Many Instruments, Heteroscedasticity and Asymptotic Optimality in the LIML Estimation [a]

Naoto Kunitomo

University of Tokyo

June 2008

at Stanford University

in Honor of Theodore W.Anderson's 90th Birthday

---

[a] This talk is based on several unpublished papers which are (or will be shortly) available at http://www.e.u-tokyo.ac.jp/cirje/research/dp. (In particular, Anderson, Kunitomo and Matsushita (2005, 2007), Kunitomo (2008), and Kunitomo and Matsushita (2008).)

Outline of Presentation

1. Introduction
2. **Asymptotic Optimality of LIML**:
   A result of Anderson-Kunitomo-Matsushita (2005, 2007)
3. Alternative Estimation Methods with Many Instruments and possibly **Persistent Heteroscedasticity**
4. Asymptotic and Finite Sample Properties of the **MLIML** Estimator
5. The **GLIML** estimation
6. An Application to **MRAAR** Test
7. Conclusions

1 Motivations of Study

1. In the traditional structural equations with instruments and the generalized method of moments (or the estimating equations) (GMM), the normal approximations for estimators and statistics based on the standard large sample theory are often not satisfactory when **the number of instruments** and/or the number of the estimating equations is **large**.
2. In the recent microeconomteric applications we sometimes have the cases when there are **many instruments** and/or **many weak instruments** including the instrumental variables methods for panels models. (Angrist and Krueger(1991), Bound et al. (1995) etc.)
3. In addition to the parametric estimation methods including the limited information maximum likelihood (**LIML**) method and the two stage least squares (**TSLS**) method, the semiparametric methods including the generalized method of moments (**GMM**) and the empirical likelihood method (**EL**) have been often discussed in recent econometrics.

Important Remarks

1. When we have **many instruments**, the finite sample distributions of TSLS (or GMM) can be quite different from the approximations based on the standard large sample theory.

2. Anderson-Kunitomo-Matsushita (2005, 2007) have developed the asymptotic theory when both the number of observations n and the number of excluded (from the structural equation of interest) instruments K_{2n} are large, i.e. the **large-K_2** asymptotic theory, and to obtain the **asymptotic optimality** results with **many instruments** under **homoscedasticity** (or **weak heteroscedasticity**) situation.

3. Some studies (Hausman et al. (2007)) argued that the LIML estimation has a serious problem when both there are many instruments and **persistent heteroscedasticity** at the same time. However, their method called **HLIM** could have a problem to be overcomed and there are still unsolved problems.

Main purposes of our study

1. We propose a simple modification of the **LIML** estimator (called the **MLIML** estimation) with **many instruments** and possibly **persistent heteroscedasticity** at the same time.

2. The MLIML estimator has often desirable (asymptotic) properties and it improves **both** the LIML estimator and the HLIM (or JLIML) estimator.

3. The MLIML estimator has often an **asymptotic optimality**.

4. We apply our approch to the testing problem and develop a new testing procedure called **MRAAR**, which is a modified **Rank-Adjusted Anderson-Rubin** procedure.

2 The structural equation model

For the *simplicity*, assume the structural equation of interest is linear for a while as

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

where K_1 and G_2 are fixed integers,

$(y_{1i}, \mathbf{y}_{2i}^{'})^{'}$: the vector of $1 + G_2$ endogenous variables (i.e. $\mathbf{E}[\mathbf{y}_{2i}u_i] \neq \mathbf{0}$),

$\mathbf{z}_{1i}$: the vector of K_1 included instrumental variables,

$\mathbf{z}_{2i}^{(n)}$: the vector of K_{2n} **excluded** instrumental variables,

$(\boldsymbol{\beta}_2^{'}, \boldsymbol{\gamma}_1^{'})^{'}$: the vector of $G_2 + K_1$ coefficients,

u_i: the disturbance term with $\mathbf{E}(u_i|\mathbf{z}_i^{(n)}) = 0$ and $\mathbf{E}(u_i^2|\mathbf{z}_i^{(n)}) = \sigma_i^2$ $(i = 1, \cdots, n)$,

- **Orthogonality conditions**: $\mathbf{E}[u_i\mathbf{z}_i^{(n)}] = \mathbf{0}$ $(i = 1, \cdots, n)$
  with $\mathbf{z}_i^{(n)} = (\mathbf{z}_{1i}^{'}, \mathbf{z}_{2i}^{(n)'})^{'}$: the vector of $K_n (= K_1 + K_{2n})$ instrumental variables

