CIRJE-F-577

On Finite Sample Properties of Alternative Estimators of Coefficients in a Structural Equation with Many Instruments

T. W. Anderson
Stanford University

Naoto Kunitomo
University of Tokyo

Yukitoshi Matsushita
University of Tokyo

July 2008

On Finite Sample Properties of Alternative Estimators of Coefficients in a Structural Equation with Many Instruments *

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

July 16, 2008

Abstract

We compare four different estimation methods for the coefficients of a linear structural equation with instrumental variables. As the classical methods we consider the limited information maximum likelihood (LIML) estimator and the two-stage least squares (TSLS) estimator, and as the semi-parametric estimation methods we consider the maximum empirical likelihood (MEL) estimator and the generalized method of moments (GMM) (or the estimating equation) estimator. Tables and figures of the distribution functions of four estimators are given for enough values of the parameters to cover most linear models of interest and we include some heteroscedastic cases and nonlinear cases. We have found that the LIML estimator has good performance in terms of the bounded loss functions and probabilities when the number of instruments is large, that is, the micro-econometric models with "many instruments" in the terminology of recent econometric literature.

Key Words

Finite Sample Properties, Empirical Likelihood, GMM, Simultaneous Equations with Many Instruments, Limited Information Maximum Likelihood, Nonlinear LIML

*TWA08-7-25-II (Final Version for JOE). This is the second part of a revised version of Discussion Paper CIRJE-F-321 (Graduate School of Economics, University of Tokyo, February 2005) which was presented at the Econometric Society World Congress 2005 at London (August 2005) and the Kyoto conference (July 2006) in the honor of Professor Kimio Morimune. We thank Cheng Hsiao, Yoichi Arai, the editor of this volume and the referees for some comments to the earlier versions.

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

1. Introduction

In some recent microeconometric applications many instrumental variables have been used in estimating important structural equations. This feature may be due to the possibility of using a large number of cross sectional data and other instrumental variables. One empirical example of this kind often cited in econometric literature is Angrist and Krueger (1991); there is some discussion by Bound et al. (1995). Because there are some distinctive aspects when the number of instrumental variables is large, we investigate the basic properties of the standard estimation methods of microeconometric models. The new development suggests reconsidering the traditional estimation methods. There is a growing recent literature on related problems; see Bekker (1994), Newey and Smith (2004), and Chao and Swanson (2005), for instance, among many others.

The study of estimating a single structural equation in econometric models has led to developing several estimation methods as 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 (Anderson and Rubin (1949, 1950)). See Anderson and Sawa (1979), Anderson, Kunitomo, and Sawa (1982), Mariano (1982), Morimune (1985), and Davidson and MacKinnon (1993), for studies of their finite sample properties, for instance. As semi-parametric estimation methods, the generalized method of moments (GMM) estimation, originally proposed by Hansen (1982), which is essentially the same as the estimating equation (EE) method, has been used in econometric applications. Also the maximum empirical likelihood (MEL) method has been proposed and has received attention (see Owen (2001)). For sufficiently large sample sizes the LIML and the TSLS estimators have approximately the same distribution in the standard large sample asymptotic theory, but their exact distributions can be quite different for the sample sizes occurring in practice. Although the GMM and the MEL estimators have approximately the same distribution under the more general heteroscedastic disturbances in the standard large sample asymptotic theory, their exact distributions can be quite different for the sample sizes occurring in practice.

The main purpose of this paper is to give information about the small sample properties of the exact cumulative distribution functions (cdf's) of these four different estimators for a wide range of parameter values; they have some asymptotic optimalities. We shall pay special attention to the finite sample properties of alternative estimators when we have *many instruments* in the simultaneous equations. Since it is quite difficult to obtain the exact densities and cdf's of these estimators, the numerical information makes possible the comparison of properties of alternative estimation methods. We intentionally use the classical estimation setting of a linear structural equation when we have a set of instrumental variables, but also we shall mention to some heteroscedastic models and nonlinar models for illustrations. It is our intention to make precise comparisons of alternative estimation procedures in the possible simplest case which has many applications. It is possible to generalize our formulation into several directions including many types of nonlinearities and heteroscedasticities as our examples. The present paper corresponds to the second part of our work on the problem and the first part (Anderson, Kunitomo and Matsushita (2007)) gave the asymptotic justification by the finite sample findings.

An important approach to the study of the finite sample properties of alternative estimators is to obtain asymptotic expansions of the exact distributions in normalized forms. The leading term in the asymptotic expansions in the standard large sample theory is the same for all estimators, but the higher-order terms are different. See Fujikoshi et al. (1982), Takeuchi and Morimune (1985), Anderson, Kunitomo and Morimune (1986), Kunitomo (1987) and their citations for the LIML and the TSLS estimators, and Kunitomo and Matsushita (2006) for the MEL and the GMM estimators. Newey and Smith (2004) considered the bias and the mean squared errors of some estimators in more general nonlinear cases. It should be noted, however, that the mean and the mean squared errors of the exact distributions of estimators are not necessarily the same as the mean and the mean squared errors of the asymptotic expansions of the distributions of estimators. In fact the LIML estimator does not possess any moment of positive integer order. (See Mariano and Sawa (1972) and Phillips (1980)). We shall investigate the exact cumulative distributions of the LIML, MEL, GMM, and TSLS estimators directly in a systematic way. We shall compare the estimators on the basis of probabilities of statistical interest, such as significance levels and confidence intervals.

In Section 2 we state the models and alternative estimation methods of unknown parameters. Then in Section 3 we shall explain our tables and figures of the finite sample distributions of alternative estimators and discuss their finite sample properties including simple heteroscedastic cases and nonlinear cases. In Section 4 we shall discuss the approximations of the distribution functions based on their asymptotic expansions. Then the conclusions of our study will be given in Section 5. Tables and figures are gathered in the Appendix.

2. Model and Alternative Estimation Methods of A Structural Equation with Instruments

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

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

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and vector $G_2 \times 1$ of endogenous variables, $\boldsymbol{z}_{1i}$ is a vector of K_1 included exogenous variables in (2.1), $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters, and u_i are mutually independent disturbance terms with $\mathcal{E}(u_i) = 0$ and $\mathcal{E}(u_i^2) = \sigma^2$ $(i = 1, \cdots, n)$. We assume that (2.1) is one equation in a system of $1 + G_2$ endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$. The vector of $K (= K_1 + K_2)$ instrumental variables $\mathbf{z}_i$ satisfies the orthogonal condition $\mathcal{E}[\boldsymbol{z}_i u_i] = \mathcal{E}[\boldsymbol{z}_i (y_{1i} - \boldsymbol{\beta}_2^{'} \boldsymbol{y}_{2i} - \boldsymbol{\gamma}_1^{'} \boldsymbol{z}_{1i})] = \mathbf{0}$ $(i = 1, \cdots, n)$. The reduced form is

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

where $\mathbf{Y} = (\mathbf{y}_i^{'})$ is the $n \times (1 + G_2)$ matrix of endogenous variables, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) = (\mathbf{z}_i^{'})$ is the $n \times K$ matrix of $K_1 + K_2$ instrument vectors $\mathbf{z}_i = (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{'})^{'}$, $\mathbf{V} = (\mathbf{v}_i^{'})$ is the $n \times (1 + G_2)$ matrix of disturbances with $\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$ and the (positive definite) covariance matrix

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} , \tag{2.3}$$

and $\boldsymbol{\Pi}$ is the $(K_1 + K_2) \times (1 + G_2)$ matrix of coefficients. The relation between the coefficients in (2.1) and (2.2) is

$$\boldsymbol{\Pi} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{pmatrix} \tag{2.4}$$

and $(\boldsymbol{\pi}_{21}, \boldsymbol{\Pi}_{22})$ is a $K_2 \times (1 + G_2)$ matrix of coefficients.

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

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

where μ and $\boldsymbol{\nu}$ are a scalar and a vector of Lagrange multipliers, and p_i $(i = 1, \cdots, n)$ is the weighted probability function to be chosen. The above maximization is the same as maximizing

$$L_{1n}(\boldsymbol{\nu},\boldsymbol{\theta}) = -\sum_{i=1}^{n} \log\left\{1 + \boldsymbol{\nu}^{'}\mathbf{z}_i \left[y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}\right]\right\} , \tag{2.5}$$

where $\hat{\mu} = n$ and $[n\hat{p}_i]^{-1} = 1 + \boldsymbol{\nu}^{'}\mathbf{z}_i[y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}]$ (see Qin and Lawless (1994) or Owen (2001)). By differentiating (2.5) with respect to $\boldsymbol{\nu}$ and combining the resulting equation for $\hat{p}_i$ $(i = 1, \cdots, n)$, we have the relations $\sum_{i=1}^{n} \hat{p}_i \mathbf{z}_i(n) \left[y_{1i} - \hat{\boldsymbol{\gamma}}_1^{'}\mathbf{z}_{1i} - \hat{\boldsymbol{\beta}}_2^{'}\mathbf{y}_{2i} \right] = \mathbf{0}$ and $\hat{\boldsymbol{\nu}} = [\sum_{i=1}^{n} \hat{p}_i u_i^2(\hat{\boldsymbol{\theta}})\mathbf{z}_i\mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} u_i(\hat{\boldsymbol{\theta}})\mathbf{z}_i]$, where $u_i(\hat{\boldsymbol{\theta}}) = y_{1i} - \hat{\boldsymbol{\gamma}}_1^{'}\mathbf{z}_{1i} - \hat{\boldsymbol{\beta}}_2^{'}\mathbf{y}_{2i}$ and $\hat{\boldsymbol{\theta}}^{'} = (\hat{\boldsymbol{\gamma}}_{1.EL}^{'}, \hat{\boldsymbol{\beta}}_{2.EL}^{'})$ is the maximum empirical likelihood (MEL) estimator of the vector of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\gamma}_1^{'}, \boldsymbol{\beta}_2^{'})^{'}$. The MEL estimator of $\boldsymbol{\theta}$ in the linear models can be written as the solution of the equations $\hat{\boldsymbol{\nu}}^{'}\sum_{i=1}^{n} \hat{p}_i\mathbf{z}_i[-(\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})] = 0$, which implies

$$\begin{aligned} & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i\mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i\, y_{1i}] \\ = \; & [\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{z}_{1i} \\ \mathbf{y}_{2i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\boldsymbol{\theta}})^2 \mathbf{z}_i\mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i(\mathbf{z}_{1i}^{'}, \mathbf{y}_{2i}^{'})]\begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.EL} \\ \hat{\boldsymbol{\beta}}_{2.EL} \end{pmatrix} . \end{aligned} \tag{2.6}$$

