; i, j = 1, \cdots, 1+G_1)$,

$\frac{\partial \mathbf{vec}}{\partial \Phi_{22.1ij}}[(\boldsymbol{\beta}, \mathbf{I}_{G_1})^{'}\mathbf{\Phi}_{22.1}(\boldsymbol{\beta}, \mathbf{I}_{G_1})]$ $(i < j; i, j = 1, \cdots, G_1)$.

In the above expression (6.18) we have used the restrictions that the random matrix $\mathbf{G}$ and a matrix $\mathbf{\Phi}_{22.1}$ are symmetric, and adopted the notation that $\mathbf{e}_i$ $(i = 1, \cdots, 1+G_1)$ are unit vectors with 1 in the i-th element and zeros in other elements. (There are $1 + [G_1] + [(G_1+1)G_1/2] + [G_1(G_1+1)/2] = (1+G_1)^2$ elements in $\mathbf{A}$ [5] .)

[5] For an illustration when $G_1 = 1$, we set a 4×1 vector $\boldsymbol{\phi}^{(1)} = (\phi_{11}^{(1)}, \phi_{21}^{(1)}, \phi_{12}^{(1)}, \phi_{22}^{(1)})^{'}$ and $\rho = \Phi_{22.1}$. Then the restrictions for consistency can be represented by a 4×4 matrix

$$\mathbf{A} = \left[(\begin{matrix} 1 \\ 0 \end{matrix}) \otimes (\begin{matrix} 1 \\ 0 \end{matrix}), \rho(\begin{matrix} 1 \\ 0 \end{matrix}) \otimes (\begin{matrix} \beta \\ 1 \end{matrix}) + \rho(\begin{matrix} \beta \\ 1 \end{matrix}) \otimes (\begin{matrix} 1 \\ 0 \end{matrix}), (\begin{matrix} 1 \\ 0 \end{matrix}) \otimes (\begin{matrix} 0 \\ 1 \end{matrix}) - (\begin{matrix} 0 \\ 1 \end{matrix}) \otimes (\begin{matrix} 1 \\ 0 \end{matrix}), (\begin{matrix} \beta \\ 1 \end{matrix}) \otimes (\begin{matrix} \beta \\ 1 \end{matrix}) \right] .$$

Now we utilize the relation $\mathbf{vec}(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3) = (\mathbf{A}_3^{'} \otimes \mathbf{A}_1)\mathbf{vec}(\mathbf{A}_2)$ for any conformable matrices $\mathbf{A}_i$ $(i = 1, 2, 3)$. Then we notice that the product of $(1 + G_1)^2 \times (1 + G_1)^2$ matrices

$$(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\mathbf{A}$$

has special structures in the first $(1 + G_1)$ rows due to the statistical relations on the coefficients in the simultaneous equations such that

$$\begin{aligned}
&[\mathbf{B}^{'}\mathbf{e}_1 \otimes \mathbf{B}^{'}\mathbf{e}_1]_{1+G_1} = \left[\begin{array}{c} 1 \\ \mathbf{B}_2^{'}\mathbf{e}_1 \end{array} \right] , \\
&[(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\frac{\partial \mathbf{vec}}{\partial \beta_i}\left((\boldsymbol{\beta}, \mathbf{I}_{G_1})^{'}\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}, \mathbf{I}_{G_1})\right)]_{1+G_1} \\
&= \left[\left(\mathbf{B}^{'}\left(\begin{array}{c} \boldsymbol{\beta}^{'} \\ \mathbf{I}_{G_1} \end{array} \right) \otimes \mathbf{B}^{'}\left(\begin{array}{c} \frac{\partial \boldsymbol{\beta}^{'}}{\partial \beta_i} \\ \mathbf{O} \end{array} \right)\right)\mathbf{vec}(\Phi_{22.1}) + \left(\mathbf{B}^{'}\left(\begin{array}{c} \frac{\partial \boldsymbol{\beta}^{'}}{\partial \beta_i} \\ \mathbf{O} \end{array} \right) \otimes \mathbf{B}^{'}\left(\begin{array}{c} \boldsymbol{\beta}^{'} \\ \mathbf{I}_{G_1} \end{array} \right)\right)\mathbf{vec}(\Phi_{22.1})\right]_{1+G_1} \\
&= \left[\begin{array}{cc} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} \end{array} \right]\mathbf{e}_{i+1} \ , \\
&[(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\frac{\partial \mathbf{vec}}{\partial \Phi_{22.1ij}}\left((\boldsymbol{\beta}, \mathbf{I}_{G_1})^{'}\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}, \mathbf{I}_{G_1})\right)]_{1+G_1} = \mathbf{0} \ \ (i, j = 1, \cdots, G_1),
\end{aligned}$$

where we have used the notation $[\ \cdot\ ]_{1+G_1}$ as the first $1 + G_1$ elements of the corresponding vectors.

For the effects of $\mathbf{H}$, we differentiate (6.16) with respect to each elements of $\boldsymbol{\Omega} = (\omega_{lm})$ $(l, m = 1, \cdots, 1 + G_1)$ and evaluated at the probability limits as $q(n)$ (and n) $\rightarrow \infty$, we have the conditions

$$\sum_{i,j=1}^{1+G_1}\left[c_1\frac{\partial \phi_k}{\partial g_{ij}}\frac{\partial g_{ij}}{\partial \omega_{lm}} + \frac{\partial \phi_k}{\partial h_{ij}}\frac{\partial h_{ij}}{\partial \omega_{lm}}\right] = 0 \ . \tag{6.19}$$

They imply the conditions $c_1\frac{\partial \phi_k}{\partial g_{lm}} + \frac{\partial \phi_k}{\partial h_{lm}} = 0$ $(k = 1, \cdots, G_1; l, m = 1, \cdots, 1 + G_1)$ and hence the coefficients of the random matrix $\mathbf{H}$ on the asymptotic covariance bound have effects only through the effects of the coefficients of the random matrix $\mathbf{G}$.

Next we shall investigate the asymptotic distributions of the linearized version of estimators in the class given by (5.9). Because we only consider the asymptotic distrubitons, we replace $c_1\sqrt{a(n)} = \sqrt{c_1c_2}\sqrt{q(n)}$ without loss of generality. Then

$$\begin{aligned}
&\left[\begin{array}{c} (\boldsymbol{\phi}^{(1)})^{'} \\ \vdots \\ (\boldsymbol{\phi}^{(G_1)})^{'} \end{array} \right]\mathbf{vec}[\sqrt{a(n)}(\frac{1}{a(n)}\mathbf{G} - \mathbf{G}_0) - \sqrt{c_1c_2}\sqrt{q(n)}(\frac{1}{q(n)}\mathbf{H} - \boldsymbol{\Omega})] \\
= \ &\left[\left(\begin{array}{c} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{array} \right)(\mathbf{e}_1^{'} \otimes \mathbf{e}_1^{'}) + \mathbf{e}_1^{'} \otimes \left(\begin{array}{c} \mathbf{e}_2^{'} \\ \vdots \\ \mathbf{e}_{1+G_1}^{'} \end{array} \right)\right]\mathbf{A}^{-1}(\mathbf{B}^{'} \otimes \mathbf{B}^{'})^{-1} \\
&\times[\mathbf{vec}[\sqrt{a(n)}\boldsymbol{\Sigma}^{1/2}(\frac{1}{a(n)}\mathbf{G}^* - \mathbf{G}_0^*)\boldsymbol{\Sigma}^{1/2} - \sqrt{c_1c_2}\sqrt{q(n)}\boldsymbol{\Sigma}^{1/2}(\frac{1}{q(n)}\mathbf{H}^* - \mathbf{I}_{1+G_1})\boldsymbol{\Sigma}^{1/2}]
\end{aligned}$$

$$= \left\{ \begin{pmatrix} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{pmatrix} (\mathbf{e}_1^{'} \otimes \mathbf{e}_1^{'})[(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\mathbf{A}]^{-1} + \left[\mathbf{e}_1^{'} \otimes \begin{pmatrix} \mathbf{e}_2^{'} \\ \vdots \\ \mathbf{e}_{1+G_1}^{'} \end{pmatrix} \right] [(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\mathbf{A}]^{-1} \right\}$$

$$\times \mathbf{vec}[\sqrt{a(n)}\boldsymbol{\Sigma}^{1/2}(\frac{1}{a(n)}\mathbf{G}^* - \mathbf{G}_0^*)\boldsymbol{\Sigma}^{1/2} - \sqrt{c_1 c_2}\sqrt{q(n)}\boldsymbol{\Sigma}^{1/2}(\frac{1}{q(n)}\mathbf{H}^* - \mathbf{I}_{1+G_1})\boldsymbol{\Sigma}^{1/2}] \,.$$

By using the fact that the matrices $\mathbf{G}^*$, $\mathbf{G}_0^*$ and $\mathbf{H}^*$ are symmetric, we rearrange the row vectors of $(\mathbf{B}^{'} \otimes \mathbf{B}^{'})\mathbf{A}$ except the first $1 + G_1$ rows such that the $i(G_1 + 1) + 1$ $(i = 1, \cdots, G_1)$th row vectors become the $(G_1 + 1) + i$th rows, respectively, after the transformation $\mathbf{R}_1$. Then it can be written as

$$\mathbf{R}_1^{-1} \begin{bmatrix} 1 & 0 \cdots 0 & \mathbf{0}^{'} & \mathbf{0}^{'} \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & \mathbf{F} & \mathbf{O} \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & -\mathbf{F} & \mathbf{O} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} ,$$

where $\mathbf{F}$ is a $G_1 \times [\frac{1}{2}G_1(G_1+1)]$ constant matrix and the $(1+G_1)^2 \times (1+G_1)^2$ matrix $\mathbf{R}_1$ has 1 and 0 in its elements. By applying the fundamental transformations of matrices to the $(1 + 2G_1)$ row vector of the upper-left-corner matrix from the right, it can be expressed as

$$\begin{bmatrix} 1 & 0 \cdots 0 & \mathbf{0}^{'} & \mathbf{0}^{'} \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & \mathbf{F} & \mathbf{O} \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & -\mathbf{F} & \mathbf{O} \end{bmatrix} = \mathbf{R}_2^{-1} \begin{bmatrix} 1 & 0 \cdots 0 & \mathbf{0}^{'} & \mathbf{0}^{'} \\ 2\mathbf{B}_2^{'}\mathbf{e}_1 & 2\mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & \mathbf{O} & \mathbf{O} \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} & -\mathbf{F} & \mathbf{O} \end{bmatrix} ,$$

where the $(1 + 2G_1) \times (1 + 2G_1)$ matrix $\mathbf{R}_2$ has 1 and 0 in its elements. Then by using the representations of the random variables in (6.14) and (6.15), the linearized version of consistent estimators is asymptotically equivalent to

$$(6.20)\ \mathbf{e}^*(\boldsymbol{\beta}) = \begin{bmatrix} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{bmatrix} \sigma^2[\sqrt{c_1}x_{11}(n) - \sqrt{c_1 c_2}h_{11}(n)]$$

$$+(\mathbf{0}, \mathbf{I}_{G_1}) \begin{bmatrix} 1 & 0 \cdots 0 \\ \mathbf{B}_2^{'}\mathbf{e}_1 & \mathbf{E}^{'}\boldsymbol{\Phi}_{22.1} \end{bmatrix}^{-1} \boldsymbol{\Sigma}^{1/2}[\mathbf{g}_1(n) - \sqrt{c_1 c_2}\mathbf{h}_1(n)]$$

$$= \sigma^2[\sqrt{c_1}x_{11}(n) - \sqrt{c_1 c_2}h_{11}(n)] \left[\begin{pmatrix} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{pmatrix} - [\mathbf{E}^{'}\boldsymbol{\Phi}_{22.1}]^{-1}\mathbf{B}_2^{'}\mathbf{e}_1 \right]$$

$$+[\mathbf{E}^{'}\boldsymbol{\Phi}_{22.1}]^{-1}[\mathbf{E}^{'}\boldsymbol{\Phi}_{22.1}^{1/2}\mathbf{w}_{21}(n) + \sqrt{c_1}\mathbf{x}_{21}(n) - \sqrt{c_1 c_2}\mathbf{h}_{21}(n)]\sigma \,.$$