- **Identifying (order) condition** : $K_{2n} \geq G_2$

Reduced Form :

$$\mathbf{Y}_2 = \mathbf{\Pi}_2(Z) + \mathbf{V}_2 \ , \ \mathbf{y}_1^{(n)} = \mathbf{\Pi}_1(Z) + \mathbf{v}_1^{(n)}$$

where

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

$$\mathbf{\Pi}_2(Z) = (\boldsymbol{\pi}_2^{'}(\mathbf{z}_i^{(n)})) : \quad (K_1 + K_{2n}) \times G_2$$

for each row $\mathbf{v}_i^{'}$ of $\mathbf{V} = (\mathbf{v}_1^{(n)}, \mathbf{V_2})$, $\mathbf{E}(\mathbf{v}_i | \mathbf{z}_i^{(n)}) = 0$, $\mathbf{E}(\mathbf{v}_i \mathbf{v}_i^{'} | \mathbf{z}_i^{(n)}) = \mathbf{\Omega}_i$ and $\sigma_i^2 = \boldsymbol{\beta}^{'} \mathbf{\Omega}_i \boldsymbol{\beta} > 0 \ (i = 1, \cdots, n)$.
In the linear case, the reduced form and the structural equation are

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V} \ , \ \mathbf{Y}\begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \mathbf{Z}_1 \boldsymbol{\gamma}_1 + \mathbf{u}$$

Estimation Methods considered

- **LIML** : the limited information maximum likelihood method developed by Anderson-Rubin (1949, 1950).
- **GMM** (the generalized method of moments : Hansen (1982), the estimating equation method) or **TSLS** (the two-stage least squares method)
- **HLIM** (or JLIML) : Jackknife-type method developed by Hausman et al. (2007).
- **MLIML** and **GLIML** : Modified LIML methods proposed by Kunitomo (2008) and, Kunitomo and Matsushita (2008).

Let $(1+G_2) \times (1+G_2)$ random matrices be

$$\mathbf{G} = \mathbf{Y}^{'}\mathbf{Z}_{2.1}^{(n)}\mathbf{A}_{22.1}^{(n)-1}\mathbf{Z}_{2.1}^{(n)'}\mathbf{Y}$$

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

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

The **LIML** estimator $\hat{\boldsymbol{\beta}}_{LI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.LI}^{'})^{'})$ for $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ is given by

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

where λ is the smallest root of $|\mathbf{G} - l\mathbf{H}| = 0$. By replacing λ by 0, we have the **TSLS** estimator $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ as

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

Asymptotic Optimality of LIML

The basic conditions of Anderson et al. (2005, 2007) are

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

$$(\mathbf{A}-\mathbf{II}') \qquad \frac{1}{d_n^2}\Pi_2^{(n)}(Z)'\mathbf{Z}_{2.1}^{(n)}\mathbf{A}_{22.1}^{(n)-1}\mathbf{Z}_{2.1}^{(n)'}\Pi_2^{(n)}(Z) \xrightarrow{p} \Phi_{22.1}$$

as $d_n \xrightarrow{p} \infty \ (n \to \infty)$, where $\Phi_{22.1}$ is a nonsingular constant matrix and $\mathcal{E}[\|\mathbf{v}_i\|^6]$ are bounded. (In the standard case $d_n^2 = n$.)

A Result of Anderson-Kunitomo-Matsushita (2005, 2007) : Define the class of consistent estimators for $\boldsymbol{\beta}_2$ by $\hat{\boldsymbol{\beta}}_2 = \phi(\frac{1}{n}\mathbf{G}, \frac{1}{q_n}\mathbf{H})$, ϕ is continuously differentiable, its derivatives are bounded at the probability limits as $K_{2n} \to \infty \ (n \to \infty)$ and $0 \le c < 1$.

(i) Then under a set of assumptions,

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

and $\Psi \geq \Psi_A^*$, where $\Psi_A^* = \Phi_{22.1}^{-1} \left[\Psi_1 + \Psi_2(c)\right] \Phi_{22.1}^{-1}$ and

$$\Psi_1 = \sigma^2 \text{plim} \frac{1}{n} \sum_{i,j=1}^{n} \boldsymbol{\pi}_{2i}(\mathbf{z}_i^{(n)}) p_{ij}^{(n)} \boldsymbol{\pi}_{2j}(\mathbf{z}_j^{(n)})^{'} ,$$

$$\Psi_2(c) = \sigma^2 \text{plim} \frac{1}{n} \sum_{i,j=1}^{n} \mathbf{E}(\mathbf{w}_{2j} \mathbf{w}_{2j}^{'} | \mathbf{z}_i^{(n)}) [p_{ij}^{(n)} - c_* q_{ij}^{(n)}]^2 ,$$

$\mathbf{P}_n = (p_{ij}^{(n)}) = \mathbf{Z}_{2.1}^{(n)} \mathbf{A}_{22.1}^{(n)-1} \mathbf{Z}_{2.1}^{(n)'}$, $\mathbf{Q}_n = (q_{ij}^{(n)}) = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}$,
$\mathbf{w}_{2i} = \mathbf{v}_{2i} - u_i(\mathbf{0}, \mathbf{I}_{G_2}) \boldsymbol{\Omega} \boldsymbol{\beta} / \sigma^2$ and $c_* = c/(1-c)$.
(ii) The LIML estimator attains the **asymptotic lower bound** Ψ^* for any $0 \leq c < 1$ when the condition **(A-VI)** holds.