The GMM estimator can be written as the solution of (2.6) when $u_i(\hat{\boldsymbol{\theta}})$ is replaced by $u_i(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\theta}}$ is a consistent initial estimator of $\boldsymbol{\theta}$ (TSLS was used) and the fixed probability weight functions as $p_i = 1/n$ $(i = 1, \cdots, n)$ (see Hayashi (2000), for instance).

In order to relate the MEL and GMM estimators and the LIML and TSLS estimators, we consider the homogeneity condition $\sum_{i=1}^{n} p_i u_i(\boldsymbol{\theta})^2\mathbf{z}_i\mathbf{z}_i^{'} = \sigma^2\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}$ and $(1/n)\sum_{i=1}^{n} u_i^2(\boldsymbol{\theta}) = \sigma^2$. The resulting maximization problem under the homogeneity restrictions requires $\hat{\boldsymbol{\nu}} = (1/\hat{\sigma}^2)[\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}]^{-1}[\sum_{i=1}^{n} u_i(\hat{\boldsymbol{\theta}})\mathbf{z}_i]$ $(u_i(\hat{\boldsymbol{\theta}}) = y_{1i} - \hat{\boldsymbol{\gamma}}_1^{'}\mathbf{z}_{1i} - \hat{\boldsymbol{\beta}}_2^{'}\mathbf{y}_{2i})$. Then by using the approximation $\log(1 + x) \sim x$, (2.5) becomes approximately

$$L_{2n}(\boldsymbol{\theta}) = (-n)\frac{\left[\sum_{i=1}^{n}\mathbf{z}_i^{'}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})\right]\left[\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}\right]^{-1}\left[\sum_{i=1}^{n}\mathbf{z}_i(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})\right]}{\sum_{i=1}^{n}(y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i})^2}, \tag{2.7}$$

which is $(-n)$ times the variance ratio in turn. The minimum of the variance ratio gives the LIML estimator $\hat{\boldsymbol{\beta}}_{LI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.LI}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$, which is the solution of

$$(2.8) \qquad (\frac{1}{n}\mathbf{G} - \frac{1}{n-K}\lambda\mathbf{H})\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0} \,,$$

where $n - K > 0$ and λ is the smallest root of $|(1/n)\mathbf{G} - l(1/(n-K))\mathbf{H}| = 0$. Here we use the notation $\mathbf{G} = \mathbf{Y}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}$, $\mathbf{H} = \mathbf{Y}^{'}\left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\right)\mathbf{Y}$, $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}^{'}\mathbf{Z}_{2.1}$, $\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ and

$$(2.9) \qquad \mathbf{A} = \begin{pmatrix} \mathbf{Z}_1^{'} \\ \mathbf{Z}_2^{'} \end{pmatrix} (\mathbf{Z}_1, \mathbf{Z}_2) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \,,$$

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

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

It minimizes the numerator of the variance ratio L_{2n} and corresponds to the GMM estimator if we put $\hat{p}_i = 1/n$ $(i = 1, \cdots, n)$ and the homogeneity condition.

The statistical methods of the LIML and TSLS estimation were originally developed by Anderson and Rubin (1949, 1950) and it is the (classical) maximum likelihood estimation method with the limited information on instrumental variables. When the disturbances are homoscedastic and normally distributed, $\mathbf{G}$ and $\mathbf{H}$ are sufficient statistics; the LIML and TSLS estimators depend only on them. The nonlinear LIML estimator can be defined by substituting $u_i(\boldsymbol{\theta}) = f_i(y_{1i}, \mathbf{z}_{1i}, \mathbf{y}_{2i}, \boldsymbol{\theta})$ for $u_i(\boldsymbol{\theta}) = y_{1i} - \boldsymbol{\gamma}_1^{'}\mathbf{z}_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{y}_{2i}$ $(i = 1, \cdots, n)$ and minimizing the variance ratio in (2.7). (The nonlinear TSLS estimator in the same way. The alternative or standard nonlinear LIML and TSLS extensions have been discussed in Chapter 8 of Amemiya (1985).)

3. Evaluation of Exact Distribution Functions and Tables

3.1 Parameterization

The evaluation of the cdf's of estimators we have used is based on simulation. In order to describe our evaluation method, we use an expanded formulation and notation of the classical study of Anderson et. al. (1982) except that here the sample size is n. We concentrate on the comparison of the estimators of the coefficient parameter of the endogenous variables and we shall investigate the finite sample distributions of the estimator expressed as

$$F(\mathbf{x}) = \Pr\left(\frac{1}{\sigma}\left[\mathbf{\Pi}_{22}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}\right]^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \le \mathbf{x}\right) \tag{3.1}$$

for $\mathbf{x} = (x_1, \cdots, x_{G_2})$. The limit of (3.1) in the large sample asymptotics is $N_{G_2}(\mathbf{0}, \mathbf{I}_{G_2})$ for any (asymptotically) efficient estimator under the homoscedasticity assymption. It is easier to interpret the distribution functions in this form rather than with some other normalization. We use the notation of the noncentrality

$$\mathbf{\Delta} = \mathbf{\Omega}_{22}^{-1/2}\mathbf{\Pi}_{22}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}\mathbf{\Omega}_{22}^{-1/2} \tag{3.2}$$

and the standardized vector of coefficients

$$\boldsymbol{\alpha} = \frac{1}{\sqrt{\omega_{11.2}}}\mathbf{\Omega}_{22}^{1/2}(\boldsymbol{\beta}_2 - \mathbf{\Omega}_{22}^{-1}\boldsymbol{\omega}_{21}) \ , \tag{3.3}$$

where $\omega_{11.2} = \omega_{11} - \boldsymbol{\omega}_{12}\mathbf{\Omega}_{22}^{-1}\boldsymbol{\omega}_{21}$.

When $G_2 = 1$ in particular, Anderson et al. (1982) have utilized the fact that the explicit distributions of (3.1) for the *normalized* LIML estimator and *normalized* TSLS estimator under the standard case (that is, the disturbances are homoscedastic and normally distributed) depend only on *the key parameters* K_2, $n-K$, α and $\delta^2 = \mathbf{\Delta}$. (See Anderson (1974) for the details.) Notice that $\mathbf{\Omega}_{22}^{-1}\boldsymbol{\omega}_{21}$ is the regression coefficient of v_{1i} on $\mathbf{v}_{2i}$ and $\omega_{11.2}$ is the conditional variance of v_{1i} given $\mathbf{v}_{2i}$. When $G_2 = 1$, we rewrite $\eta = -\alpha/\sqrt{1+\alpha^2} = (\omega_{12} - \omega_{22}\beta_2)/[\sigma\sqrt{\omega_{22}}]$ $(\omega_{22} = \mathbf{\Omega}_{22})$, which is the correlation coefficient between the two random variables u_i and v_{2i} (or y_{2i}) and it is the coefficient of simultaneity in the structural equation of the simultaneous equations system. The numerator of the noncentrality parameter δ^2 represents the additional explanatory power due to $\mathbf{y}_{2i}$ over $\mathbf{z}_{1i}$ in the structural equation and its denominator is the error variance of $\mathbf{y}_{2i}$. Hence the noncentrality δ^2 determines how well the equation is defined in the simultaneous equations system, and $n-K$ is the degrees of freedom of $\mathbf{H}$ which estimates $\mathbf{\Omega}$ in the LIML method; it is not relevant to the TSLS method. The normalization part in (3.1) can also be written as the square root of $[\delta^2/(1+\alpha^2)] \times [\omega_{22}/\omega_{11.2}]$. The distribution of (3.1) does not depend on the units of measurements of y_{1i} and y_{2i}.

Some econometricians use the terminology *many weak instruments* for the cases when K_2 is large while δ^2 is not that large as n such that $\delta^2/n \to 0$ and $\delta^2/K_2 \to a$ (> 0). We have tried to choose the key parameter values to make useful interpretations.

