CIRJE-F-576

# An Optimal Modification of the LIML Estimation for Many Instruments and Persistent Heteroscedasticity

Naoto Kunitomo
University of Tokyo

July 2008; Revised in May 2009

# An Optimal Modification of the LIML Estimation for Many Instruments and Persistent Heteroscedasticity *

Naoto Kunitomo †

April 15, 2009

**Abstract**

We consider the estimation of coefficients of a structural equation with many instrumental variables in a simultaneous equation system. It is mathematically equivalent to an *estimating equation estimation* or a *reduced rank regression* in the statistical linear models when the number of restrictions or the dimension increases with the sample size. As a semi-parametric method, we propose a class of modifications of the limited information maximum likelihood (LIML) estimator to improve its asymptotic properties as well as the small sample properties for many instruments and persistent heteroscedasticity. We show that an asymptotically optimal modification of the LIML estimator, which is called AOM-LIML, improves the LIML estimator and other estimation methods. We give a set of sufficient conditions for an asymptotic optimality when the number of instruments or the dimension is large with persistent heteroscedasticity including a case of *many weak instruments*.

**Key Words**

Estimation of Structural Equation, Estimating Equation, Reduced Rank Regression, Many Instruments, Persistent Heteroscedasticity, Asymptotic Optimality.

---

*K09-4-15. This is a revised version of Discussion Paper CIRJE-F-576, "Improving the LIML Estimation with Many Instruments and Persistent Heteroscedasticity," (Graduate School of Economics, University of Tokyo, July 2008). The paper was prepared and presented at the conference in honor of Theodore W.Anderson's 90th Birthday held at Stanford University. I thank Yukitoshi Matsushita for useful discussions and providing some figures.

†Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1 Tokyo, JAPAN, kunitomo@e.u-tokyo.ac.jp

## 1. Introduction

In recent analysis of micro-econometric data many explanatory or instrumental variables are sometimes used in estimating an important structural equation. Then there have been increasing interest and research on the estimation of a structural equation in a system of simultaneous equations when the number of instruments (the number of exogenous variables excluded from the structural equation), say $K_2$, is large relative to the sample size, say $n$. Asymptotic distributions of estimators and test criteria have been investigated on the basis when both $K_2 \to \infty$ and $n \to \infty$. These asymptotic distributions are used as approximations to the distributions of the estimators and criteria when $K_2$ and $n$ are large. The early studies on the case of many instruments, which we call *the large-$K_2$ asymptotic theory* or *the many instruments asymptotics*, are Kunitomo (1980, 1981, 1982, 1987), Morimune (1983) and Bekker (1994). Several semi-parametric estimation methods have been developed including the estimating equation method (or the generalized method of moments (GMM) in econometrics) and the maximum empirical likelihood (MEL) method (see Hayashi (2000), Qin and Lawless (1994) and Owen (2001)). However, it has been recently recognized in econometrics that the classical Limited Information Maximum Likelihood (LIML) estimation, originally developed by Anderson and Rubin (1949, 1950), has some advantage with many instruments in micro-econometric applications. (The LIML estimation can be regarded as a simplified version of the MEL estimation.) There has been a growing literature in econometrics on the problem of many instruments including Chao and Swanson (2005), Anderson, Kunitomo and Matsushita (2005, 2007), Hansen, Hausman and Newey (2008) and their references. This problem is mathematically equivalent to *an estimating equation estimation* or *a reduced rank regression* with the statistical linear models when the number of restrictions or the dimension increases with the sample size.

For sufficiently large sample sizes the LIML estimator and the Two-Stage Least Squares (TSLS) estimator have approximately the same distribution in the standard large-sample asymptotic theory, but their exact distributions can be quite different for the sample size occurring in practice with many instruments. Anderson et al. (2007) have shown that the LIML estimator has an asymptotic optimum property when $K_2$ and $n$ are large under a set of conditions. On the other hand, the JIVE (Jackknife Instrumental Variables Estimation) method has been proposed and its properties has been investigated. (See Angrist, Imbens and Krueger (1999), Chao and Swanson (2004), for instance.) Also Hausman, Newey, Wountersen, Chao and Swanson (2007) proposed the jackknife version of the LIML estimator (called JMIML or HLIM) and the Fuller modification. They suggested that the JLIML estimator improves the bias property of the LIML estimator in case of *the persistent heteroscedasticity*, which we shall define precisely.

The main purpose of this paper is to propose an asymptotically optimal modification of the LIML estimator, which we shall call *AOM-LIML* as an abbreviation. We show that the AOM-LIML estimator improves some properties of the LIML estimator and its possible modifications including the JIVE (Jackknife Instrumental Variables Estimators), the JLIML estimator. The AOM-LIML estimator has good asymptotic properties and it often attains the lower bound of the asymptotic variance in a class of estimators when the disturbances are heteroscedastic and there are *many instruments* or *many weak instruments*. We relate the AOM-LIML estimator to other estimations methods known and show that the JLIML estimator is asymptotically equivalent to the AOM-LIML estimator. The results of this paper lead to a new light on the asymptotic efficiency when there are many incidental parameters (i.e. the number of instruments is large) and the disturbances have persistent heteroscedasticity.

In Section 2 we state the structural equation model and the alternative estimation methods of unknown parameters in simultaneous equation models with possibly many instruments. Then in Section 3 we develop a new way of improving the LIML estimation and discuss a set of sufficient conditions for the asymptotic normality and the asymptotic lower bound when the number of instruments is large with the persistent heteroscedasticity. We shall give a small number of numerical evidence on the finite sample properties of the LIML, the AOM-LIML and JLIML estimators.

when there are many incidental parameters. Finally, some brief concluding remarks will be given in Section 4. The proof of our theorems will be given in Section 5. For an illustration of our results in Section 3.3, we shall give some figures in Appendix.