Remark : When $c = 0$, $\Psi_2(c) = \mathbf{O}$ as in the standard large sample theory. Under the normality and the homoscedasticity with $c = 0$, Ψ^* is the (Fisher) information bound. The **asymptotic normality** of LIML holds with the moment condition.

Modifications of LIML

One of **key conditions** used by Anderson et al. (2005, 2007) is

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

where $p_{ii}^{(n)} = (\mathbf{Z}_{2.1}^{(n)}\mathbf{A}_{22.1}^{(n)-1}\mathbf{Z}_{2.1}^{(n)'})_{ii}\ (i=1,\cdots,n)$.
If $(1/n)\sum_{i=1}^{n}\mathbf{\Omega}_i \xrightarrow{p} \mathbf{\Omega}$, then we have the **Weak Heteroscedasticity** condition :

$$(\mathbf{WH}) \qquad \operatorname*{plim}_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n}p_{ii}^{(n)}\mathbf{\Omega}_i - c\mathbf{\Omega}\right] = \mathbf{O}$$

We say **Persistent Heteroscedasticity** (PH) if (WH) is not satisfied.

Remark : There may be a question whether (WH) or (PH) would be relevant in many applications.

Lemma 1 : Let $\mathbf{P}_n = (p_{ij}^{(n)}) = \mathbf{Z}_{2.1}^{(n)} \mathbf{A}_{22.1}^{(n)-1} \mathbf{Z}_{2.1}^{(n)'}$ and $\mathbf{Q}_n = (q_{ij}^{(n)}) = \mathbf{I}_n - \mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}^{'}$. Then $0 \leq p_{ii}^{(n)} < 1 \ (i = 1, \cdots, n)$ and $0 < q_{ii}^{(n)} \leq 1 \ (i = 1, \cdots, n)$. Condition (A-I) implies

$$\bar{p}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} p_{ii}^{(n)} = \frac{K_{2n}}{n} \longrightarrow c \ ,$$

$$\bar{q}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} q_{ii}^{(n)} = 1 - \frac{K_n}{n} \longrightarrow 1 - c \ ,$$

where $c_n = K_{2n}/n$.

Remark : When K_{2n} is a fixed number (the standard asymptotic theory), $c = 0$ automatically.

In order to remove **Persistent Heteroscedasticity**, we propose a modification of the LIML estimator (we may call **MLIML**) such that $\hat{\boldsymbol{\beta}}_{MLI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.MLI}^{'})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2^{'})^{'}$ is the solution of

$$(\frac{1}{n}\mathbf{G}_M - \frac{1}{q_n}\lambda_n \mathbf{H}_M)\hat{\boldsymbol{\beta}}_{MLI} = \mathbf{0}$$

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

$$|\frac{1}{n}\mathbf{G}_M - l\frac{1}{q_n}\mathbf{H}_M| = 0 \ ,$$

where $\mathbf{G}_M = \mathbf{Y}^{'}\mathbf{P}_n^{*}\mathbf{Y}$, $\mathbf{H}_M = \mathbf{Y}^{'}\mathbf{Q}_n^{*}\mathbf{Y}$, and $\mathbf{P}_n^{*} = (p_{ij}^{*})$ and $\mathbf{Q}_n^{*} = (q_{ij}^{*})$ are defined by $p_{ij}^{*} = p_{ij}^{(n)}$ $(i \neq j); p_{ii}^{*} = K_{2n}/n$ and $q_{ij}^{*} = q_{ij}^{(n)}$ $(i \neq j)$; $q_{ii}^{*} = 1 - K_n/n(i, j = 1, \cdots, n)$.

Remark : More generally, we can take p_{ii}^{*} and q_{ii}^{*} $(i = 1, \cdots)$ such that they satisfy $\bar{p}^{*} \sim K_{2n}/n$, $\bar{q}^{*} \sim 1 - K_n/n$ and (A-VI). The **JLIML** or (HLIM) estimation $(\hat{\boldsymbol{\beta}}_{HLI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.HLI}^{'})^{'}))$ removes the diagonal elements such that $p_{ii}^{*} = q_{ii}^{*} = 0$ $(i = 1, \cdots, n)$ and it is not possible to do this.