3.2 Simulation Procedure

By using Monte Carlo simulations we obtain empirical cdf's of estimators of the coefficients of the endogenous variables in the structural equation as follows. We generate a set of random numbers by using a system of (2.1) and

$$\mathbf{y}_{2i} = \mathbf{\Pi}_2^{'}\mathbf{z}_i + \mathbf{v}_{2i} \ , \tag{3.4}$$

where $\mathbf{z}_i \sim N(0, \mathbf{I}_K)$, $u_i \sim N(0, \sigma^2)$, $\mathbf{\Pi}_2$ is a $K \times G_2$ matrix of coefficients and $v_{2i} \sim N_{G_2}(0, \mathbf{\Omega}_{22})$ with $\mathcal{E}(u_i \mathbf{v}_{2i}) = \boldsymbol{\omega}_{21} - \mathbf{\Omega}_{22}\boldsymbol{\beta}_2$ $(i = 1, \cdots, n)$. Since the model of (2.1) and (3.4) is consistent with the reduced form (2.2), we have $u_i = v_{1i} - \boldsymbol{\beta}_2^{'}\mathbf{v}_{2i}$, $\sigma^2 = \omega_{11} - 2\boldsymbol{\beta}_2^{'}\boldsymbol{\omega}_{12} + \boldsymbol{\beta}_2^{'}\mathbf{\Omega}_{22}\boldsymbol{\beta}_2$, and $\mathbf{z}_i$ are independent of u_i and $\mathbf{v}_{2i}$ $(i = 1, \cdots, n)$ in the homoscedastic case. We take a set of true values of parameters $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_1$, σ^2, $\mathbf{\Omega}$ to satisfy the restrictions in (2.1) and (3.4) given the value of $\boldsymbol{\alpha}$, and then we control the elements of $\mathbf{\Delta}$ by setting values for the $(1 + K_2)$-vectors $\mathbf{\Pi}_2 = (\boldsymbol{\pi}_{2j})$. [1]

Following Owen (2001) the maximization in the MEL estimation has been done in 2-steps; the inner loop for the numerical calculation of Lagrange multiplier in (2.5) and the outer loop for the minimization with respect to the unknown parameters. We have used the derivative based maximization routine in the inner loop and a simplex-method based optimization algorithm in the outer loop by utilizing (2.6). There is a non-trivial computational problem on the MEL estimation when the noncentrality parameter is near to zero, which is pointed out by Mittelhammer et al. (2006), for instance. Therefore we have made computations for cases where we did not have a problem in numerical convergence. [2] For the LIML and TSLS estimators as (2.8) and (2.10), there are simple ways to express $\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2$ in terms of two matrices $\mathbf{G}$ and $\mathbf{H}$ (see Anderson et al. (1982) for instance). For each simulation we generated a set of random variables from the disturbance terms and exogenous variables. In each simulation the number of repetitions was 5,000.

[1] In order to examine whether our results strongly depend on the specific values of parameters, however, we have done several simulations for other values of parameters.

[2] The numerical convergence of the outer loop is not guaranteed in the MEL estimation (Owen (2001)) and we have confirmed that there can be some problems in the extreme cases.

As an illustration we give Figure 1A for the distribution functions of estimators in normalized terms when $G_2 = 2$ and the disturbances are normally distributed. They are given in terms of two marginal distributions of (3.1); $F_1(x_1) = F(x_1, \infty)$ and $F_2(x_2) = F(\infty, x_2)$, respectively. Each limiting distribution is $N(0,1)$. [3] In this example we have set the parameters : $n - K = 300, K_2 = 30$, $\boldsymbol{\alpha} = (1,1)'$ $(\boldsymbol{\omega}_{12} = (\rho, \rho), \rho = -1/\sqrt{3})$, $\mathbf{\Omega}_{22} = \mathbf{I}_2$ and

$$\mathbf{\Delta} = \begin{pmatrix} 100 & 1.5 \\ 1.5 & 50 \end{pmatrix} .$$

Although we have investigated aspects of the distributions of four estimators when $G_2 > 1$ as in Figure 1A, each depends on many parameters and we may require too many tables and figures to obtain useful information in a systematic way. Thus, from now on, we shall give tables (Tables 1A-9A) and figures (Figures 2A-11A, 1B-3B) only for $G_2 = 1$. We control the values of key parameters in order to compare the distributions of estimators in a limited number of cases in a systematic way.

In order to investigate the effects of nonnormal disturbances on the distributions of estimators, we used many nonnormal distributions, but we only report two cases when the distributions of the disturbances are skewed or fat-tailed. As the first case we have generated a set of random variables $(y_{1i}, \mathbf{y}_{2i}, \mathbf{z}_i)$ by using (2.1) and (3.4), $u_i = -(\chi_i^2(3) - 3)/\sqrt{6}$, and $\chi_i^2(3)$ are χ^2-random variables with 3 degrees of freedom. As the second case, we took the t-distribution with 5 degrees of freedom for the disturbance terms. Also in order to investigate the effects of heteroscedastic disturbances on the distributions of estimators, we have considered the form of $\mathcal{E}(u_i^2) = \sigma^2(\mathbf{z}_i)$ and in particular $u_i = \|\mathbf{z}_i\| u_i^*$ $(i = 1, \cdots, n)$, and u_i^* $(i = 1, \cdots, n)$ are homoscedastic disturbance terms as the typical example.

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 spline functions to estimate their percentiles. The distributions are tabulated in standardized terms because this form of tabulation makes comparisons and interpolation easier. Each table includes three quartiles, the 5 and 95 percentiles and the interquartile range of the distribution. 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) es-

[3] It is possible to give figures for each components of coefficient vectors. Then the normalizations for the components become messy and the comparison with the limiting distribution may become less clear.

timator, which corresponds to the GMM estimator when $\hat{u}_i^2$ is replaced by a constant (namely σ^2), that is, the variance-covariance matrix is homoscedastic and known. The exact distribution of the TSLS estimator has been studied and tabulated extensively by Anderson and Sawa (1979). We do not report the details of our results, but we have found that the differences between the exact cdf and its estimate are less than 0.005 in most cases and the maximum difference is about 0.008. Hence our estimates of the cdf's are quite accurate; we have accuracy to two digits.

3.3 Distributions of the MEL and LIML Estimators

For $\alpha = 0$, the densities of the LIML and MEL estimators are close to symmetric. (See Table 8A and Figure 9A). 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. (See Tables 1A-3A and Figures 2A-4A.) For given α, δ^2, and n, the asymmetry increases with K_2. The main finding from the tables is that the distributions of the MEL and LIML estimators are roughly symmetric around the true parameter value and they are almost median-unbiased. This finite sample property holds even when K_2 is fairly large. At the same time, their distributions have relatively long tails. As $\delta^2 \to \infty$, the distributions approach $N(0,1)$; however, for small values of δ^2 there are appreciable probabilities outside the range of 3 or 4 times ASD(asymptotic standard deviation)'s. (When δ^2 is extremely small, we cannot ignore the tail probabilities for practical purposes. See Table 9A.) As δ^2 increases, the spread of the normalized distribution decreases. Also the distribution of the LIML estimator has slightly tighter tails than that of the MEL estimator. For given α,K_2, and δ^2, the spread decreases as n increases and it tends to increase with K_2 and decrease with α .

Also we have found that some of our findings on the MEL estimator are also pointed out by Guggenberger (2005) in this subsection.

3.4 Distributions of the GMM and TSLS Estimators

We have included tables of the distributions of the GMM and TSLS estimators. However, since they are quite similar in most cases, we have included only the distribution of the GMM estimator in many figures. The most striking feature of the distributions of the GMM and TSLS estimators is that they are skewed towards the left for $\alpha > 0$ (and towards the right for $\alpha < 0$), and the distortion increases with α and K_2 . The MEL and LIML estimators are close to median-unbiased in each case while the GMM and TSLS estimators are biased. As K_2 increases, this bias becomes more serious; for $K_2 = 10, 30$ and 100 , 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 their distributions have 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 $K_2 = 10 \sim 100$ (Figures 3A, 4A and 3B), 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. These observations on the distributions of the MEL estimator and the GMM estimator are analogous 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.

3.5 Effect of the difference between the structural coefficient and the error regression

Before the development of inference for the model of simultaneous equations, the structural coefficient, say, β_2, was estimated by the sample regression y_1 on y_2, that is, by Ordinary Least Squares (OLS). That estimation procedure could result in very biased estimates. The LIML and TSLS estimators were developed to improve the OLS estimator, but the TSLS estimation ignores the information on $\mathbf{\Omega}$ in (2.3).

Table 8A and Figure 9A compare estimation procedures for some different values of α. The bias of the (normalized) TSLS increases with α; the median goes from 0 at $\alpha = 0$ to -2.22 for $\alpha = 5$. The interquartile range goes from 1.19 at $\alpha = 0$ to .86 at $\alpha = 5$ as compared to LIML (from 1.56 to 1.37). However, the 95 percentile goes from 1.46 to

-1.02; that is, if α is as large as 1, the probability of a negative estimator is greater than .95 when α is large. In effect, α is a *nuisance* parameter. It can have a large effect on the bias of the TSLS estimator. In a sense the TSLS has the defect of the OLS estimator, but not as extreme.