We notice that each elements of $x_{11}(n), h_{11}(n), \mathbf{w}_{21}(n), \mathbf{x}_{21}(n)$, and $\mathbf{h}_{21}(n)$ are asymptotically independent normal variables. Hence we have the covariance matrix

$$(6.21)\quad \mathbf{E}[\mathbf{e}^*(\boldsymbol{\beta})\mathbf{e}^*(\boldsymbol{\beta})^{'}]$$

$$= 2\sigma^4 c_1(1 + c_2)\mathbf{a}\mathbf{a}^{'} + \sigma^2[\boldsymbol{\Phi}_{22.1}^{-1} + c_1(1 + c_2)\boldsymbol{\Phi}_{22.1}^{-1}\mathbf{E}^{'-1}\mathbf{E}^{-1}\boldsymbol{\Phi}_{22.1}^{-1}],$$

where

$$\mathbf{a} = \begin{pmatrix} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{pmatrix} - [\mathbf{E}^{'}\mathbf{\Phi}_{22.1}]^{-1}\mathbf{B}_2^{'}\mathbf{e}_1$$

Then by minimizing (6.21) with respect to $\phi_{11}^{(k)}$ $(k = 1, \cdots, G_1)$, that is by $\mathbf{a}$, we establish the next proposition, which gives the lower bound of the asymptotic covariance matrix for the class of consistent estimators we have investigated.

Lemma A.1 : For (2.1) and (2.3), suppose all conditions in *Theorem 5.1* hold. Define the class of consistent estimators in the form of (5.9). Then the covariance matrix (6.21) is minimized in the sense of positive definiteness when

$$\begin{pmatrix} \phi_{11}^{(1)} \\ \vdots \\ \phi_{11}^{(G_1)} \end{pmatrix} = [\mathbf{E}^{'}\mathbf{\Phi}_{22.1}]^{-1}\mathbf{B}_2^{'}\mathbf{e}_1 \quad , \tag{6.22}$$

where $\mathbf{E}^{'} = \mathbf{B}_2^{'}(\beta, \mathbf{I}_{G_1})^{'}$. The resulting lower bound of the asymptotic covariance matrix for the normalized consistent estimator $\hat{e}_\beta = [\mathbf{\Pi}_{22}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}]^{1/2}(\hat{\beta} - \beta)$ can be expressed as

$$\mathbf{\Psi}^{*} = \sigma^2\mathbf{I}_{G_1} + c_1(1+c_2)\sigma^2\mathbf{\Phi}_{22.1}^{-1/2}\mathbf{E}^{'-1}\mathbf{E}^{-1}\mathbf{\Phi}_{22.1}^{-1/2} \ . \tag{6.23}$$

Finally, we consider a $(1+G_1)\times(1+G_1)$ matrix $\mathbf{C} = (\mathbf{c}_1, \mathbf{C}_2)$ such that $\mathbf{c}_1 = \mathbf{\Omega}^{1/2}\mathbf{b}_0$ and $\mathbf{C}_2 = \mathbf{\Omega}^{1/2}\mathbf{B}_2$. Because the matrix $\mathbf{C}$ is a block diagonal matrix, we have the relations $\mathbf{c}_1^{'}\mathbf{C}_2 = \mathbf{0}^{'}$ and

$$\mathbf{I}_{1+G_1} = \mathbf{C}(\mathbf{C}^{'}\mathbf{C})^{-1}\mathbf{C}^{-1} = (\mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{b}_0)^{-1}\mathbf{\Omega}^{1/2}\mathbf{b}_0\mathbf{b}_0^{'}\mathbf{\Omega}^{1/2} + \mathbf{\Omega}^{1/2}\mathbf{B}_2[\mathbf{B}_2^{'}\mathbf{\Omega}\mathbf{B}_2]^{-1}\mathbf{B}_2^{'}\mathbf{\Omega}^{1/2} \ .$$

Then we use the simple relation that $\mathbf{J}_2^{'}\mathbf{\Omega}\mathbf{B}\mathbf{J}_2 = \mathbf{J}_2^{'}\mathbf{B}^{'-1}\mathbf{B}^{'}\Omega\mathbf{B}\mathbf{J}_2 = \mathbf{J}_2^{'}\mathbf{B}^{'-1}\mathbf{J}_2[\mathbf{B}_2^{'}\mathbf{\Omega}\mathbf{J}_2]$, which is eventually the same as $\mathbf{E}^{-1}\mathbf{B}_2^{'}\mathbf{\Omega}\mathbf{J}_2$. Then we find

$$\begin{aligned} & \mathbf{J}_2^{'}\mathbf{\Omega}\mathbf{J}_2 - \sigma^{-2}\mathbf{J}_2^{'}\mathbf{\Omega}\mathbf{b}_0\mathbf{b}_0^{'}\mathbf{\Omega}\mathbf{J}_2 \\ = \ & \mathbf{J}_2^{'}\mathbf{\Omega}\mathbf{B}_2[\mathbf{B}_2^{'}\mathbf{\Omega}\mathbf{B}_2]^{-1}\mathbf{B}_2^{'}\mathbf{\Omega}\mathbf{J}_2 = \mathbf{E}^{'-1}\mathbf{E}^{-1} \ , \end{aligned} \tag{6.24}$$

and we can confirm that two expressions of the asymptotic covariance matrix in (5.3) and (6.23) are the same. Hence we have completed the proof of *Theorem 5.2*.
(**Q.E.D.**)

APPENDIX B : TABLES AND FIGURES

Notes on Tables

In Tables the distributions are tabulated in the standardized terms, that is, of (3.2). 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 (3.2) 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 (3.2). (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 line 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.

		$n-K=30, K_2=3, \alpha=1$							
		$\delta^2=10$				$\delta^2=30$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.003	0.006	0.001	0.001	0.001	0.000	0.000	0.000
-2.5	0.006	0.005	0.008	0.002	0.003	0.002	0.002	0.000	0.002
-2	0.023	0.009	0.016	0.010	0.014	0.007	0.012	0.011	0.016
-1.4	0.081	0.045	0.062	0.079	0.087	0.053	0.066	0.078	0.091
-1	0.159	0.127	0.146	0.205	0.219	0.133	0.145	0.187	0.195
-0.8	0.212	0.188	0.209	0.295	0.305	0.193	0.204	0.262	0.270
-0.6	0.274	0.264	0.281	0.394	0.394	0.263	0.274	0.346	0.356
-0.4	0.345	0.347	0.357	0.496	0.492	0.339	0.353	0.434	0.443
-0.2	0.421	0.423	0.437	0.591	0.585	0.420	0.429	0.519	0.521
0	0.500	0.499	0.511	0.671	0.663	0.500	0.503	0.601	0.597
0.2	0.579	0.569	0.576	0.738	0.731	0.576	0.574	0.675	0.669
0.4	0.655	0.631	0.635	0.792	0.786	0.646	0.640	0.740	0.732
0.6	0.726	0.686	0.687	0.839	0.832	0.706	0.699	0.793	0.784
0.8	0.788	0.735	0.733	0.876	0.869	0.760	0.751	0.838	0.829
1	0.841	0.775	0.772	0.901	0.896	0.805	0.794	0.877	0.868
1.4	0.919	0.835	0.833	0.939	0.936	0.874	0.866	0.930	0.927
2	0.977	0.891	0.888	0.972	0.968	0.935	0.931	0.971	0.968
2.5	0.994	0.922	0.917	0.985	0.982	0.962	0.956	0.986	0.984
3	0.999	0.941	0.938	0.991	0.989	0.978	0.974	0.994	0.994
X05	-1.65	-1.37	-1.49	-1.55	-1.59	-1.40	-1.52	-1.55	-1.64
L.QT	-0.67	-0.63	-0.68	-0.90	-0.93	-0.64	-0.66	-0.83	-0.85
MEDN	0	0.00	-0.03	-0.40	-0.38	0.00	-0.01	-0.24	-0.26
U.QT	0.67	0.87	0.88	0.21	0.27	0.76	0.80	0.44	0.47
X95	1.65	3.21	3.45	1.53	1.62	2.14	2.37	1.64	1.66
IQR	1.35	1.50	1.56	1.11	1.19	1.40	1.46	1.27	1.31

Table 1:

		$n-K=30, K_2=3, \alpha=1$							
		$\delta^2=50$				$\delta^2=100$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2.5	0.006	0.001	0.001	0.001	0.001	0.002	0.003	0.003	0.003
-2	0.023	0.008	0.012	0.013	0.016	0.011	0.016	0.015	0.019
-1.4	0.081	0.057	0.065	0.079	0.087	0.062	0.069	0.079	0.086
-1	0.159	0.138	0.146	0.181	0.188	0.143	0.150	0.173	0.183
-0.8	0.212	0.197	0.205	0.247	0.257	0.200	0.208	0.237	0.245
-0.6	0.274	0.264	0.271	0.326	0.332	0.267	0.272	0.309	0.313
-0.4	0.345	0.340	0.343	0.413	0.413	0.341	0.342	0.388	0.390
-0.2	0.421	0.420	0.417	0.502	0.491	0.420	0.419	0.470	0.474
0	0.500	0.500	0.490	0.584	0.569	0.500	0.498	0.552	0.556
0.2	0.579	0.577	0.567	0.657	0.648	0.578	0.573	0.629	0.629
0.4	0.655	0.648	0.638	0.723	0.716	0.651	0.644	0.698	0.696
0.6	0.726	0.712	0.698	0.782	0.772	0.716	0.707	0.758	0.758
0.8	0.788	0.767	0.751	0.833	0.818	0.774	0.763	0.811	0.812
1	0.841	0.814	0.799	0.871	0.859	0.823	0.812	0.855	0.853
1.4	0.919	0.885	0.875	0.925	0.920	0.897	0.884	0.918	0.915
2	0.977	0.947	0.938	0.970	0.965	0.958	0.951	0.971	0.967
2.5	0.994	0.973	0.968	0.985	0.983	0.982	0.977	0.988	0.986
3	0.999	0.986	0.984	0.993	0.993	0.992	0.990	0.995	0.995
X05	-1.65	-1.46	-1.50	-1.58	-1.62	-1.47	-1.54	-1.59	-1.63
L.QT	-0.67	-0.64	-0.66	-0.80	-0.82	-0.65	-0.67	-0.77	-0.79
MEDN	0	0.00	0.03	-0.20	-0.18	0.00	0.01	-0.14	-0.14
U.QT	0.67	0.73	0.80	0.49	0.52	0.71	0.75	0.55	0.57
X95	1.65	2.01	2.17	1.67	1.75	1.90	1.98	1.71	1.74
IQR	1.35	1.37	1.46	1.29	1.34	1.36	1.42	1.32	1.36