3 Asymptotic Properties of MLIML

Theorem 1: Let $\mathbf{z}_i^{(n)} (i = 1, 2, \cdots, n)$ be a set of $K_n \times 1$ vectors $(K_n = K_1 + K_{2n})$. Let $\mathbf{v}_i$ satisfy $\mathcal{E}(\mathbf{v}_i | \mathbf{z}_i^{(n)}) = \mathbf{0}$, $\mathcal{E}(\mathbf{v}_i \mathbf{v}_i^{'} | \mathbf{z}_i^{(n)}) = \boldsymbol{\Omega}_i$ and $\mathcal{E}[\|\mathbf{v}_i\|^6]$ are bounded. Suppose Conditions (A-I), $(\mathrm{A} - \mathrm{II}^{'})$ with $d_n = n^{1/2}$ and

$$\frac{1}{n} \Pi_2^{(n)}(Z)^{'} \mathbf{P}_n^* \Pi_2^{(n)}(Z) \xrightarrow{p} \Phi_{22.1}^* > 0 \ .$$

Then

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

where

$$\Psi^* = \Phi_{22.1}^{*-1} \left[\Psi_1^* + \Psi_2^*(c) \right] \Phi_{22.1}^{*-1} \ ,$$

$$\Psi_1^* = \text{plim}\frac{1}{n}\sum_{i,j,k=1}^{n} \boldsymbol{\pi}_{2i}(\mathbf{z}_i^{(n)}) p_{ij}^* p_{jk}^* \sigma_j^2 \boldsymbol{\pi}_{2k}(\mathbf{z}_k^{(n)})^{'} ,$$

$$\Psi_2^*(c) = \text{plim}\frac{1}{n}\sum_{i,j=1}^{n} \left[\sigma_i^2 \mathbf{E}(\mathbf{w}_{2j}\mathbf{w}_{2j}^{'}|\mathbf{z}_i^{(n)}) + \mathbf{E}(\mathbf{w}_{2j}u_j|\mathbf{z}_i^{(n)})\mathbf{E}(\mathbf{w}_{2j}^{'}u_j|\mathbf{z}_i^{(n)})\right] \times \left[p_{ij}^* - c_* q_{ij}^*\right]^2 ,$$

provided that Ψ_1^* and $\Psi_2^*(c)$ converge in probability as $n \to \infty$, $c_* = c/(1-c)$ and $\mathbf{w}_{2i} = \mathbf{v}_{2i} - u_i(\mathbf{0}, \mathbf{I}_{G_2})\boldsymbol{\Omega}\boldsymbol{\beta}/\sigma^2$ $(i = 1, \cdots, n)$.

Corollary 1: Assume

$$\Phi_{22.1H} = \text{plim} \frac{1}{n} \sum_{i,j=1, i \neq j}^{n} \boldsymbol{\pi}_{2i}(\mathbf{z}_i^{(n)}) p_{ij}^{(n)} \boldsymbol{\pi}_{2j}(\mathbf{z}_j^{(n)})'$$

is nonsingular. Then under the assumptions of Theorem 1,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.HLI} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \Psi_H) ,$$

where (the asymptotic covariances of HLIM)

$$\Psi_H = \Phi_{22.1H}^{-1} \left[\Psi_{1H} + \Psi_2^*(c) \right] \Phi_{22.1H}^{-1}$$

and

$$\Psi_{1H} = \text{plim} \frac{1}{n} \sum_{i,j,k=1, i \neq j, j \neq k}^{n} \boldsymbol{\pi}_{2i}(\mathbf{z}_i^{(n)}) p_{ij}^{(n)} p_{jk}^{(n)} \sigma_j^2 \boldsymbol{\pi}_{2k}(\mathbf{z}_k^{(n)})' .$$

For the weak instruments case, we have an inequality

$$\Phi_{22.1H}^{-1} \Psi_2^*(c) \Phi_{22.1H}^{-1} > \Phi_{22.1}^{*-1} \Psi_2^*(c) \Phi_{22.1}^{*-1}$$

Corollary 2: Assume $\max_{1 \leq i \leq n} |\sigma_i^2 - \sigma^2| \xrightarrow{p} 0$ (we may write $(\mathrm{WH})'$) and $\Phi_{22.1H}$ is nonsingular. Then

$$\Phi_{22.1H}^{-1} \Psi_{1H} \Phi_{22.1H}^{-1} > \Phi_{22.1}^{*-1} \Psi_1^* \Phi_{22.1}^{*-1}$$

and $\Psi_H > \Psi^*$.

Theorem 2: Define the class of consistent estimators for $\boldsymbol{\beta}_2$ by

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

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

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

and $\Psi \geq \Psi^*$, where

$$\Psi^* = \Phi_{22.1}^{-1} \left[\Psi_1^* + \Psi_2^*(c)\right] \Phi_{22.1}^{-1} \ .$$

(ii) The MLIML estimator attains the lower bound Ψ^*.