3.6 Effects of Nonnormality 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 in the form of $\mathcal{E}(u_i^2) = \sigma^2(\mathbf{z}_i)$. We use the normalization

$$(3.5) \qquad F(\mathbf{x}) = \Pr\left(\left[\mathbf{\Pi}_{22}'\mathbf{Q}_{22.1}\mathbf{\Pi}_{22}\right]^{1/2} (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \leq \mathbf{x} \right) ,$$

where $\mathbf{Q}_{22.1} = \mathbf{Z}_2' \left[\mathbf{R} - \mathbf{R}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{R}\right]\mathbf{Z}_2$, $\mathbf{R} = \mathbf{Z}(\mathbf{Z}'\mathbf{\Sigma}\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{\Sigma} = (\text{diag } \sigma^2(\mathbf{z}_i))$. (When $\sigma^2(\mathbf{z}_i) = \sigma^2$, we have $\mathbf{Q}_{22.1} = \sigma^{-2}\mathbf{A}_{22.1}$ in (3.1).) The limit of (3.5) is $N_{G_2}(\mathbf{0}, \mathbf{I}_{G_2})$ for the MEL and GMM estimators in the large sample asymptotics. In this case the asymptotic variance-covariance matrix for the LIML and TSLS estimators could be slightly larger than those of the MEL and GMM estimators.

Among many tables we show Tables 5A-7A and Figures 7A-8A as the representative heteroscedastic cases ($\sigma^2(\mathbf{z}_i) = \|\mathbf{z}_i\|^2$) by following Hayashi (2000). Also we show Table 4A and Figures 5A-6A as the representative nonnormal disturbances which we have chosen (a χ^2-type and t distributions). From our tables the comparison of the distributions of four estimators are approximately valid even if the distributions of disturbances are different from normal and they are heteroscedasitic in the sense we have specified above. The bias and skewness of the distributions have relatively large effects and they often dominate the nonnormality and heteroscedasticity. Thus the effects of heteroscedasticity and nonnormality of disturbances on the exact distributions of alternative estimators have the secondary importance in our setting.

When the disturbance terms are heteroscedastic with many instruments, Anderson, Kunitomo and Matsushita (2007) assumed the 6th order moments condition for the disturbances and the key condition

$$(3.6) \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)} - c_n\right]^2 = 0 ,$$

where $p_{ii}^{(n)} = (\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}')_{ii}$ and $c_n = K_2/n$. They have shown that the LIML estimator has still desirable asymptotic properties. The typical (two) examples satisfying this condition are (i) the case of $c_n \to 0$ ($p_{ii}^{(n)} \to 0$) and (ii) the case when we have dummy variables which have 1 or -1 in their all components so that $(1/n)\mathbf{A}_{22.1} = \mathbf{I}_{K_2}$ and $p_{ii}^{(n)} = K_2/n$ ($i = 1, \cdots, n$). When (3.6) is not satisfied with many instruments, the LIML estimator may have some biases in extreme cases. Kunitomo (2008) has considered some modifications of the LIML estimation. [4]

3.7 On Nonliner LIML Estimator

There is an interesting question if our observations on alternative estimators are specific for the *linear* structural equation models or not. Although there can be many possible nonlinear models, we shall report only some results on two nonlinear cases with two endogenous variables ($y_{1i}, y_{2i}; i = 1, \cdots, n$). The first case of nonlinearity is a linear structural equation with nonlinear instruments

$$\mathcal{E}\left[\begin{pmatrix} \mathbf{w}_i \\ \mathbf{w}_i^{\otimes 2} \end{pmatrix}(y_{1i} - \beta_2\, y_{2i})|\mathbf{w}_i\right] = \mathbf{0}\ , \tag{3.7}$$

where $y_{2i} = (\mathbf{w_i}', \mathbf{w}_i^{\otimes 2})\mathbf{\Pi}_2 + v_{2i}$, $\mathbf{w}_i = (w_{1i}, \cdots, w_{mi})'$, $\mathbf{w}_i^{\otimes 2} = (w_{1i}^2, \cdots, w_{mi}^2)'$, $\mathbf{w}_i$ follows $N(\mathbf{0}, \mathbf{I}_K)$ and $\mathcal{E}[\,\cdot\,|\mathbf{w}_i]$ is the conditional expectation given $\mathbf{w}_i$. We have set $m = 5$ and then the number of instruments $K = 10$.

As the second nonlinear example, we have a nonlinear structural equation

$$\mathcal{E}\left[\mathbf{z}_i(y_{1i} - \beta_2\, y_{2i}^2)|\mathbf{z}_i\right] = \mathbf{0}\ , \tag{3.8}$$

where $y_{2i} = \mathbf{z_i}'\mathbf{\Pi}_2 + v_{2i}$, $\mathbf{z}_i$ follows $d(1, \cdots, 1)' + N_K(\mathbf{0}, \mathbf{\Omega}_z)$ (d is a constant) and

$$\mathbf{\Omega}_z = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{K-1} \end{bmatrix}\ .$$

In this case we have set the number of instruments $K = 10$ and used $\mathbf{z}_i^{\otimes 2}$ as the instrumental variables in the nonlinear estimation.

[4] When we have time series data for the simultaneous equations model, some parametric models for the heteroscedasticities of disturbances have been developed as one referee has pointed out. We have examined some possibilities including the stationary GARCH models and have found that the essential conclusions on alternative estimators are unchanged. In order to obtain the limiting normality for the LIML estimator, we need to require some moment conditions for the disturbances and thus we need careful analysis on the stationarity, for instance. Since the problem is related to the vast growing concerns in time series econometrics, we did not discuss them in detail. See McAleer (2005), for instance.

We have used the nonlinear LIML, MEL and GMM estimators which are mentioned at the end of Section 2. We give the cdfs of the nonlinear LIML, MEL and GMM estimators for Case 1 and Case 2 as Figures 10A and 11A, respectively. We have normalized the estimators of the coefficient β_2 as in the linear cases such that we can compare the finite sample properties of the alternative estimators. (As the parameters, α was constructed as before and δ^2 has been constructed so that the resulting normalized LIML estimator has the limiting N(0,1) distribution in the large sample asymptotics, for instance.) Since the evaluation methods of cdfs are basically the same as the linear cases,we have omitted the details.

The most important observation is the fact that the finite sample properties of the nonlinear LIML, MEL and GMM estimators are similar to the ones we have discussed for the linear cases.

4. Discussion on Distributions of Estimators

4.1 The Moments and Monte Carlo Experiments

We have mentioned the fact that some estimators do not necessarily have the exact moments under reasonable assumptions. The first moment of a scalar random variable X is said to be infinite or is said not to exist if for any given positive constant c there is a constant a such that

$$\int_{-a}^{a} |x| dF(x) > c \ ,$$

where $F(\cdot)$ is the cdf of X. In this case $\mathcal{E}(X)$ is not defined as a finite number. However, nevertheless a Monte Carlo experiment can be conducted and the sample mean of the sample calculated. What kind of conclusion can be drawn ? As a simple illustration of the problem of interpreting Monte Carlo experiments, we take i.i.d. observations X_i $(i = 1, \cdots, n)$ from $N(\theta^{-1}, 1)$ when $\theta \neq 0$. As a reasonable estimator of θ we take $\hat{\theta} = \bar{X}_n^{-1}$ and $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then $\sqrt{n} \left[\theta^{-2}\right] \left[\bar{X}_n^{-1} - \theta\right] \xrightarrow{d} N(0, 1)$, but we can calculate the sample bias and MSE in Monte Carlo experiments even though we know that MSE is $+\infty$. We have confirmed the fact in our experiments that the sample bias and MSE of $\hat{\theta}$ calculated from Monte Carlo experiments are *not stable* and they are *not reliable* quantities even on average with many replications. The extension of the above example to the problem of estimating simultaneous equations can be made. However, it suggests that before conducting a Monte Carlo experiment a mathematical study should be carried out to verify that the parameter being estimated actually exists. (See Mariano and Sawa (1972) as an early development.)

Our method of analysis in this paper is free from this issue because we compare the estimators on the basis of probabilities.

4.2 Approximations of Finite Sample Distributions

The exact distribution functions of alternative estimators in the general case are very complicated, but it is possible to derive the asymptotic expansions of the density functions of alternative estimators as shown by Anderson (1974), Anderson and Sawa (1973), and Kunitomo and Matsushita (2006). Although the asymptotic expansions in the general case ($G_2 \geq 1$) look complicated even for the linear simultaneous equations, they give some useful information in particular when $G_2 = 1$ and the disturbances are normally distributed. In the (standard) large sample asymptotics, the noncentrality (or concentration) parameter divided by n is assumed to approach a limit as $n \longrightarrow \infty$. It is convenient to use the noncentrality parameter given by

$$\mu^2 = (1+\alpha^2)\frac{\mathbf{\Pi}_{22}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}}{\omega_{22}} = (1+\alpha^2)\delta^2 \tag{4.1}$$

and the *semi-parametric parameter* given by

$$\tau = 2\frac{1+\alpha^2}{\omega_{22}}(1,\mathbf{0})Q_{11}^{-1}\mathbf{Q}\mathbf{D}^{'}\mathbf{F}\mathbf{D}\mathbf{Q}Q_{11}^{-1}\begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} , \tag{4.2}$$

where $\mathbf{Q} = (\mathbf{D}^{'}\mathbf{M}\mathbf{D})^{-1}$, $Q_{11} = \left(\mathbf{\Pi}_{22}^{'}\mathbf{M}_{22.1}\mathbf{\Pi}_{22}\right)^{-1}$, $\mathbf{M}_{22.1} = \mathrm{plim}_{n\rightarrow\infty}n^{-1}\mathbf{A}_{22.1}$, $\mathbf{M} = \mathrm{plim}_{n\rightarrow\infty}n^{-1}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}$, $\mathbf{F} = \mathrm{plim}_{n\rightarrow\infty}n^{-1}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{'}\left[\mathbf{M}^{-1} - \mathbf{D}(\mathbf{D}^{'}\mathbf{M}\mathbf{D})^{-1}\mathbf{D}^{'}\right]\mathbf{z}_i\mathbf{z}_i^{'}$ and a $(K_1+K_2)\times(G_2+K_1)$ coefficient matrix $\mathbf{D} = \left[\mathbf{\Pi}_2, (\mathbf{I}_{G_2}, \mathbf{O})^{'}\right]$. We need this semi-parametric factor because we estimate the variance-covariance matrix in the MEL and GMM estimation.