Table 2:

					$n-K=100, K_2=10, \alpha=1$				
		$\delta^2=30$				$\delta^2=50$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.001
-2.5	0.006	0.001	0.005	0.008	0.013	0.002	0.002	0.009	0.012
-2	0.023	0.011	0.020	0.063	0.072	0.012	0.018	0.047	0.066
-1.4	0.081	0.064	0.089	0.250	0.260	0.065	0.090	0.210	0.226
-1	0.159	0.149	0.178	0.457	0.458	0.147	0.174	0.392	0.394
-0.8	0.212	0.207	0.233	0.564	0.559	0.205	0.231	0.492	0.488
-0.6	0.274	0.275	0.296	0.662	0.651	0.272	0.294	0.592	0.578
-0.4	0.345	0.349	0.362	0.742	0.731	0.345	0.360	0.679	0.661
-0.2	0.421	0.425	0.433	0.811	0.798	0.423	0.428	0.752	0.733
0	0.500	0.500	0.507	0.865	0.854	0.500	0.497	0.815	0.794
0.2	0.579	0.572	0.574	0.905	0.895	0.574	0.565	0.867	0.845
0.4	0.655	0.637	0.635	0.935	0.925	0.644	0.631	0.904	0.888
0.6	0.726	0.696	0.688	0.957	0.947	0.705	0.691	0.933	0.920
0.8	0.788	0.747	0.733	0.970	0.962	0.759	0.742	0.953	0.942
1	0.841	0.791	0.777	0.979	0.973	0.806	0.786	0.968	0.959
1.4	0.919	0.859	0.845	0.991	0.985	0.876	0.856	0.986	0.980
2	0.977	0.923	0.906	0.998	0.995	0.939	0.925	0.997	0.994
2.5	0.994	0.952	0.938	1.000	0.998	0.967	0.959	0.999	0.998
3	0.999	0.971	0.959	1.000	1.000	0.982	0.977	1.000	0.999
X05	-1.65	-1.51	-1.66	-2.07	-2.11	-1.49	-1.68	-1.98	-2.09
L.QT	-0.67	-0.67	-0.75	-1.40	-1.42	-0.66	-0.74	-1.31	-1.33
MEDN	0	0.00	-0.02	-0.92	-0.92	0.00	0.01	-0.77	-0.77
U.QT	0.67	0.81	0.88	-0.38	-0.35	0.76	0.83	-0.18	-0.15
X95	1.65	2.36	2.76	0.52	0.63	2.11	2.35	0.76	0.89
IQR	1.35	1.48	1.62	1.02	1.08	1.42	1.57	1.12	1.19

Table 3:

		$n-K=100, K_2=10, \alpha=1$							
		$\delta^2=100$				$\delta^2=300$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.000	0.001	0.002	0.000	0.000	0.001	0.001
-2.5	0.006	0.003	0.004	0.009	0.014	0.003	0.005	0.008	0.011
-2	0.023	0.013	0.020	0.046	0.055	0.016	0.022	0.036	0.043
-1.4	0.081	0.066	0.080	0.178	0.188	0.071	0.088	0.132	0.148
-1	0.159	0.148	0.166	0.328	0.337	0.152	0.171	0.251	0.260
-0.8	0.212	0.205	0.224	0.414	0.419	0.207	0.228	0.323	0.330
-0.6	0.274	0.271	0.290	0.501	0.505	0.272	0.290	0.402	0.403
-0.4	0.345	0.345	0.359	0.586	0.586	0.343	0.354	0.486	0.478
-0.2	0.421	0.423	0.429	0.667	0.660	0.421	0.425	0.571	0.555
0	0.500	0.500	0.504	0.739	0.726	0.500	0.500	0.645	0.628
0.2	0.579	0.575	0.577	0.804	0.785	0.578	0.572	0.712	0.694
0.4	0.655	0.647	0.642	0.855	0.832	0.652	0.635	0.776	0.753
0.6	0.726	0.712	0.698	0.893	0.869	0.718	0.693	0.831	0.806
0.8	0.788	0.769	0.748	0.922	0.901	0.778	0.748	0.873	0.853
1	0.841	0.818	0.793	0.944	0.927	0.829	0.798	0.908	0.892
1.4	0.919	0.893	0.869	0.974	0.968	0.905	0.880	0.956	0.944
2	0.977	0.955	0.941	0.992	0.991	0.966	0.952	0.985	0.979
2.5	0.994	0.979	0.970	0.998	0.998	0.988	0.975	0.995	0.990
3	0.999	0.990	0.985	0.999	1.000	0.996	0.989	0.999	0.997
X05	-1.65	-1.54	-1.61	-1.97	-2.04	-1.57	-1.65	-1.86	-1.94
L.QT	-0.67	-0.66	-0.72	-1.17	-1.22	-0.66	-0.73	-0.99	-1.03
MEDN	0	0.00	-0.01	-0.59	-0.61	0.00	0.00	-0.36	-0.34
U.QT	0.67	0.73	0.81	0.05	0.08	0.70	0.81	0.31	0.39
X95	1.65	1.90	2.11	1.06	1.18	1.78	1.97	1.32	1.47
IQR	1.35	1.39	1.53	1.22	1.30	1.36	1.54	1.30	1.42

Table 4:

		$n-K=300, K_2=30, \alpha=1$							
		$\delta^2=50$				$\delta^2=100$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.002	0.033	0.042	0.000	0.001	0.022	0.035
-2.5	0.006	0.003	0.010	0.154	0.165	0.002	0.008	0.105	0.124
-2	0.023	0.019	0.034	0.400	0.417	0.016	0.030	0.302	0.309
-1.4	0.081	0.085	0.110	0.750	0.736	0.077	0.100	0.600	0.593
-1	0.159	0.170	0.195	0.893	0.870	0.160	0.182	0.773	0.758
-0.8	0.212	0.227	0.248	0.933	0.918	0.217	0.235	0.839	0.821
-0.6	0.274	0.231	0.308	0.959	0.950	0.280	0.299	0.889	0.873
-0.4	0.345	0.359	0.369	0.977	0.971	0.351	0.364	0.924	0.913
-0.2	0.421	0.430	0.432	0.988	0.983	0.425	0.428	0.951	0.941
0	0.500	0.500	0.494	0.994	0.991	0.500	0.495	0.968	0.962
0.2	0.579	0.567	0.556	0.997	0.996	0.573	0.561	0.980	0.977
0.4	0.655	0.630	0.616	0.999	0.998	0.641	0.626	0.988	0.986
0.6	0.726	0.687	0.670	1.000	1.000	0.703	0.684	0.993	0.992
0.8	0.788	0.739	0.718	1.000	1.000	0.758	0.736	0.995	0.996
1	0.841	0.783	0.756	1.000	1.000	0.806	0.779	0.997	0.998
1.4	0.919	0.852	0.822	1.000	1.000	0.879	0.853	1.000	1.000
2	0.977	0.920	0.895	1.000	1.000	0.946	0.922	1.000	1.000
2.5	0.994	0.953	0.931	1.000	1.000	0.974	0.957	1.000	1.000
3	0.999	0.972	0.953	1.000	1.000	0.988	0.977	1.000	1.000
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 5:

		$n-K=100, K_2=10, \alpha=1, u_i=(\chi^2(3)-3)/\sqrt{6}$							
		$\delta^2=30$				$\delta^2=50$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.003	0.000	0.001	0.000	0.001	0.000	0.001
-2.5	0.006	0.001	0.007	0.011	0.012	0.002	0.004	0.012	0.010
-2	0.023	0.009	0.017	0.057	0.053	0.013	0.015	0.059	0.044
-1.4	0.081	0.058	0.067	0.243	0.211	0.067	0.069	0.218	0.188
-1	0.159	0.139	0.150	0.431	0.394	0.151	0.150	0.390	0.356
-0.8	0.212	0.196	0.201	0.535	0.492	0.208	0.206	0.484	0.448
-0.6	0.274	0.261	0.263	0.636	0.593	0.273	0.275	0.583	0.545
-0.4	0.345	0.333	0.334	0.724	0.689	0.343	0.348	0.669	0.634
-0.2	0.421	0.409	0.409	0.796	0.772	0.423	0.424	0.742	0.712
0	0.500	0.487	0.491	0.852	0.838	0.504	0.505	0.804	0.780
0.2	0.579	0.561	0.568	0.895	0.885	0.579	0.582	0.858	0.835
0.4	0.655	0.631	0.637	0.927	0.919	0.648	0.649	0.899	0.883
0.6	0.726	0.694	0.699	0.951	0.946	0.710	0.708	0.929	0.920
0.8	0.788	0.744	0.753	0.968	0.964	0.764	0.760	0.952	0.948
1	0.841	0.784	0.795	0.979	0.977	0.808	0.804	0.969	0.965
1.4	0.919	0.855	0.860	0.992	0.991	0.878	0.875	0.988	0.987
2	0.977	0.921	0.922	0.998	0.999	0.939	0.936	0.997	0.997
2.5	0.994	0.951	0.953	0.999	1.000	0.968	0.962	0.999	0.999
3	0.999	0.970	0.967	1.000	1.000	0.982	0.979	1.000	1.000
X05	-1.65	-1.46	-1.51	-2.04	-2.02	-1.52	-1.53	-2.06	-1.96
L.QT	-0.67	-0.63	-0.64	-1.38	-1.31	-0.67	-0.67	-1.32	-1.24
MEDN	0	0.04	0.02	-0.87	-0.78	-0.01	-0.01	-0.77	-0.69
U.QT	0.67	0.83	0.79	-0.33	-0.26	0.75	0.76	-0.17	-0.09
X95	1.65	2.49	2.43	0.59	0.64	2.17	2.24	0.78	0.82
IQR	1.35	1.46	1.43	1.05	1.05	1.42	1.43	1.14	1.15

Table 6:

	$n-K=100, K_2=10, \alpha=1, u_i=t(5)$								
			$\delta^2=30$				$\delta^2=50$		
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.002	0.002	0.001	0.000	0.001	0.000	0.002
-2.5	0.006	0.001	0.006	0.013	0.012	0.001	0.005	0.011	0.010
-2	0.023	0.011	0.022	0.062	0.054	0.012	0.016	0.053	0.047
-1.4	0.081	0.060	0.082	0.246	0.218	0.064	0.070	0.204	0.183
-1	0.159	0.143	0.162	0.442	0.408	0.139	0.156	0.374	0.346
-0.8	0.212	0.200	0.222	0.550	0.513	0.192	0.210	0.475	0.444
-0.6	0.274	0.270	0.290	0.650	0.613	0.257	0.273	0.573	0.545
-0.4	0.345	0.348	0.362	0.738	0.702	0.333	0.344	0.662	0.635
-0.2	0.421	0.428	0.437	0.810	0.777	0.413	0.419	0.742	0.718
0	0.500	0.504	0.508	0.864	0.841	0.492	0.495	0.807	0.789
0.2	0.579	0.577	0.572	0.904	0.890	0.563	0.565	0.861	0.842
0.4	0.655	0.643	0.632	0.933	0.925	0.630	0.631	0.900	0.885
0.6	0.726	0.703	0.690	0.955	0.949	0.698	0.690	0.929	0.918
0.8	0.788	0.753	0.739	0.971	0.967	0.757	0.742	0.952	0.944
1	0.841	0.794	0.781	0.982	0.978	0.805	0.787	0.967	0.963
1.4	0.919	0.856	0.847	0.993	0.990	0.875	0.860	0.985	0.985
2	0.977	0.921	0.909	0.998	0.997	0.942	0.927	0.997	0.997
2.5	0.994	0.953	0.940	0.999	0.999	0.968	0.960	1.000	0.999
3	0.999	0.971	0.958	1.000	1.000	0.983	0.977	1.000	1.000
X05	-1.65	-1.47	-1.65	-2.07	-2.03	-1.51	-1.55	-2.02	-1.97
L.QT	-0.67	-0.65	-0.72	-1.39	-1.32	-0.62	-0.67	-1.28	-1.22
MEDN	0	-0.01	-0.02	-0.89	-0.82	0.02	0.01	-0.75	-0.69
U.QT	0.67	0.79	0.85	-0.37	-0.28	0.77	0.83	-0.18	-0.12
X95	1.65	2.44	2.76	0.54	0.62	2.12	2.33	0.78	0.86
IQR	1.35	1.44	1.56	1.02	1.05	1.39	1.50	1.10	1.10

Table 7:

	$n-K=30, K_2=3, \alpha=1, u_i=\|Z_i\|\epsilon_i$								
			$\delta^2=30$				$\delta^2=100$		
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2.5	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001
-2	0.023	0.003	0.007	0.007	0.009	0.007	0.012	0.010	0.015
-1.4	0.081	0.042	0.052	0.066	0.071	0.049	0.067	0.066	0.080
-1	0.159	0.121	0.130	0.176	0.173	0.121	0.149	0.154	0.176
-0.8	0.212	0.180	0.186	0.255	0.249	0.179	0.204	0.223	0.242
-0.6	0.274	0.250	0.251	0.342	0.335	0.249	0.270	0.303	0.318
-0.4	0.345	0.328	0.330	0.431	0.426	0.329	0.344	0.387	0.401
-0.2	0.421	0.412	0.412	0.524	0.519	0.410	0.426	0.473	0.488
0	0.500	0.495	0.494	0.612	0.609	0.493	0.506	0.560	0.569
0.2	0.579	0.574	0.571	0.687	0.685	0.576	0.584	0.644	0.646
0.4	0.655	0.644	0.641	0.749	0.750	0.653	0.658	0.714	0.716
0.6	0.726	0.704	0.703	0.807	0.806	0.720	0.720	0.774	0.778
0.8	0.788	0.761	0.756	0.855	0.851	0.779	0.776	0.827	0.833
1	0.841	0.806	0.797	0.890	0.883	0.828	0.825	0.870	0.875
1.4	0.919	0.870	0.860	0.932	0.930	0.899	0.892	0.930	0.926
2	0.977	0.928	0.924	0.969	0.971	0.955	0.947	0.973	0.968
2.5	0.994	0.955	0.956	0.984	0.985	0.977	0.974	0.987	0.987
3	0.999	0.973	0.975	0.991	0.992	0.988	0.988	0.994	0.996
X05	-1.65	-1.35	-1.42	-1.49	-1.53	-1.39	-1.51	-1.52	-1.57
L.QT	-0.67	-0.60	-0.60	-0.81	-0.80	-0.60	-0.66	-0.73	-0.78
MEDN	0	0.01	0.02	-0.25	-0.24	0.02	-0.02	-0.14	-0.17
U.QT	0.67	0.76	0.78	0.40	0.40	0.70	0.71	0.52	0.51
X95	1.65	2.40	2.40	1.65	1.62	1.93	2.05	1.62	1.70
IQR	1.35	1.36	1.38	1.21	1.20	1.29	1.36	1.25	1.29

Table 8:

	$n-K=100, K_2=10, \alpha=1, u_i=\|Z_i\|\epsilon_i$								
			$\delta^2=50$				$\delta^2=100$		
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.002
-2.5	0.006	0.000	0.003	0.010	0.011	0.002	0.003	0.009	0.013
-2	0.023	0.011	0.017	0.056	0.057	0.014	0.020	0.045	0.054
-1.4	0.081	0.068	0.080	0.227	0.224	0.066	0.080	0.177	0.190
-1	0.159	0.153	0.167	0.410	0.406	0.147	0.165	0.340	0.335
-0.8	0.212	0.213	0.227	0.509	0.500	0.205	0.221	0.433	0.417
-0.6	0.274	0.279	0.292	0.606	0.590	0.276	0.282	0.523	0.502
-0.4	0.345	0.350	0.362	0.693	0.679	0.355	0.350	0.608	0.588
-0.2	0.421	0.430	0.439	0.767	0.755	0.436	0.422	0.689	0.663
0	0.500	0.508	0.511	0.827	0.816	0.517	0.491	0.759	0.729
0.2	0.579	0.581	0.579	0.873	0.866	0.592	0.563	0.816	0.787
0.4	0.655	0.646	0.642	0.909	0.904	0.660	0.629	0.862	0.840
0.6	0.726	0.708	0.703	0.938	0.933	0.722	0.689	0.897	0.885
0.8	0.788	0.761	0.754	0.959	0.953	0.776	0.743	0.925	0.918
1	0.841	0.805	0.794	0.973	0.967	0.820	0.792	0.947	0.942
1.4	0.919	0.870	0.860	0.988	0.984	0.889	0.869	0.975	0.970
2	0.977	0.935	0.923	0.998	0.997	0.952	0.935	0.994	0.991
2.5	0.994	0.964	0.953	1.000	0.999	0.978	0.966	0.998	0.997
3	0.999	0.981	0.971	1.000	1.000	0.991	0.984	1.000	1.000
X05	-1.65	-1.52	-1.60	-2.05	-2.05	-1.52	-1.64	-1.96	-2.03
L.QT	-0.67	-0.69	-0.73	-1.34	-1.34	-0.67	-0.70	-1.20	-1.22
MEDN	0	-0.02	-0.03	-0.82	-0.80	-0.04	0.03	-0.65	-0.60
U.QT	0.67	0.76	0.78	-0.25	-0.21	0.70	0.83	-0.03	0.07
X95	1.65	2.23	2.43	0.71	0.76	1.97	2.20	1.03	1.09
IQR	1.35	1.44	1.51	1.10	1.12	1.37	1.53	1.18	1.29

Table 9:

	$n-K=300, K_2=30, \alpha=1, u_i=\|Z_i\|\epsilon_i$								
	$\delta^2=50$					$\delta^2=100$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.002	0.035	0.045	0.000	0.000	0.022	0.030
-2.5	0.006	0.005	0.010	0.159	0.169	0.002	0.005	0.111	0.118
-2	0.023	0.021	0.031	0.424	0.419	0.016	0.024	0.314	0.312
-1.4	0.081	0.082	0.100	0.763	0.743	0.075	0.091	0.614	0.599
-1	0.159	0.171	0.188	0.901	0.888	0.161	0.178	0.788	0.767
-0.8	0.212	0.226	0.241	0.943	0.931	0.217	0.231	0.853	0.830
-0.6	0.274	0.286	0.299	0.969	0.959	0.284	0.295	0.899	0.880
-0.4	0.345	0.351	0.364	0.983	0.976	0.355	0.361	0.931	0.918
-0.2	0.421	0.422	0.428	0.991	0.986	0.427	0.427	0.954	0.944
0	0.500	0.493	0.491	0.995	0.993	0.499	0.495	0.970	0.962
0.2	0.579	0.560	0.551	0.998	0.996	0.568	0.563	0.981	0.975
0.4	0.655	0.622	0.605	0.999	0.999	0.635	0.626	0.988	0.984
0.6	0.726	0.678	0.660	1.000	1.000	0.696	0.683	0.993	0.991
0.8	0.788	0.729	0.710	1.000	1.000	0.752	0.732	0.996	0.995
1	0.841	0.775	0.757	1.000	1.000	0.801	0.777	0.998	0.998
1.4	0.919	0.845	0.826	1.000	1.000	0.880	0.849	1.000	1.000
2	0.977	0.912	0.893	1.000	1.000	0.941	0.923	1.000	1.000
2.5	0.994	0.947	0.931	1.000	1.000	0.972	0.958	1.000	1.000
3	0.999	0.970	0.952	1.000	1.000	0.987	0.976	1.000	1.000
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 10:

		$n-K=1000, K_2=100, \alpha=1, \delta^2=100$					
		$u_i=N(0,1)$			$u_i=\|Z_i\|\epsilon_i$		
x	normal	LIML	TSLS	GMM	LIML	TSLS	GMM
-3	0.001	0.002	0.817	0.803	0.002	0.834	0.808
-2.5	0.006	0.009	0.961	0.947	0.010	0.965	0.952
-2	0.023	0.032	0.994	0.993	0.035	0.998	0.995
-1.4	0.081	0.111	1.000	1.000	0.111	1.000	1.000
-1	0.159	0.191	1.000	1.000	0.200	1.000	1.000
-0.8	0.212	0.243	1.000	1.000	0.253	1.000	1.000
-0.6	0.274	0.306	1.000	1.000	0.308	1.000	1.000
-0.4	0.345	0.370	1.000	1.005	0.368	1.000	1.000
-0.2	0.421	0.434	1.000	1.000	0.433	1.000	1.000
0	0.500	0.499	1.000	1.000	0.497	1.000	1.000
0.2	0.579	0.564	1.000	1.000	0.557	1.000	1.000
0.4	0.655	0.623	1.000	1.000	0.617	1.000	1.000
0.6	0.726	0.677	1.000	1.000	0.673	1.000	1.000
0.8	0.788	0.728	1.000	1.000	0.722	1.000	1.000
1	0.841	0.774	1.000	1.000	0.765	1.000	1.000
1.4	0.919	0.851	1.000	1.000	0.839	1.000	1.000
2	0.977	0.920	1.000	1.000	0.912	1.000	1.000
2.5	0.994	0.957	1.000	1.000	0.949	1.000	1.000
3	0.999	0.979	1.000	1.000	0.972	1.000	1.000
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 11:

				$n-K=300, K_2=30, \delta^2=100$					
			$\alpha=0$				$\alpha=5$		
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.006	0.014	0.000	0.001	0.000	0.000	0.088	0.108
-2.5	0.006	0.017	0.030	0.003	0.005	0.000	0.000	0.319	0.333
-2	0.023	0.044	0.064	0.012	0.016	0.007	0.011	0.636	0.622
-1.4	0.081	0.115	0.146	0.055	0.066	0.055	0.067	0.883	0.863
-1	0.159	0.193	0.226	0.130	0.146	0.137	0.155	0.953	0.943
-0.8	0.212	0.244	0.277	0.187	0.203	0.196	0.213	0.971	0.966
-0.6	0.274	0.300	0.333	0.252	0.271	0.265	0.280	0.983	0.979
-0.4	0.345	0.364	0.391	0.328	0.346	0.340	0.352	0.991	0.986
-0.2	0.421	0.430	0.449	0.415	0.425	0.420	0.430	0.994	0.992
0	0.500	0.500	0.507	0.502	0.506	0.500	0.506	0.996	0.995
0.2	0.579	0.570	0.567	0.585	0.587	0.577	0.578	0.997	0.998
0.4	0.655	0.636	0.627	0.666	0.666	0.648	0.643	0.999	0.999
0.6	0.726	0.700	0.684	0.744	0.737	0.712	0.704	0.999	1.000
0.8	0.788	0.756	0.736	0.814	0.800	0.768	0.761	1.000	1.000
1	0.841	0.807	0.784	0.870	0.855	0.815	0.806	1.000	1.000
1.4	0.919	0.885	0.862	0.943	0.931	0.886	0.876	1.000	1.000
2	0.977	0.956	0.937	0.987	0.983	0.950	0.940	1.000	1.000
2.5	0.994	0.983	0.970	0.998	0.997	0.976	0.970	1.000	1.000
3	0.999	0.994	0.987	1.000	1.000	0.989	0.983	1.000	1.000
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 12:

		$n-K=30, K_2=3, \alpha=5$							
		$\delta^2=30$				$\delta^2=100$			
x	normal	LIML	MEL	TSLS	GMM	LIML	MEL	TSLS	GMM
-3	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2.5	0.006	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002
-2	0.023	0.001	0.002	0.004	0.009	0.006	0.009	0.010	0.013
-1.4	0.081	0.031	0.042	0.067	0.084	0.049	0.058	0.073	0.080
-1	0.159	0.117	0.132	0.207	0.220	0.134	0.142	0.177	0.186
-0.8	0.212	0.182	0.197	0.294	0.304	0.189	0.200	0.244	0.253
-0.6	0.274	0.257	0.270	0.389	0.394	0.253	0.263	0.322	0.330
-0.4	0.345	0.341	0.352	0.479	0.487	0.330	0.339	0.405	0.408
-0.2	0.421	0.424	0.428	0.561	0.571	0.409	0.412	0.489	0.490
0	0.500	0.502	0.507	0.644	0.644	0.492	0.492	0.573	0.573
0.2	0.579	0.580	0.583	0.720	0.712	0.576	0.572	0.656	0.651
0.4	0.655	0.655	0.650	0.778	0.767	0.650	0.644	0.720	0.715
0.6	0.726	0.713	0.706	0.824	0.815	0.715	0.704	0.778	0.771
0.8	0.788	0.768	0.756	0.862	0.856	0.773	0.761	0.829	0.820
1	0.841	0.810	0.794	0.893	0.883	0.821	0.807	0.867	0.858
1.4	0.919	0.869	0.859	0.930	0.924	0.891	0.880	0.924	0.917
2	0.977	0.930	0.917	0.965	0.963	0.953	0.944	0.969	0.965
2.5	0.994	0.958	0.948	0.983	0.980	0.981	0.976	0.989	0.986
3	0.999	0.975	0.966	0.990	0.989	0.993	0.989	0.995	0.994
X05	-1.96	-1.27	-1.34	-1.48	-1.55	-1.39	-1.45	-1.54	-1.58
L.QT	-0.67	-0.62	-0.65	-0.89	-0.93	-0.61	-0.64	-0.78	-0.81
MEDN	0	-0.01	-0.02	-0.35	-0.37	0.02	0.02	-0.17	-0.18
U.QT	0.67	0.73	0.78	0.30	0.34	0.72	0.76	0.50	0.53
X95	1.96	2.35	2.54	1.73	1.76	1.96	2.09	1.69	1.78
IQR	1.35	1.34	1.43	1.19	1.26	1.33	1.40	1.29	1.33

Table 13:

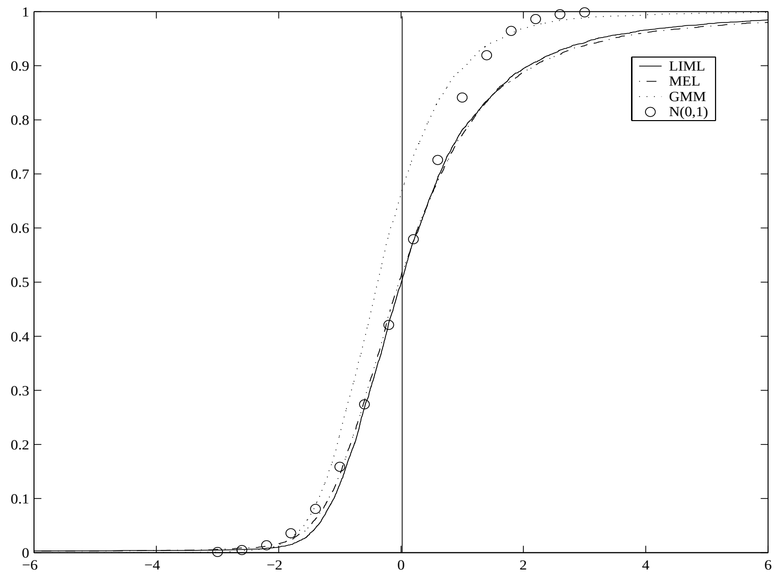

Figure 1: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 10$

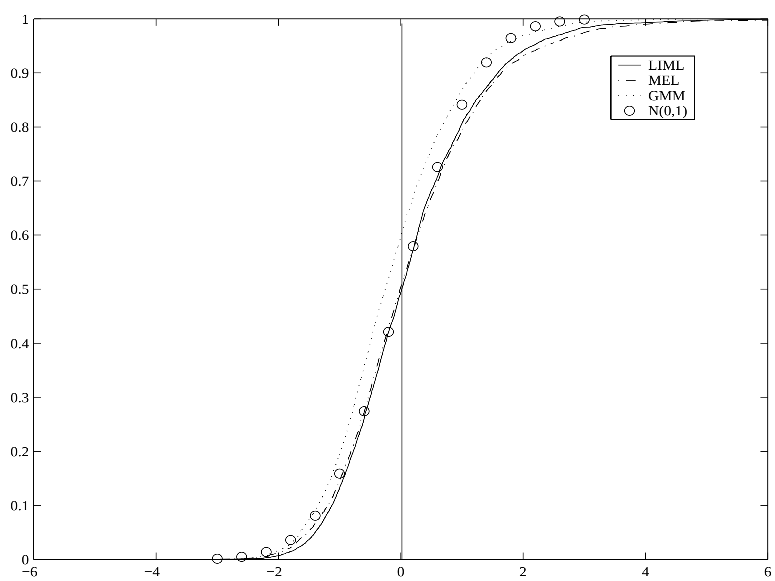

Figure 2: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 30$

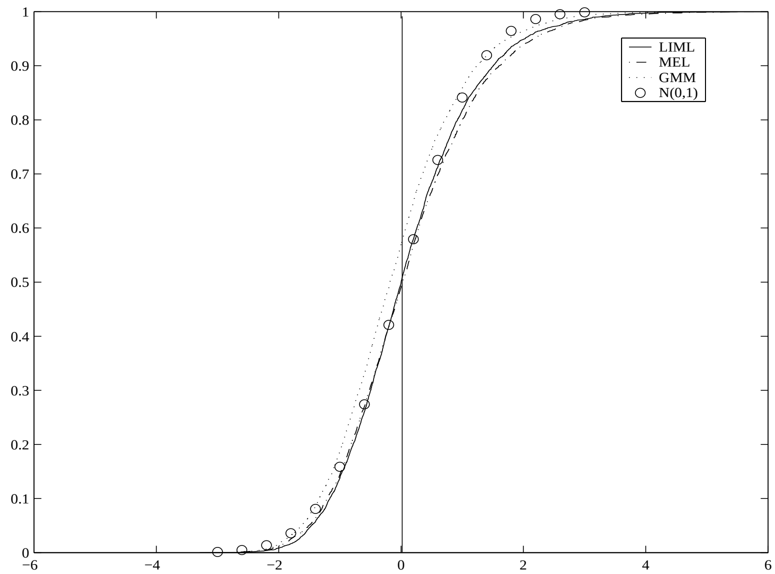

Figure 3: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 50$

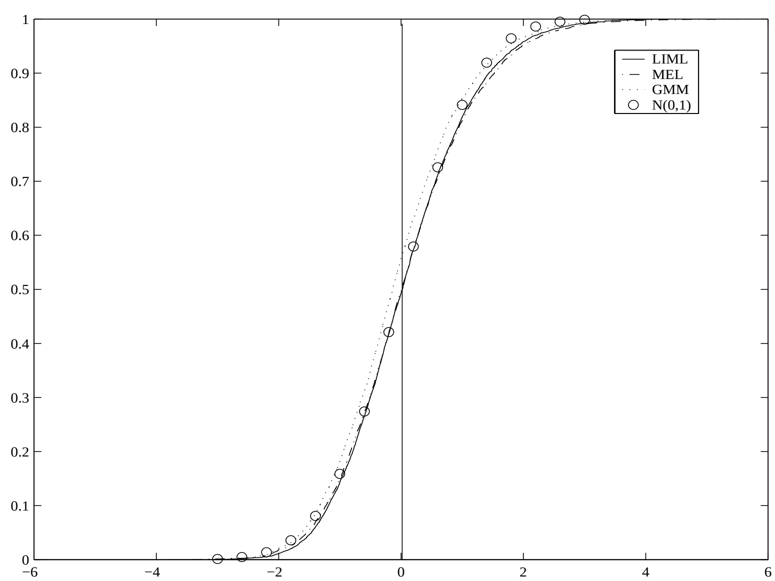

Figure 4: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 100$

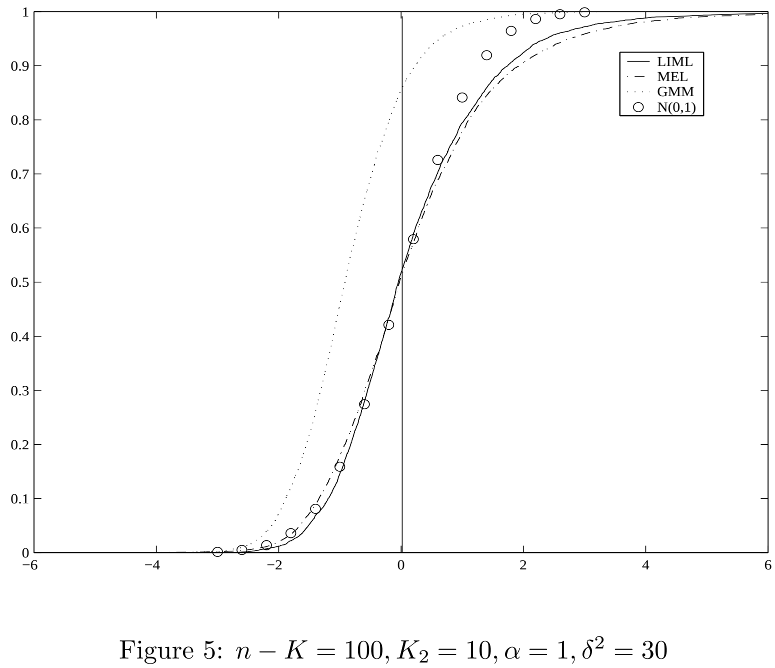

Figure 5: $n-K=100, K_2=10, \alpha=1, \delta^2=30$

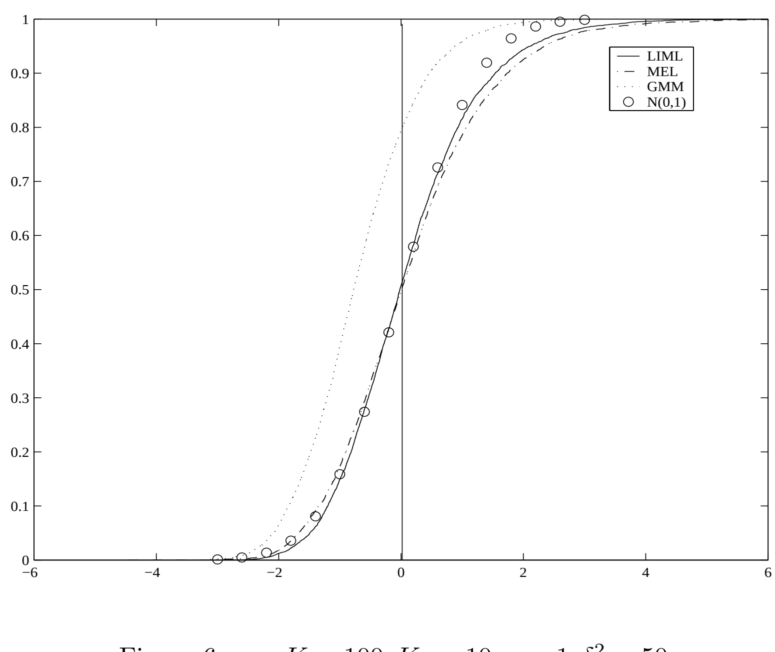

Figure 6: $n-K=100, K_2=10, \alpha=1, \delta^2=50$

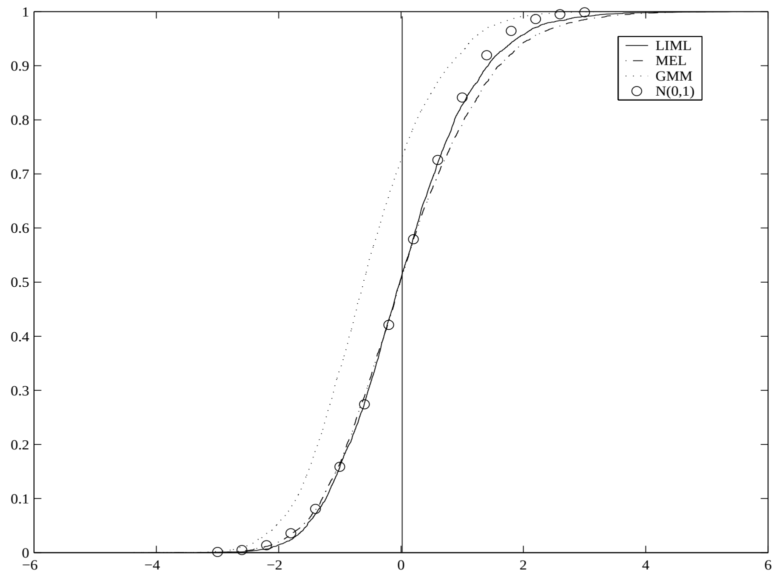

Figure 7: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 100$

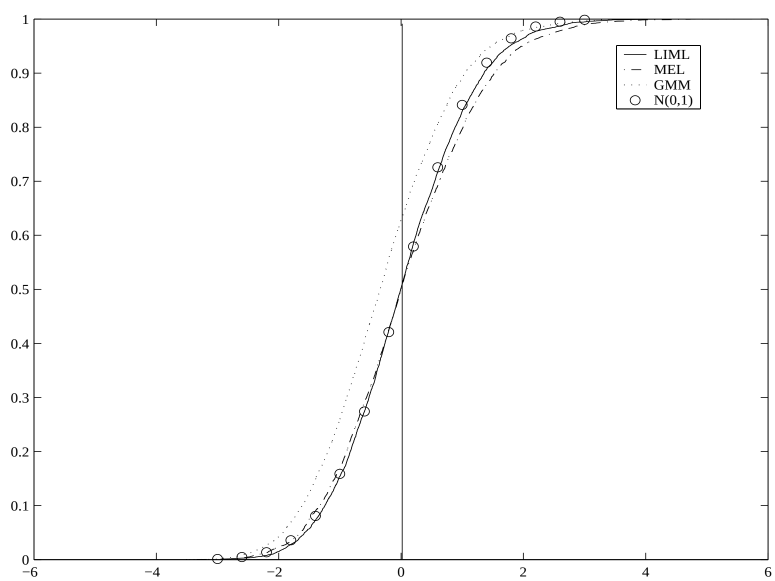

Figure 8: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 300$

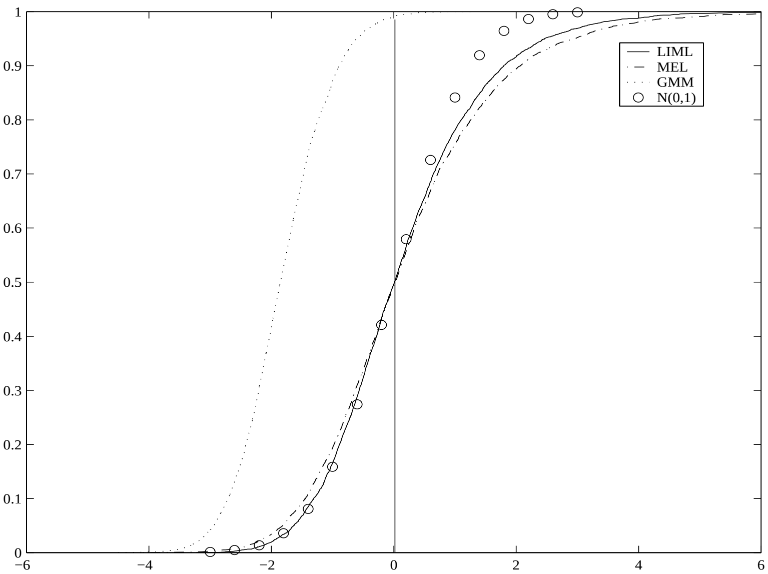

Figure 9: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 50$

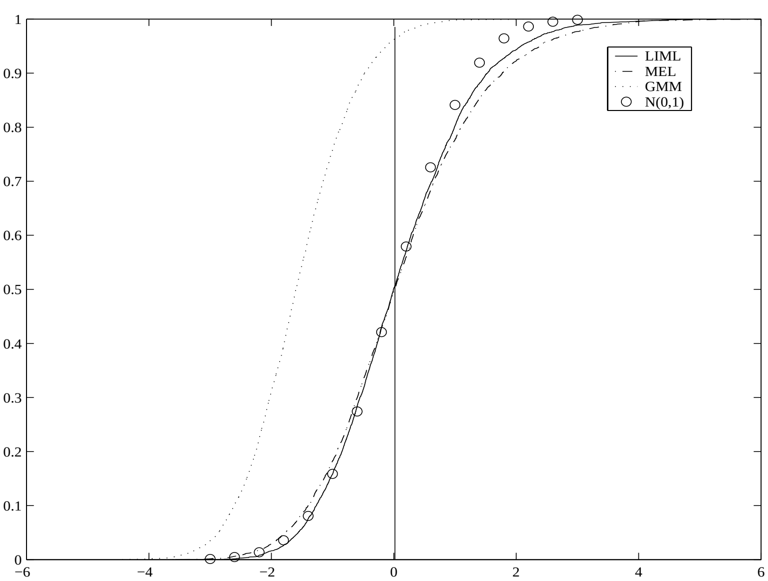

Figure 10: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 100$

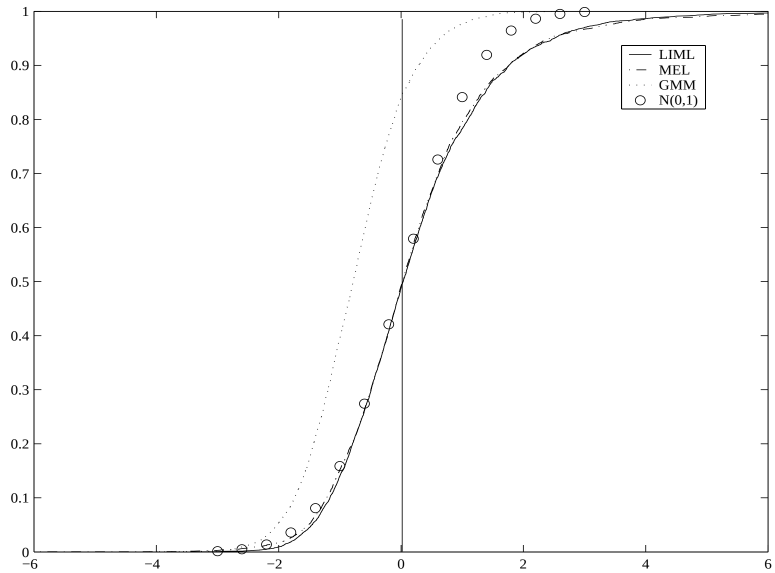

Figure 11: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 30, u_i = \frac{\chi^2(3)-3}{\sqrt{6}}$

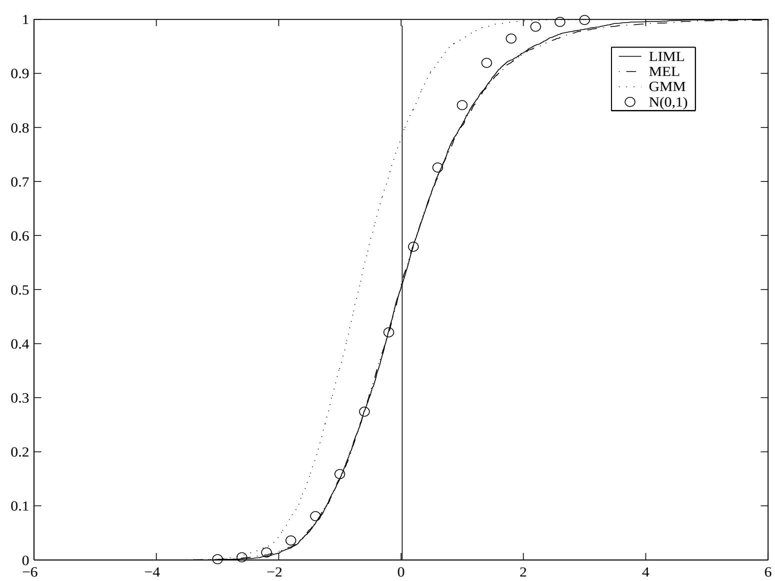

Figure 12: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = \frac{\chi^2(3)-3}{\sqrt{6}}$