4 Some finite sample properties

Some Figures of Alternative Estimators

CDF of LIML

CDF of JLIML (or HLIM)

CDF of MLIML (or KLIML)

o stands for CDF of N(0,1)

Key Parameters : $(G_2 = 1)$ $n - K$, K_2, δ^2 (the noncentrality parameter),
$\alpha = \sqrt{\frac{\omega_{22}}{\omega_{11.2}}}(\beta_2 - \frac{\omega_{12}}{\omega_{22}})$ (the normalized coefficient)

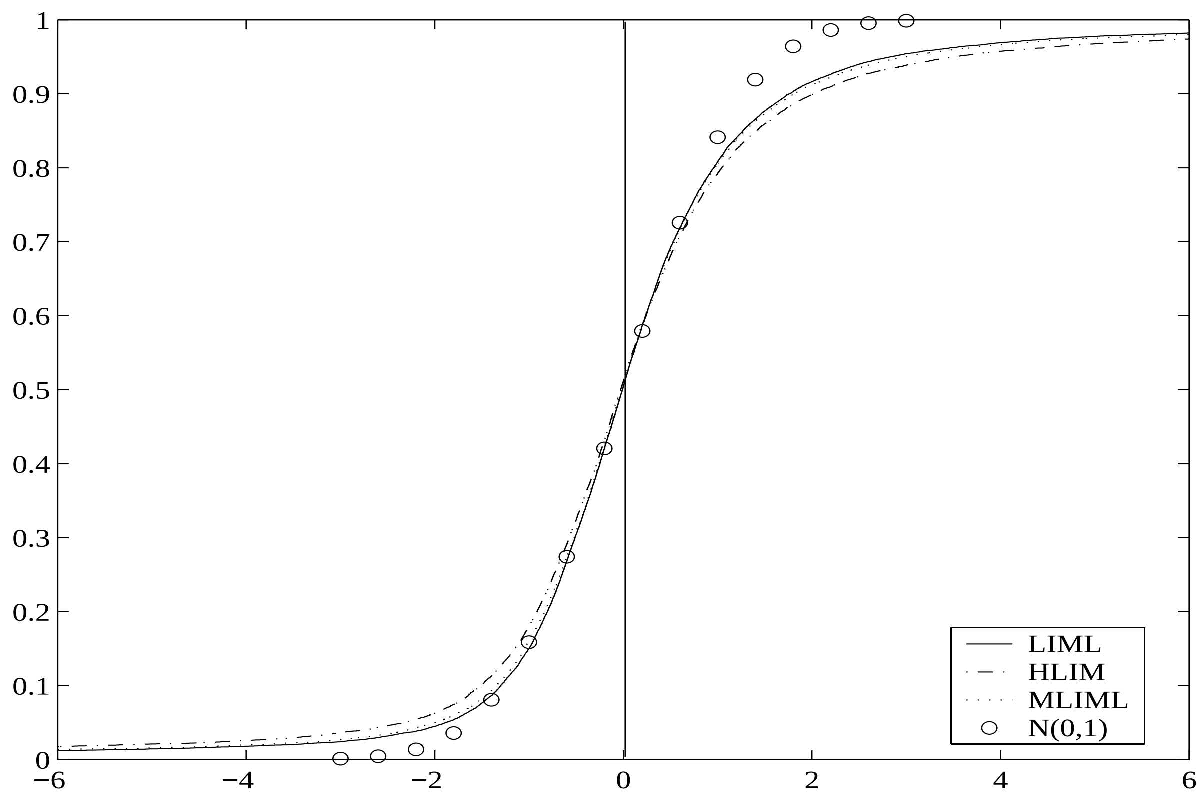

Figure 1: CDF of Standardized estimators: $n - K = 20, K_2 = 30, \alpha = 0.5, \delta^2 = 30, u_i = N(0,1)$