For the GMM estimator, an asymptotic expansion of its distribution function as $n \rightarrow \infty$ (K_2 is fixed) when the disturbances are normally distributed ($N(0,\sigma^2)$) and μ^2 is proportional to n is given by

$$\mathbf{P}\left(\frac{\sqrt{\mathbf{\Pi}_{22}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}}}{\sigma}(\hat{\beta}_{2.GMM} - \beta_2) \leq x\right) \tag{4.3}$$

$$
\begin{aligned}
= \quad & \Phi(x) + \Big\{ -\frac{\alpha}{\mu}[x^2 - (K_2 - 1)] \\
& -\frac{1}{2\mu^2}[(\tau + (K_2 - 1)^2\alpha^2 - (K_2 - 1))x + (1 - 2K_2\alpha^2)x^3 + \alpha^2 x^5] \Big\} \phi(x) + O(\mu^{-3}) \ .
\end{aligned}
$$

For the MEL estimator, an asymptotic expansion of its distribution function as $n \to \infty$ (K_2 is fixed) is given by

$$
\begin{aligned}
(4.4) \qquad & \mathbf{P}\left(\frac{\sqrt{\mathbf{\Pi}_{22}'\mathbf{A}_{22.1}\mathbf{\Pi}_{22}}}{\sigma}(\hat{\beta}_{2.MEL} - \beta_2) \le x \right) \\
= \quad & \Phi(x) + \Big\{ -\frac{\alpha}{\mu}x^2 - \frac{1}{2\mu^2}[(\tau + K_2 - 1)x + (1 - 2\alpha^2)x^3 + \alpha^2 x^5] \Big\} \phi(x) \\
& +O(\mu^{-3}),
\end{aligned}
$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and the density function of the standard normal distribution, respectively.

The asymptotic expansions of the distributions of the TSLS and LIML estimators are (4.3) and (4.4), respectively, with $\tau = 0$. See Anderson and Sawa (1973) and Anderson (1974). They agree with Fujikoshi et al. (1982) for the LIML and TSLS estimators ($G_2 \ge 1$). Because $\tau > 0$, the contribution due to the semiparametric methods is that we have the additional term τ/μ^2 to the asymptotic mean squared errors (AMSE). As a numerical illustration we give Figures 1B in *Appendix*), which show the finite sample distributions and the approximate distributions of the LIML, MEL and GMM estimators in normalized forms as (3.1). Since the limiting distributions of the above estimators are $N(0,1)$ in the large sample asymptotics, they are denoted by "o" as the bench mark.

4.3 An Alternative Approximation

As we have shown in Anderson, Kunitomo and Matsushita (2007) (Part-I of our study), there is an alternative asymptotic theory for the case when the number of excluded instruments K_2 (say K_{2n}) is dependent on the sample size n. (Kunitomo (1980, 1982) and Morimune (1983) were the earlier developers of this theory. For more recent developments, see Bekker (1994), and Chao and Swanson (2005), for instance.) We consider a typical approximation when $K_{2n}/n \to c$ $(0 \le c < 1)$ (and μ^2/n is approximately a constant) as $n \to \infty$. For the LIML estimator, an asymptotic expansion of its distribution function when $G_2 = 1$ as $n \to \infty$ is

$$\mathbf{P}\left(\sqrt{n}(\hat{\beta}_{2.LI} - \beta_2) \leq x\right) = \Phi_\Psi(x) + \frac{1}{\sqrt{n}}\left\{-\frac{|\mathbf{\Omega}|^{1/2}}{\sigma^2}\alpha x^2\right\}\phi_\Psi(x) + O(n^{-1}), \tag{4.5}$$

where $\mathbf{\Omega} = \mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'})$, $\Phi_\Psi(\cdot)$ and $\phi_\Psi(\cdot)$ are the cdf and the density function of $N(0, \Psi)$, respectively,

$$\Psi = \sigma^2\Phi_{22.1}^{-1} + c_*\Phi_{22.1}^{-1}|\mathbf{\Omega}|\Phi_{22.1}^{-1}\ , \tag{4.6}$$

$c_* = c/(1-c)$, and $\Phi_{22.1} = \lim_{n\to\infty}(1/n)\mathbf{\Pi}_{22}^{'}\mathbf{A}_{22.1}\mathbf{\Pi}_{22}$.

By setting $x = 0$ in (4.3) for the GMM estimator, we have $1/2 + \alpha(K_2 - 1)/[\mu\sqrt{2\pi}] + O(\mu^{-3})$. By setting $x = 0$ in (4.4) and (4.5) for the MEL and LIML estimators, we have $1/2 + O(\mu^{-3})$ and $1/2 + O(n^{-1})$, respectively. When $\alpha \neq 0$, the bias of the GMM estimator (and the TSLS estimator) is proportional to K_2/μ, which increases rapidly if K_2 is large in comparison to the noncentrality μ^2. If μ^2 is proportional to K_2, for instance, the left hand side of probability is far from 1/2 whenever $\alpha \neq 0$. On the other hand, the MEL and the LIML estimators are almost median-unbiased and this property holds even if K_{2n} is proportional to n. As numerical illustrations, we give the approximations based on the asymptotic expansions in (4.5) up to $O(n^{-1/2})$ as *A.exp* (large-K_2) in Table 1B and Figures 2B-3B, which gives several approximations of the finite sample distributions of the LIML estimator when K_2 is relatively large. As we may expect, in these cases the normal approximation based on large-K_2 theory (discussed in Part-I) is better than the normal approximation based on the standard large sample theory. (In Table 1B and Figures 2B-3B we have used $h = 1 + (n/\mu^2)\left[K_2/(n-K)\right]$, which is approximately $\Psi \times \Phi_{22.1}/\sigma^2$, for the normalized variance for the limiting distribution.) Also we find that the approximations based on (4.5) are even better than the normal approximations. This observation gives an important implication for the testing problem (see Matsushita (2006)).

5. Conclusions

First, the distributions of the MEL and GMM estimators are asymptotically equivalent in the sense of the limiting distributions in the standard large sample asymptotic theory, but their exact distributions are substantially different in finite samples. The relation of their distributions are quite similar to the distributions of the LIML and TSLS estimators. The MEL and LIML estimators are to be preferred to the GMM and TSLS estimators if K_2 is large. In some microeconometric models and models on panel data, it is often a common feature that K_2 is fairly large. For such situations we have shown (Anderson, Kunitomo and Matsushita (2007)) that the LIML estimator has asymptotic optimality in the large K_2-asymptotics sense. It seems that we need some stronger conditions for the MEL estimator, but its finite sample properties are often similar to the corresponding LIML estimator [5] .

Second, the large-sample normal approximation in the large K_2 asymptotic theory is relatively accurate for the MEL and LIML estimators. Hence the usual methods with asymptotic standard deviations give often 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 normality when K_2 is large, in particular.

Third, it is recommended to use the probability of concentration as a criterion of comparisons because some estimators do not possess any (exact) moments and hence we expect to have *unstable and unreliable* values of the sample bias and mean squared errors of such estimators in Monte Carlo simulations. This is the reason why we directly considered the finite sample distribution functions of alternative estimation methods. The probability criterion we have adopted roughly corresponds to the bounded loss function.

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 (1949,1950). Other estimation methods including the TSLS, the GMM, and the MEL estimation methods have been developed with several different motivations and purposes. Now we have some practical situations in econometric applications where the LIML estimation has advantage over other estimation methods. [6]

---

[5] We have reported the results for estimation problem, but they have a number of important implications for testing problem. See Morimune (1989), Matsushita (2006), and Anderson and Kunitomo (2007), for instance.

[6] Although we have investigated the nonlinear models and the heteroscedastic models to some extents,

References

[1] Amemiya, T. (1985), *Advanced Econometrics*, Blackwell.

[2] Anderson, T.W. (1974), "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System," *Journal of the American Statistical Asociation*, Vol. 69, 565-573.

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

[4] Anderson, T.W. and N. Kunitomo (2007), "On Likelihood Ratio Tests of Structural Coefficients : Anderson-Rubin (1949) revisited," Discussion Paper CIRJE-F-499, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007).

[5] Anderson, T.W. and N. Kunitomo and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models," Discussion Paper CIRJE-F-399, Graduate School of Economics, University of Tokyo.