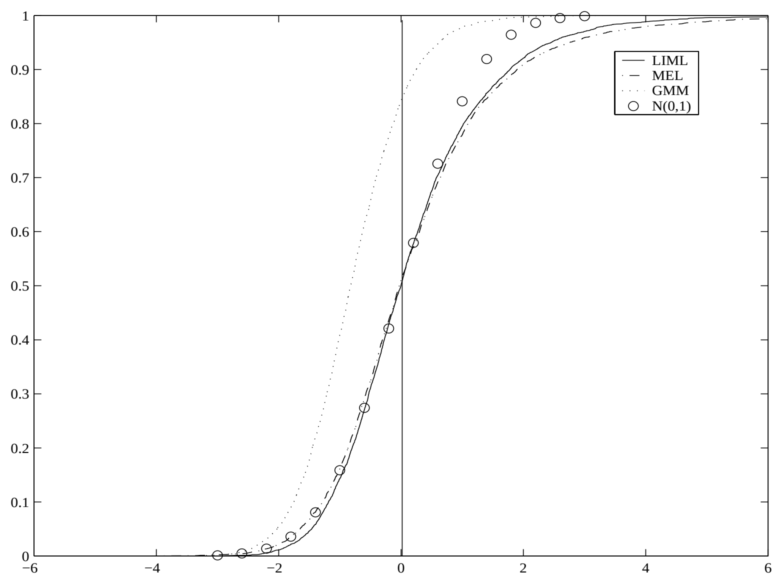

Figure 13: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 30, u_i = t(5)$

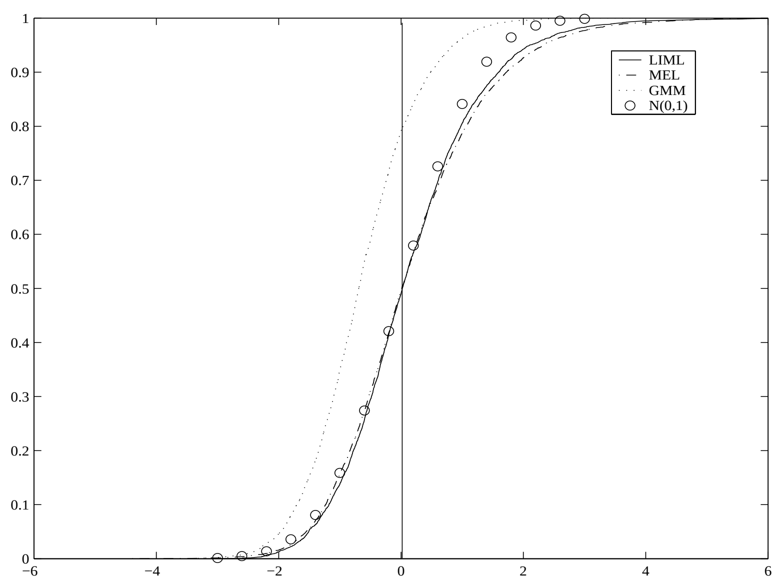

Figure 14: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = t(5)$

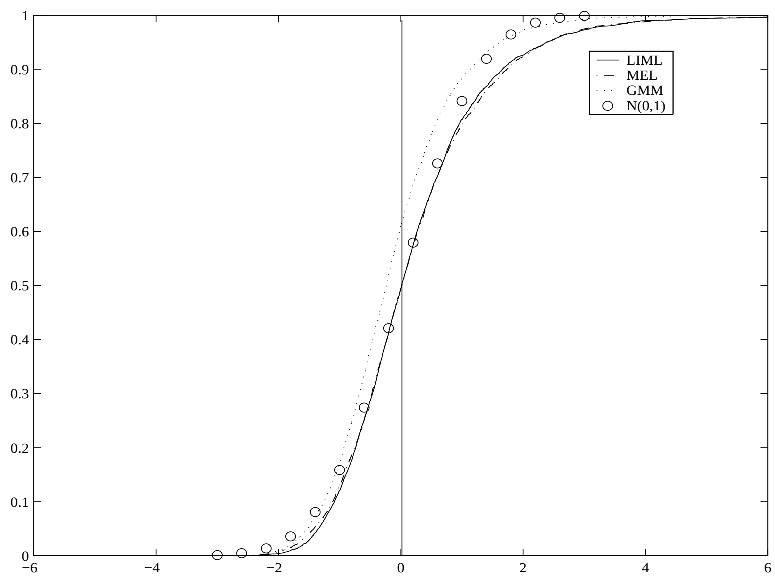

Figure 15: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 30, u_i = \|z_i\|\epsilon_i$

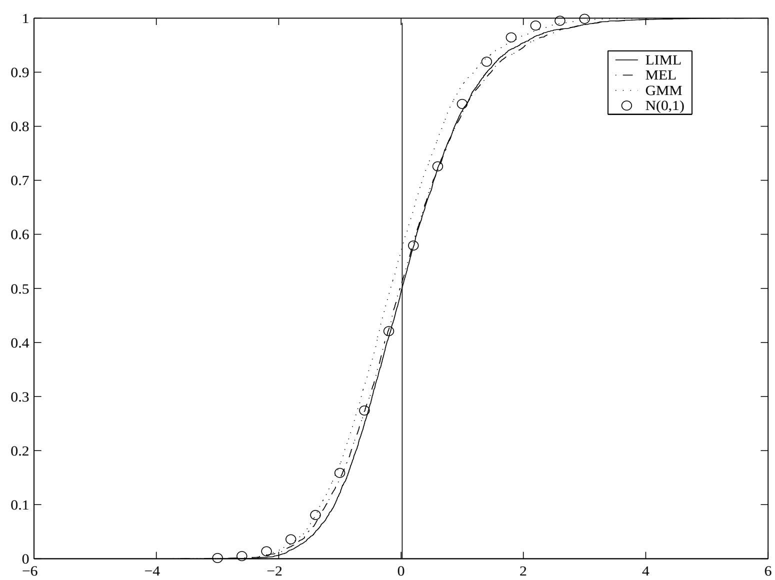

Figure 16: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

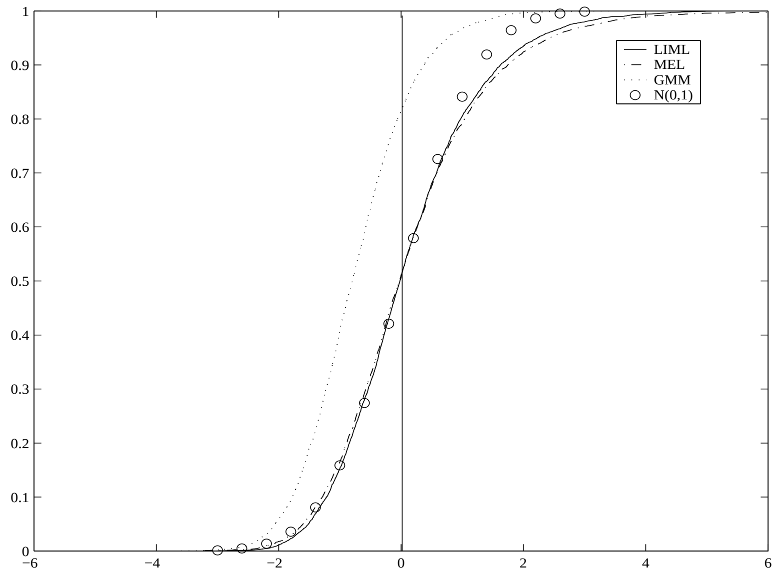

Figure 17: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = \|z_i\|\epsilon_i$

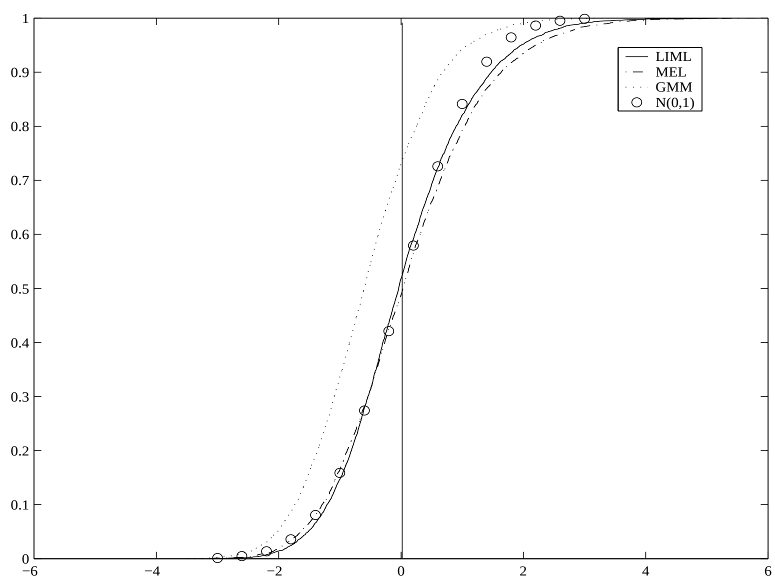

Figure 18: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

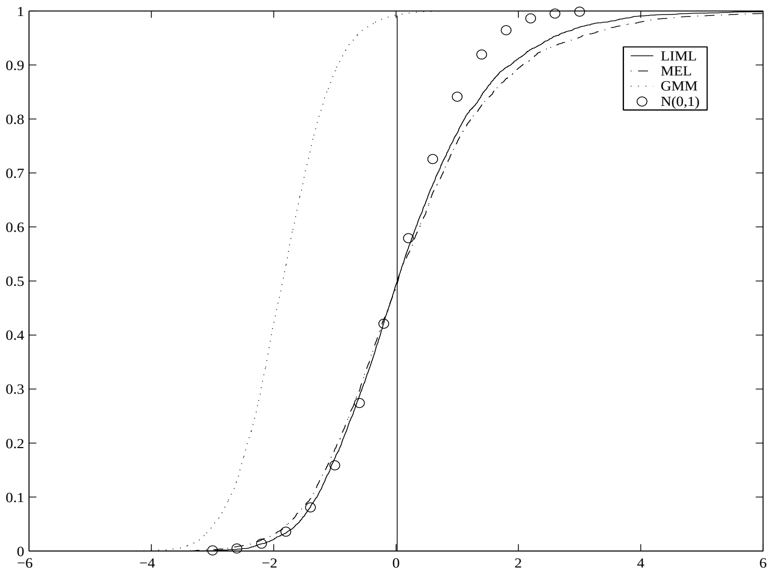

Figure 19: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 50, u_i = \|z_i\|\epsilon_i$