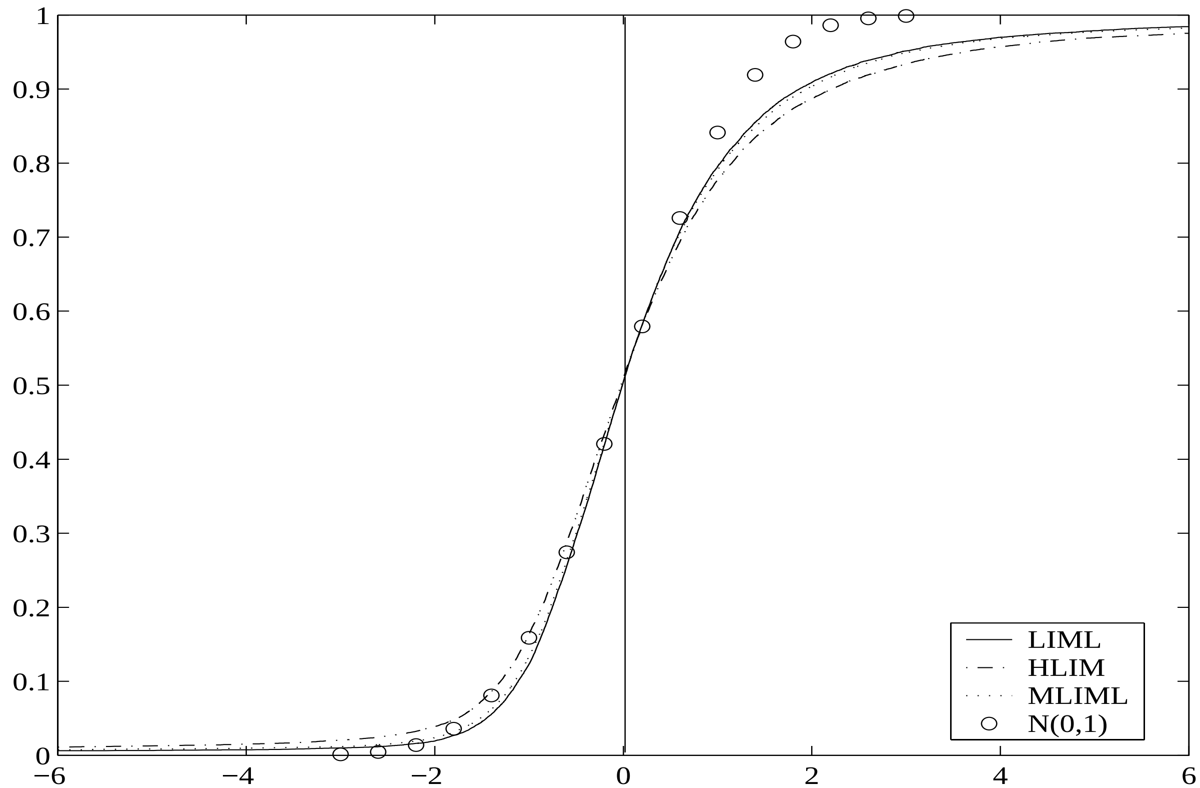

Figure 2: CDF of Standardized estimators: $n - K = 20, K_2 = 30, \alpha = 1, \delta^2 = 30, u_i = N(0,1)$

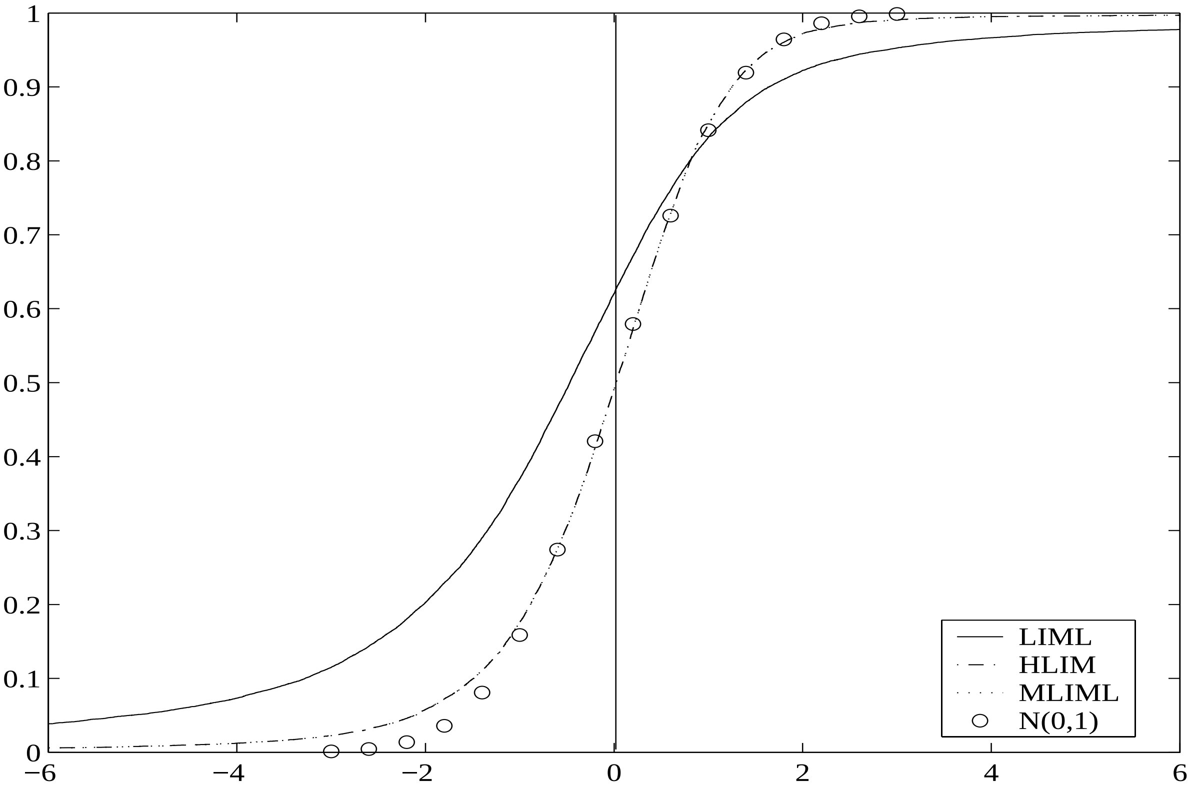

Figure 3: CDF of Standardized estimators: Heteroscedastic disturbances in Hausman et.al (2007), $n = 100, K = 10, \delta^2 = 30$