[6] Anderson, T.W. , N. Kunitomo and Y. Matsushita (2007), "On Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments," Discussion Paper CIRJE-F-542, Graduate School of Economics, University of Tokyo.

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

[8] Anderson, T.W. and 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.

the issue remains on the condition that our conclusions are valid in more general situations including the case of extremely weak instruments. There are many related important issues, which should be future topics.

[9] 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.

[10] Anderson, T.W. and T. Sawa (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansion," *Econometrica*, Vol. 41, 683-714.

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

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

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

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

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

[16] Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

[17] 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.

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

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

[20] Guggenberger, P. (2005), “Finite-Sample Evidence Suggesting a heavy Tail Problem of the Generalized Empirical Likelihhood Estimator,” Unpublished Manuscript, Department of Economics, UCLA.

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

[22] 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.

[23] 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.

[24] Kunitomo, N. (2008), “Improving the LIML estimation with many instruments and Persistent Heteroscedasticities,” Discussion Paper CIRJE-F-576, Graduate School of Economics, University of Tokyo.

[25] Kunitomo, N. and Matsushita, Y. (2006) , “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. Revised.

[26] Mariano, R.S. (1982), ”Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equations Case,” *International Economic Review*, Vol. 23, 503-533.

[27] 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.

[28] Matsushita, Y. (2006), “Estimation and Testing in a Structural Equation with Pos-

sibly Many Instruments," Unpublished Ph.D. Dissertation, Graduate School of Economics, University of Tokyo.

[29] McAleer, M. (2005), "Automated Inference and Learning in Modeling Financial Volatility," *Econometric Theory*, Vol. 21, 232-261.

[30] Mittelhammer, R., G. Judge, and R. Schoenberg (2005), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods," in *Identification and Inference for Econometric Models*, Cambridge University Press, Andrews, D. and J. Stock ed., 285-305.

[31] 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.

[32] Morimune, K. (1985), *Estimation and Testing in Econometric Models*, Kyoritsu (in Japanese), Tokyo.

[33] Morimune, K. (1989), "t Test in a Structural Equation," *Econometrica*, Vol.57-6, 1341-1360.

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

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

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

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

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

[39] Takeuchi, K. and Morimune, K. (1985), "Third Order Efficincy of the Extended Maximum Likelihood Estimator in a Simultaneous Equation System," *Econometrica*, Vol. 53, 177-200.

APPENDIX : TABLES AND FIGURES

Notes on Tables

In Tables 1A-9A the distributions are tabulated in the standardized terms, that is, of (3.1) or (3.5). 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 (3.1) or (3.5) for the MEL and GMM estimators in the standard large sample asymptotic theory are $N(0,1)$ as $n \to \infty$, we added the standard normal case as the bench mark. In Table 1B we also give the normal approximations based on the large-K_2 theory and the approximations based on the asymptotic expansions.

Notes on Figures

In Figures the cdf's of the LIML, MEL and GMM estimators are shown in the standardized terms, that is, of (3.1) or (3.5) in linear models. (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.) Figure 1A with $G_2 = 2$ gives two marginal distributions in (3.1) or (3.5) and other figures are with $G_2 = 1$. The dotted lines were used for the distributions of the GMM estimator. For the comparative purpose we give the standard normal distribution as the bench mark for each case. In Figures 2B-3B we also give the normal approximations based on the large-K_2 theory and the approximations based on the asymptotic expansions. We have used the similar method for heteroscedastic cases and nonlinear cases.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 9A: $\alpha = 1$

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

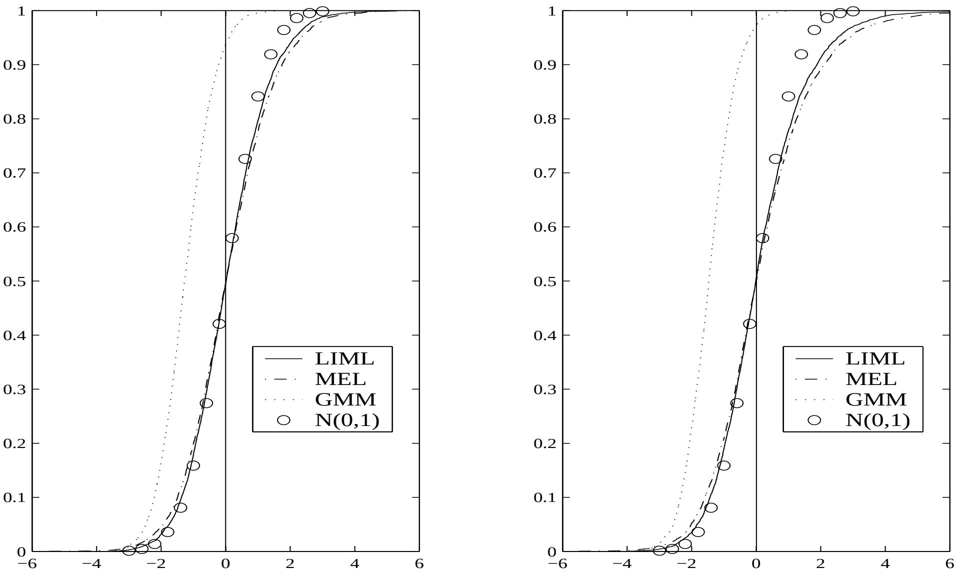

Figure 1A: CDF of Standardized estimators: $n - K = 300, K_2 = 30, \alpha = (1,1)', \mathbf{\Delta} = (100, 1.50; 1.50, 50), u_i \sim N(0,1)$