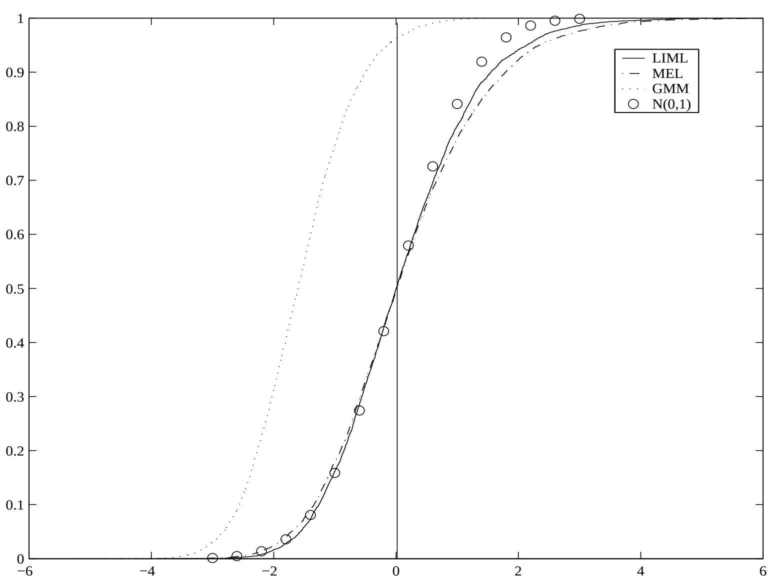

Figure 20: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

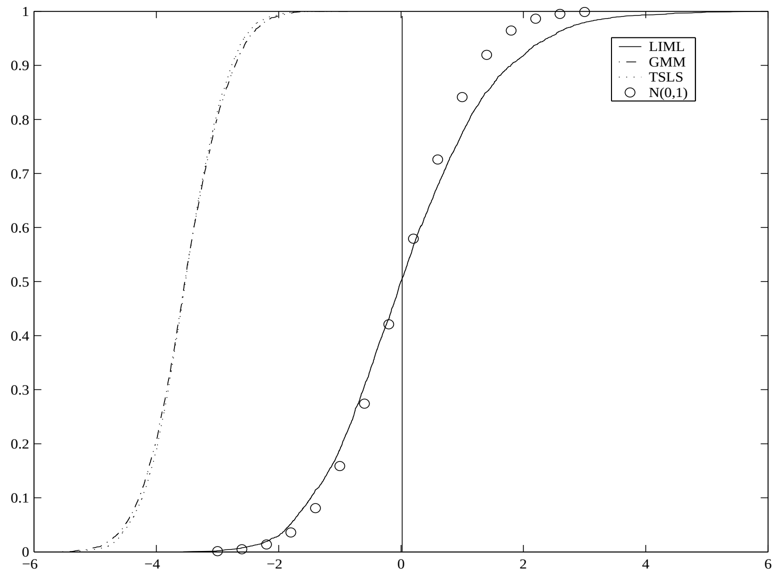

Figure 21: $n - K = 1000, K_2 = 100, \alpha = 1, \delta^2 = 100$

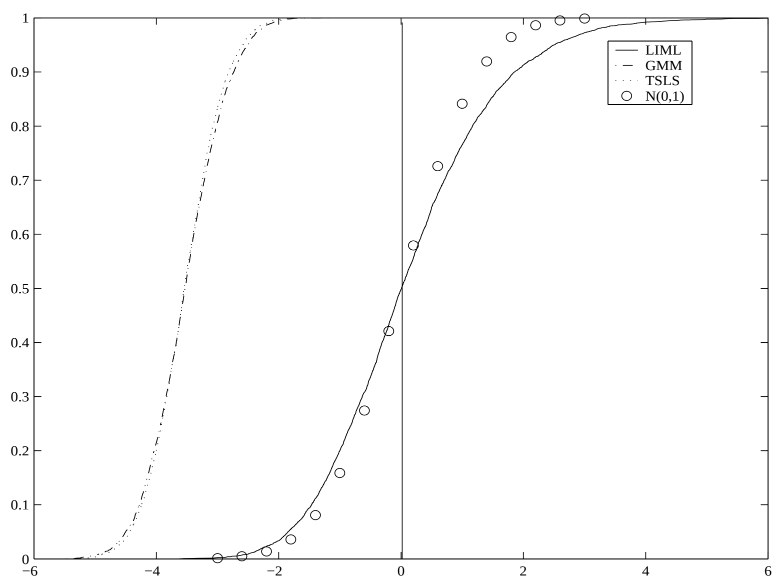

Figure 22: $n - K = 1000, K_2 = 100, \alpha = 1, \delta^2 = 100, u_i = \|z_i\|\epsilon_i$

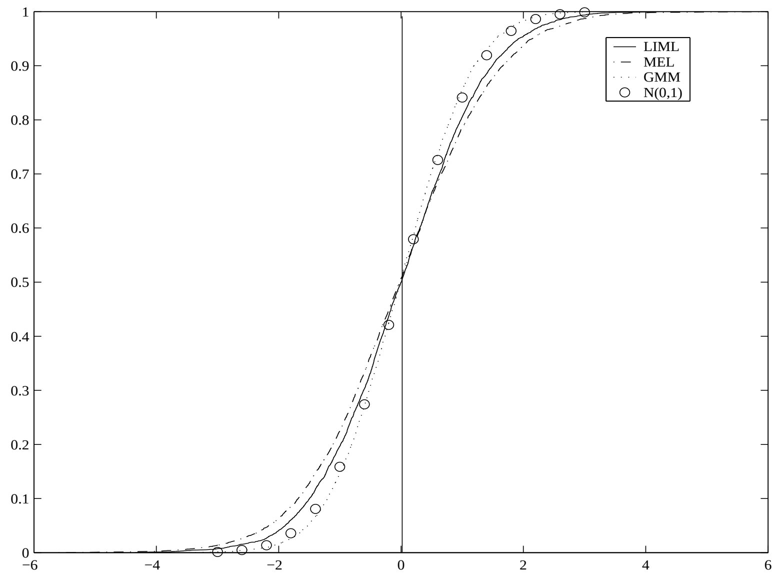

Figure 23: $n - K = 300, K_2 = 30, \alpha = 0, \delta^2 = 100$

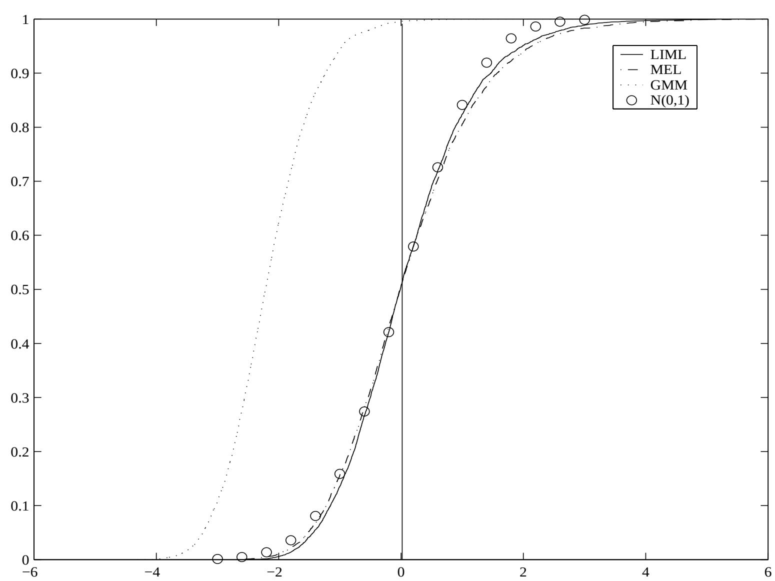

Figure 24: $n - K = 300, K_2 = 30, \alpha = 5, \delta^2 = 100$

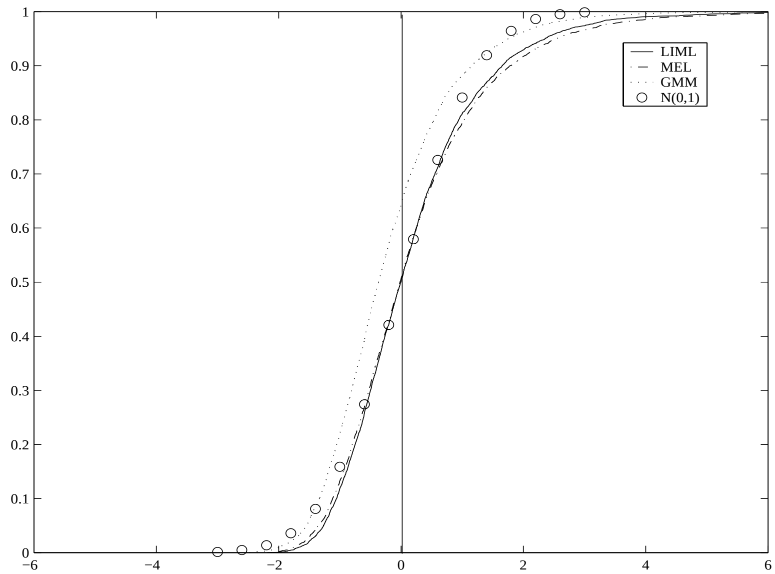

Figure 25: $n - K = 30, K_2 = 3, \alpha = 5, \delta^2 = 30$

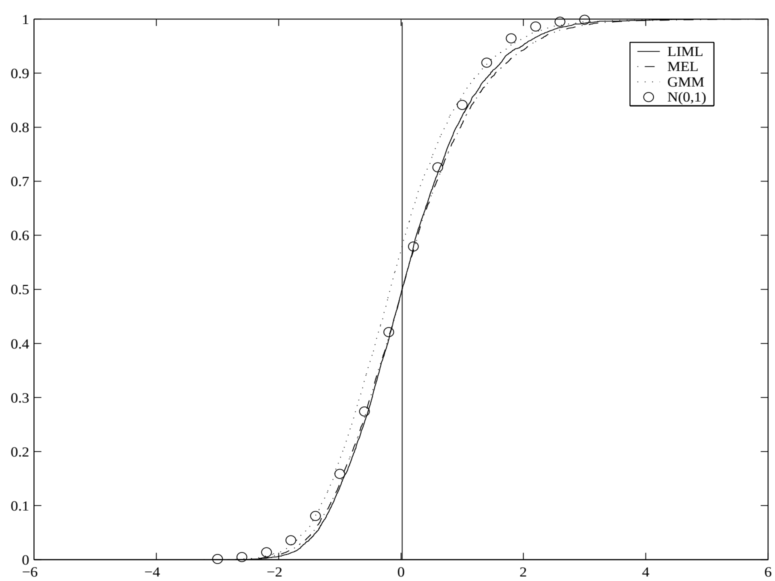

Figure 26: $n - K = 30, K_2 = 3, \alpha = 5, \delta^2 = 100$