5 GLIML

For the simplicity, consider the case when $\sigma_i^2 = \sigma^2(\mathbf{z}_i^{(n)})$ depend on a finite subset of $\mathbf{z}_i^{(n)}$for each i (say, the dimension is a fixed d) and assume (WH).

Let

$$\tilde{\mathbf{P}}(Z) = \mathbf{Z}\left[\sum_{i=1}^{n}\hat{\sigma}_i^2(\beta_{-1})\mathbf{z}_i^{(n)}\mathbf{z}_i^{(n)'}\right]^{-1}\mathbf{Z}'$$

and

$$\hat{\sigma}_i^2(\beta_{-1}) = \sum_{j=1}^{k(n)} w_{ij}\hat{u}_{j(i)}^2(\beta_{-1})\ ,$$

where the indices $j(i)$ are chosen such that $\sum_j w_i j = 1, w_{ii} = 0$, $\rho(\sum_{k=1}^{d}|z_{ik} - z_{jk}|^2/s_k^2)$ are in the nearest neighborhood (NN) method. (See Robinson (1987) for the linear regression case.) Then we can construct the GLIML estimator $\hat{\boldsymbol{\beta}}_{2.GLI}$ by the LIML method.

6 A MRAAR Test

Anderson-Rubin (AR) test (Anderson and Rubin (1949, 1950)) is to reject $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ if

$$\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}\boldsymbol{\beta}_0} > \frac{K_{2n}}{q_n} F_{K_{2n},q_n}(\epsilon) \ ,$$

where $F_{K_{2n},q_n}(\epsilon)$ denotes the $1-\epsilon$ significance point of the F-distribution with K_{2n} and q_n degrees of freedom.

RAAR test (Anderson and Kunitomo (2007)) is that the null hypothesis that H_0 is rejected if

$$\frac{1 + \dfrac{\hat{\boldsymbol{\beta}}^{'}\mathbf{G}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\mathbf{H}\hat{\boldsymbol{\beta}}}}{1 + \dfrac{\boldsymbol{\beta}_0^{'}\mathbf{G}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}\boldsymbol{\beta}_0}} < c(K_{2n}, q_n) \ ,$$

where $q_n = n - K_n$. Moreira (2003) arrived at a similar statistic (when $\boldsymbol{\Omega}$ is known), which is a *conditional likelihood statistic*.

MRAAR test (Kunitomo and Matsushita (2008)) is to reject $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ if

$$\mathrm{MRAAR} = (-1)(n - K_n) \log \left[\frac{1 + \dfrac{\hat{\boldsymbol{\beta}}^{'} \mathbf{G}_M \hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'} \mathbf{H}_M \hat{\boldsymbol{\beta}}}}{1 + \dfrac{\boldsymbol{\beta}_0^{'} \mathbf{G}_M \boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'} \mathbf{H}_M \boldsymbol{\beta}_0}} \right] > c^*(K_{2n}, q_n) \ ,$$

where $c^*(K_{2n}, q_n)$ is a constant.

Theorem 1 (Kunitomo and Matsushita (2008)) : Let $\mathbf{z}_i^{(n)} (i = 1, 2, \cdots, n)$ be a set of $K_n \times 1$ vectors ($K_n = K_1 + K_{2n}$). Let $\mathbf{v}_i$ satisfy $\mathcal{E}(\mathbf{v}_i | \mathbf{z}_i^{(n)}) = \mathbf{0}$, $\mathcal{E}(\mathbf{v}_i \mathbf{v}_i^{'} | \mathbf{z}_i^{(n)}) = \boldsymbol{\Omega}_i$ and $\mathcal{E}[\|\mathbf{v}_i\|^6]$ are bounded. Suppose Conditions (A-I) and

$$\frac{1}{n} \Pi_2(Z)^{(n)'} \mathbf{P}_n^* \Pi_2^{(n)}(Z) \xrightarrow{p} \Phi_{22.1}^* > 0 \ .$$