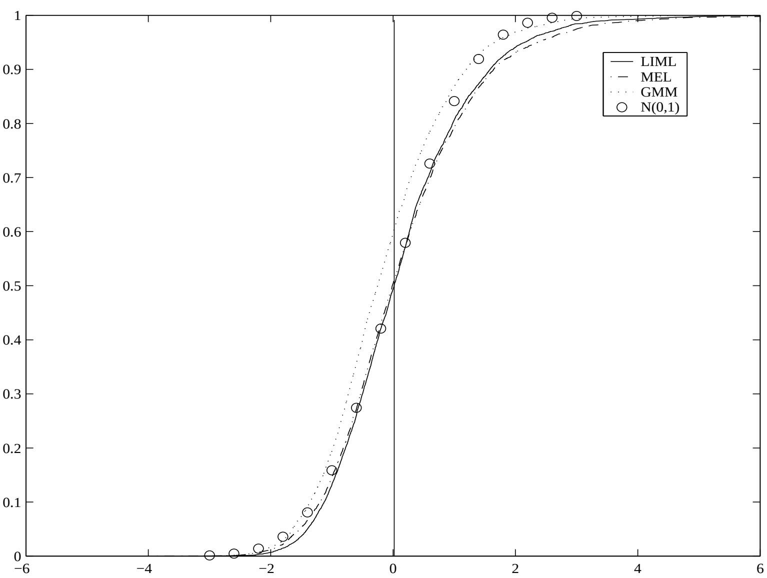

Figure 2A: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 30$

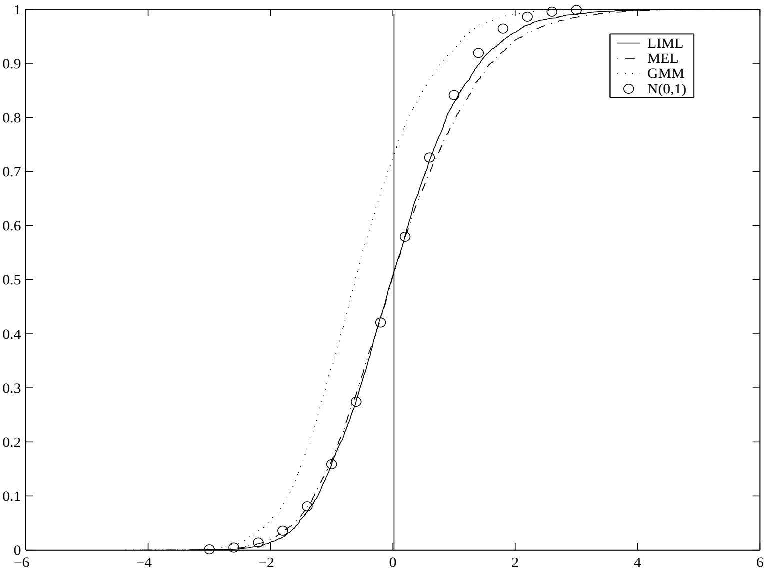

Figure 3A: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 100$

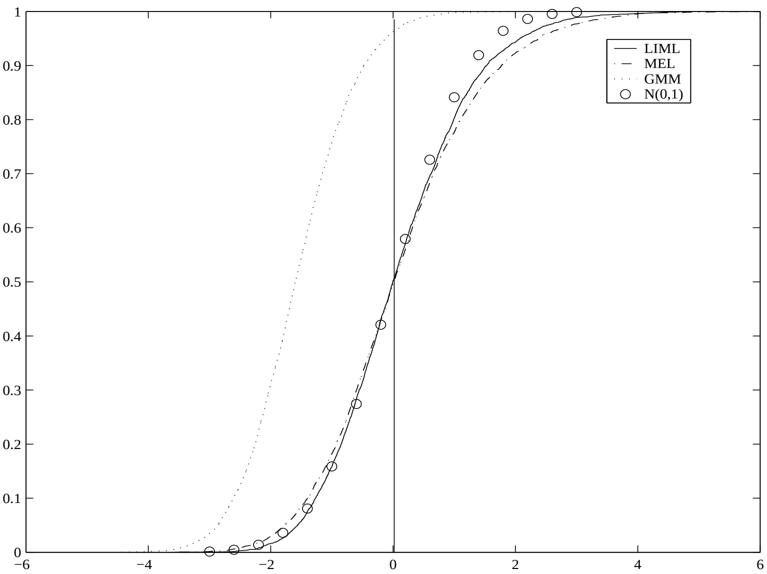

Figure 4A: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 100$

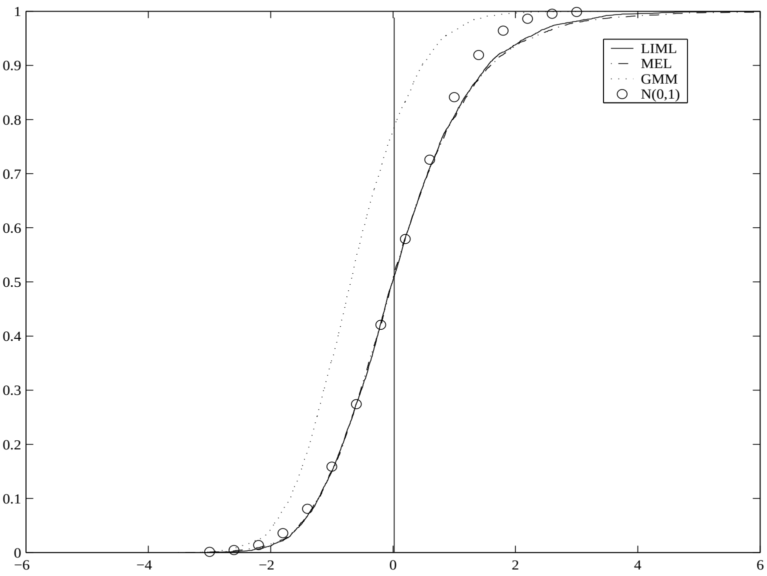

Figure 5A: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = -\frac{\chi^2(3)-3}{\sqrt{6}}$

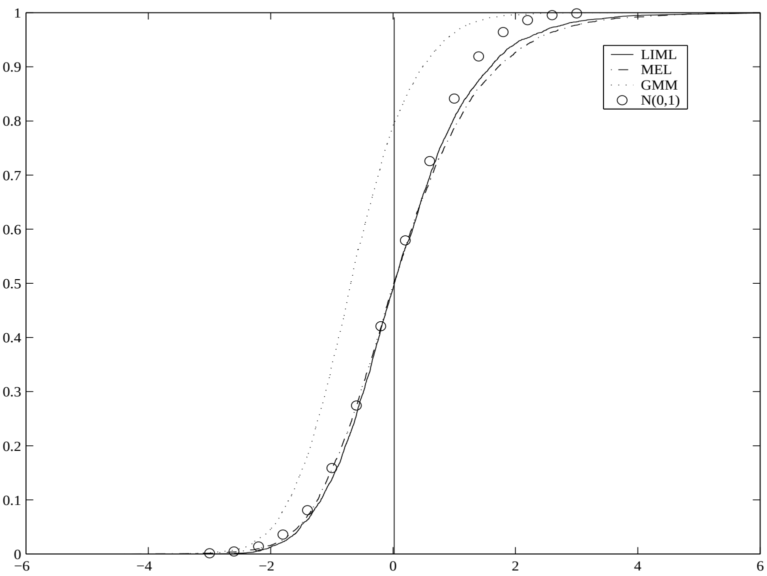

Figure 6A: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50, u_i = t(5)$

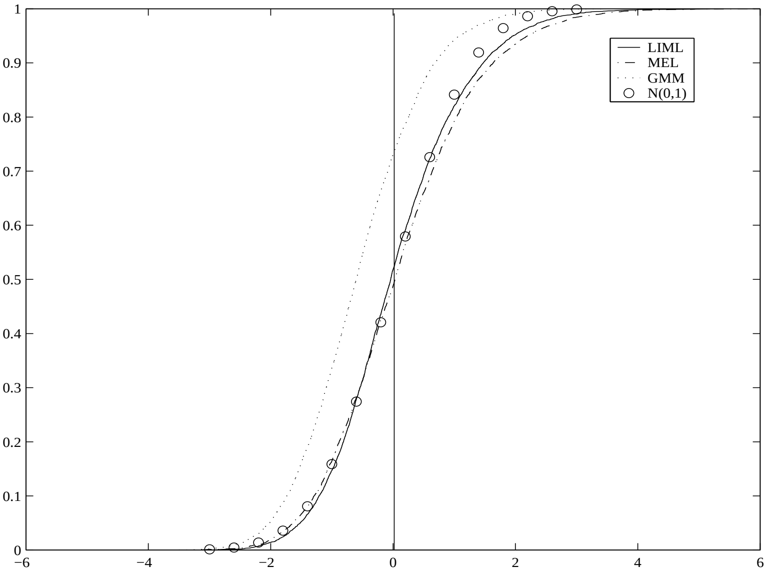

Figure 7A: $n-K=100, K_2=10, \alpha=1, \delta^2=100, u_i=\|z_i\|\epsilon_i$

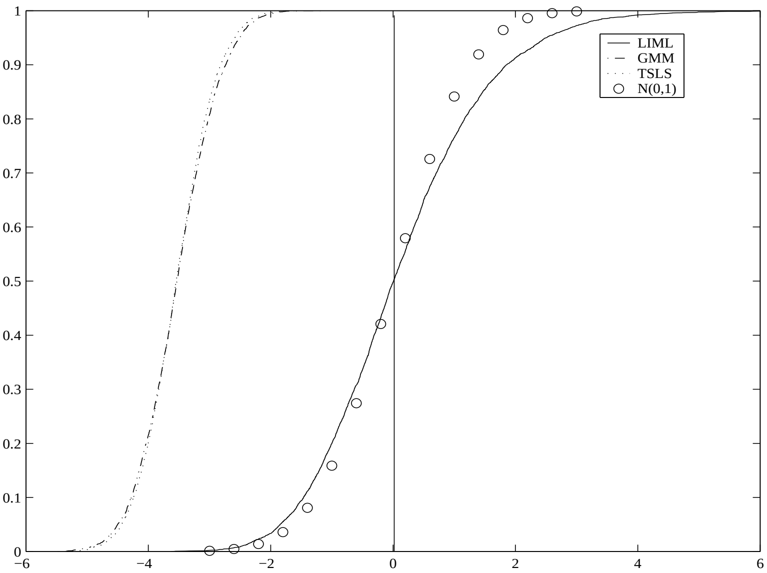

Figure 8A: $n-K=1000, K_2=100, \alpha=1, \delta^2=100, u_i=\|z_i\|\epsilon_i$

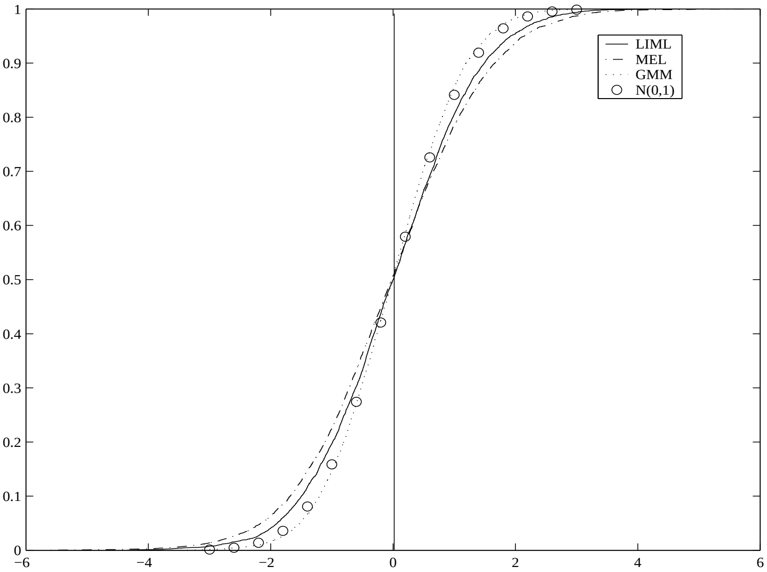

Figure 9A: $n - K = 300, K_2 = 30, \alpha = 0, \delta^2 = 100$

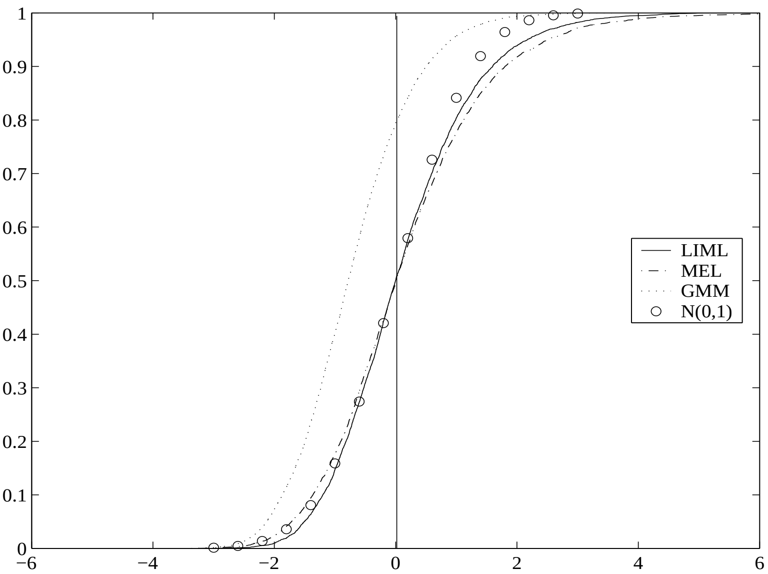

Figure 10A: Nonlinear case I: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50$