Then

$$\mathrm{MRAAR}_n - \mathrm{MRAAR}_{1n}^* \xrightarrow{p} 0 \ ,$$

and

$$
\begin{aligned}
& \mathrm{MRAAR}^*_{1n} \\
= \quad & \frac{1}{\sigma_0^2}\mathbf{u}^{'}\left[\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)+(\mathbf{P}^*_n-c_*\mathbf{Q}^*_n)\mathbf{W}_2\right]\left[\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)\right]^{-1} \\
& \times\left[\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n+\mathbf{W}_2^{'}(\mathbf{P}^*_n-c_*\mathbf{Q}^*_n)\right]\mathbf{u}\ ,
\end{aligned}
$$

where $\sigma_0^2=\boldsymbol{\beta}_0^{'}\mathbf{\Omega}\boldsymbol{\beta}_0$ and $\mathbf{W}_2=(\mathbf{w}_{2i}^{'})$ with $\mathbf{w}_{2i}=\mathbf{v}_{2i}-u_i(\mathbf{0},\mathbf{I}_{G_2})\mathbf{\Omega}\boldsymbol{\beta}_0/\sigma_0^2$. Furthermore, MRAAR^*_{1n} is asymptotically equivalent to $\mathrm{MRAAR}^*_n=\mathbf{\Lambda}_{1n}+\mathbf{\Lambda}_{2n}(c)+2\mathbf{\Lambda}_{3n}(c)$, where

$$
\begin{aligned}
\mathbf{\Lambda}_{1n} &= \frac{1}{\sigma_0^2}\mathbf{u}^{'}\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)\left[\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)\right]^{-1}\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n\mathbf{u}\ , \\
\mathbf{\Lambda}_{2n}(c) &= \frac{1}{\sigma_0^2}\mathbf{u}^{'}(\mathbf{P}^*_n-c_*\mathbf{Q}^*_n)\mathbf{W}_2\left[\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)\right]^{-1} \\
&\quad \times\mathbf{W}_2^{'}(\mathbf{P}^*_n-c_*\mathbf{Q}^*_n)\mathbf{u}\ , \\
\mathbf{\Lambda}_{3n}(c) &= \frac{1}{\sigma_0^2}\mathbf{u}^{'}(\mathbf{P}^*_n-c_*\mathbf{Q}^*_n)\mathbf{W}_2\left[\mathbf{\Pi}_2^{(n)}(Z)^{'}\mathbf{P}^*_n\mathbf{\Pi}_2^{(n)}(Z)\right]^{-1}\mathbf{\Pi}_2^{(n)'}\mathbf{P}^*_n\mathbf{u}\ .
\end{aligned}
$$

If $c=0$, then $\mathbf{\Lambda}_{2n}(c)=o_p(1)$ and $\mathbf{\Lambda}_{3n}(c)=o_p(1)$.

Corollary 1 (Kunitomo and Matsushita (2008)) : Assume the condition (A-VI). Then the limiting distribution of MRAAR_n under $\mathbf{H}_0$ is χ^2 with G_2 degrees of freedom if and only if $c = 0$.

Remark 1 : There is a simple way to calculate the null distribution as a weighted χ^2-type distribution.

Remark 2 : The significance levels of MRAAR calculated by the weighted χ^2- type limiting distribution are more accurate than most of existing tests when there are many instruments and persistent heteroscedasticity at the same time.

7 Concluding Remarks

1. We propose a particular modification of the **LIML** estimation method called **MLIML**, which is similar to the original LIML (**Anderson-Rubin (1949, 1950)**) or the HLIM estimators in some situations.
2. With many instruments and persist heteroscedasticities, the MLIML estimator has good asymptotic properties in the **large-**K_2 asymptotic theory. Under certain conditions it often attains the efficiency bound asymptotically.
3. We can apply our method to the testing problem and develop **MRAAR**, (which is a modification of *Rank-Adjusted Anderson-Rubin* test).
4. The finite sample behaviors of the MLIML, the GLIML estimators, their nonlinear extensions and the MRAAR test are currently under investigation. The question whether **(WH)** or **(PH)** would be relevant is also currently under investigation.

Some references

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.

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.

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

Anderson, T.W., N. Kunitomo, and Y. Matsushita (2008) ”On Finite Sample Properties of Alternative Estimators with Many Instruments,” Unpublished Manuscript.

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).

Hansen, C. J. Hausman, and Newey, W. K. (2006), "Many Instruments, Weak Instruments, and Microeconometric Practice," Unpublished Manuscript.

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

N. Kunitomo (2008) "Improving the LIML Estimation with Many Instruments and Persistent Heteroscedasticity," Discussion Paper CIRJE-F-??, Graduate School of Economics, University of Tokyo.

N. Kunitomo and Matshushita (2008) "Improving the Rank-Adjusted Anderson-Rubin Test with Many Instruments and Heteroscedasticity," Discussion Paper CIRJE-F-??, Graduate School of Economics, University of Tokyo.