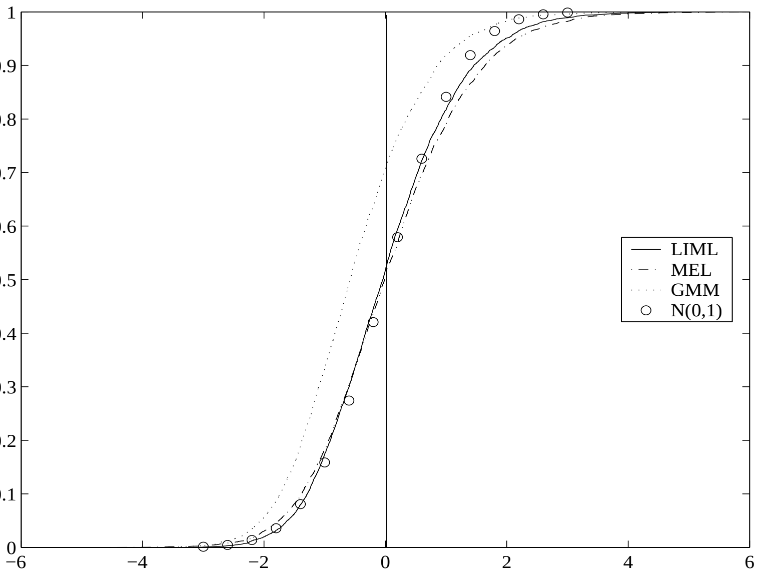

Figure 11A: Nonlinear case II: $n - K = 100, K_2 = 10, \alpha = 1, \delta^2 = 50$

	$n-K=300, K_2=30, \alpha=1, \delta^2=50$						
	LIML	MEL	GMM	$N(0,1)$	$N(0,h)$	A.exp	Difference
x					(large-K_2)	(large-K_2)	("A.exp"-LIML)
-3	0.000	0.002	0.042	0.001	0.005	0.000	0.000
-2.5	0.003	0.010	0.165	0.006	0.015	0.000	-0.003
-2	0.019	0.034	0.417	0.023	0.041	0.011	-0.008
-1.4	0.085	0.110	0.736	0.081	0.112	0.080	-0.005
-1	0.170	0.195	0.870	0.159	0.193	0.169	-0.001
-0.8	0.227	0.248	0.918	0.212	0.244	0.226	-0.001
-0.6	0.291	0.308	0.950	0.274	0.301	0.291	-0.001
-0.4	0.359	0.369	0.971	0.345	0.364	0.359	0.000
-0.2	0.430	0.432	0.983	0.421	0.431	0.430	0.000
0	0.500	0.494	0.991	0.500	0.500	0.500	0.000
0.2	0.567	0.556	0.996	0.579	0.569	0.568	0.001
0.4	0.630	0.616	0.998	0.655	0.636	0.631	0.000
0.6	0.687	0.670	1.000	0.726	0.699	0.688	0.001
0.8	0.739	0.718	1.000	0.788	0.756	0.739	0.000
1	0.783	0.756	1.000	0.841	0.807	0.783	0.000
1.4	0.852	0.822	1.000	0.919	0.888	0.855	0.003
2	0.920	0.895	1.000	0.977	0.959	0.928	0.008
2.5	0.953	0.931	1.000	0.994	0.985	0.964	0.011
3	0.972	0.953	1.000	0.999	0.995	0.985	0.013
X05	-1.63	-1.82	-2.95	-1.65	-1.90	-1.59	0.04
L.QT	-0.75	-0.79	-2.30	-0.67	-0.78	-0.72	0.03
MEDN	0.00	0.02	-1.85	0.00	0.00	0.00	0.00
U.QT	0.85	0.97	-1.37	0.67	0.78	0.85	0.00
X95	2.48	2.94	-0.60	1.65	1.90	2.27	-0.21
IQR	1.60	1.76	0.93	1.35	1.56	1.57	-0.03

See Section 4.2.

Table 1B:

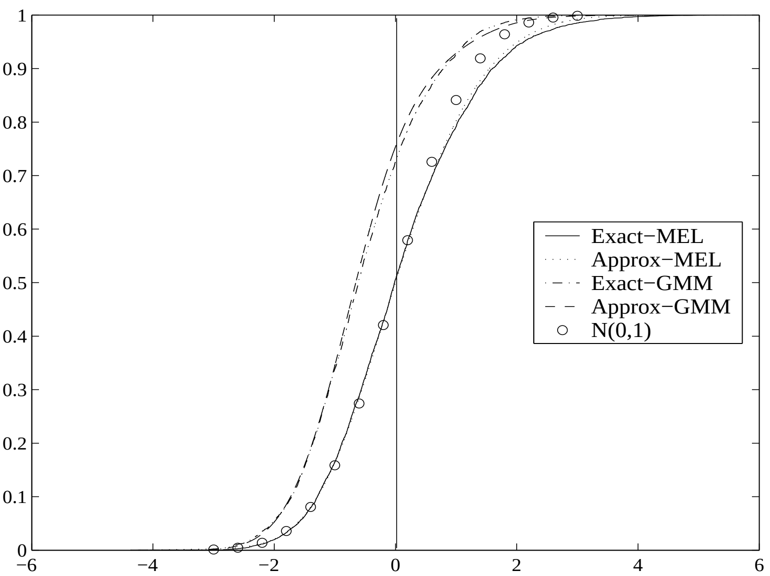

Figure 1B: $n-K=100, K_2=10, \alpha=1, \delta^2=100$

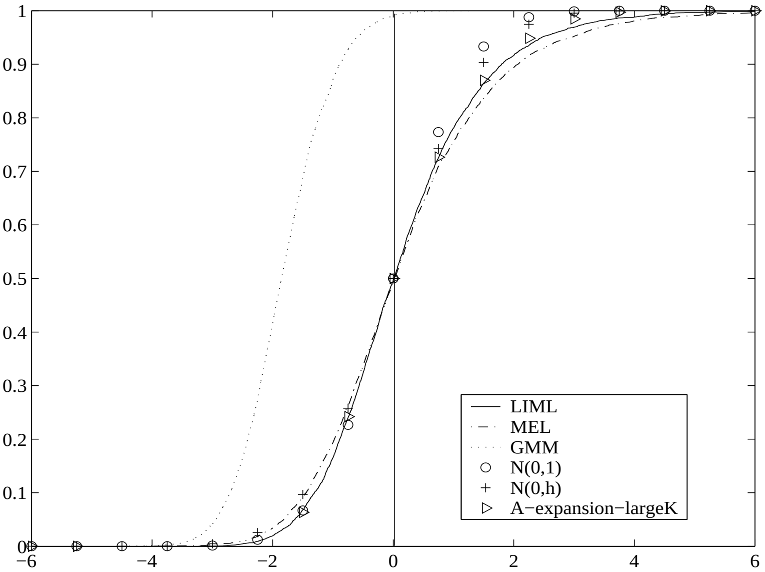

Figure 2B: $n - K = 300, K_2 = 30, \alpha = 1, \delta^2 = 50$

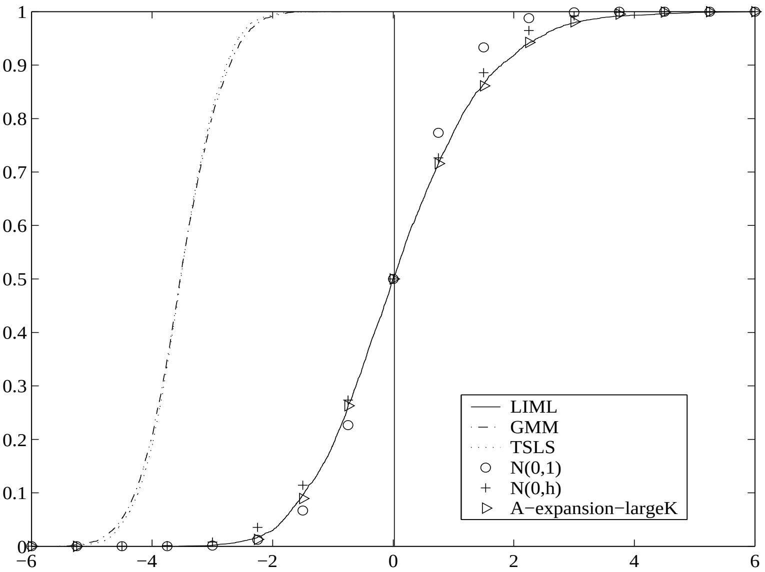

Figure 3B: $n - K = 1000, K_2 = 100, \alpha = 1, \delta^2 = 100$