## 2. Alternative Estimation Methods of A Structural Equation with Many Instruments

Let a single linear structural equation be

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

where $y_{1i}$ and $\boldsymbol{y}_{2i}$ are a scalar and a vector of $G_2$ endogenous variables, respectively ($K_1$ and $G_2$ are fixed integers); $\boldsymbol{z}_{1i}$ is a vector of $K_1$ (included) exogenous variables, $\boldsymbol{\gamma}_1$ 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 $\mathcal{E}(u_i|\mathbf{z}_i^{(n)}) = 0$ and $\mathcal{E}(u_i^2|\mathbf{z}_i^{(n)}) = \sigma_i^2$ with the $K_n \times 1$ instrumental variables $\mathbf{z}_i^{(n)}$ $(i = 1, \cdots, n)$. We assume that (2.1) is the structural equation in a system of $1 + G_2$ endogenous variables $\boldsymbol{y}_i' = (y_{1i}, \boldsymbol{y}_{2i}')'$ and $\mathbf{Y} = (\mathbf{y}_1^{(n)}, \mathbf{Y}_2^{(n)})$ is an $n \times (1 + G_2)$ vector of their observations. As a typical situation we consider

$$\mathbf{Y}_2^{(n)} = \mathbf{\Pi}_{2n}^{(z)} + \mathbf{V}_2^{(n)} \ , \tag{2.2}$$

where $\mathbf{\Pi}_{2n}^{(z)} = (\boldsymbol{\pi}_{2i}'(\mathbf{z}_i^{(n)}))$ is an $n \times G_2$ matrix, each row $\boldsymbol{\pi}_{2i}'(\mathbf{z}_i^{(n)})$ depends on $K_n \times 1$ vector $\mathbf{z}_i^{(n)}$, $\mathbf{V}_2^{(n)}$ is an $n \times G_2$ matrix, $\mathbf{v}_1^{(n)} = \mathbf{u} + \mathbf{V}_2^{(n)} \boldsymbol{\beta}_2$ and $\mathbf{V} = (\mathbf{v}_1^{(n)}, \mathbf{V}_2^{(n)})$. $\mathbf{V} = (\mathbf{v}_i')$ is an $n \times (1+G_2)$ matrix of disturbances (the i-th row $\mathbf{v}_i'$ is a $1 \times (1+G_2) \times 1$ vector) with $\mathcal{E}(\mathbf{v}_i|\mathbf{z}_i^{(n)}) = \mathbf{0}$ and

$$\mathcal{E}(\mathbf{v}_i \mathbf{v}_i' | \mathbf{z}_i^{(n)}) = \mathbf{\Omega}_i = \begin{bmatrix} \omega_{11.i} & \boldsymbol{\omega}_{2.i}' \\ \boldsymbol{\omega}_{2.i} & \mathbf{\Omega}_{22.i} \end{bmatrix} . \tag{2.3}$$

The formulation of (2.1) and (2.2) includes the statistical linear models as special cases. We write

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi}_n + \mathbf{V} \ , \tag{2.4}$$

$\mathbf{\Pi}_n$ is a $(1 + G_2) \times K_n$ matrix of coefficients and the $n \times K_n$ matrix $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_{2n}) = (\mathbf{z}_i^{(n)'})$ (the i-th row $\mathbf{z}_i^{(n)'} = (\mathbf{z}_{1i}', \mathbf{z}_{2i}^{(n)'})$ is the vector of $K_n$ $(= K_1 + K_{2n})$ instruments).

When $\boldsymbol{\gamma}_1 = \mathbf{0}$, the rank of $\mathbf{\Pi}_n$ in (2.3) is $G_2$ and it is a reduced rank regression model. See Anderson (1984) for the classical arguments on the relations among statistical models with *different names* including the linear functional relationships, the simultaneous equations models, the errors-in-variables models and factor models.

Since we assume that the vector of $K_n$ ($K_n = K_1 + K_{2n}$) instruments $\mathbf{z}_i^{(n)}$ satisfy the orthogonal condition

$$\mathcal{E}[u_i \boldsymbol{z}_i^{(n)}] = \mathbf{0}\ (i = 1, \cdots, n)\ , \tag{2.5}$$

the model of (2.1) and (2.2) is the same as *an estimation equation* problem well-known in statistics, but we shall mainly investigate the situation when the number of orthogonal conditions ($K_n$) increases with the sample size $n$. This situation has been called the case of *many instruments* in recent econometrics. The relation between (2.1) and (2.2) gives $u_i = (1, -\boldsymbol{\beta}_2^{'})\mathbf{v}_i$ and

$$\sigma_i^2 = (1, -\boldsymbol{\beta}_2^{'})\mathbf{\Omega}_i \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{\beta}^{'}\mathbf{\Omega}_i\boldsymbol{\beta}\ , \tag{2.6}$$

where $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$. Since we are interested in the analysis of a large number of cross-section micro-data as typical applications, we impose the condition

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{\Omega}_i \xrightarrow{p} \mathbf{\Omega} \tag{2.7}$$

and $\mathbf{\Omega}$ is a positive definite (constant) matrix. Then

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \xrightarrow{p} \sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} > 0\ . \tag{2.8}$$

Define the $(1+G_2)\times(1+G_2)$ matrices by

$$\mathbf{G} = \mathbf{Y}^{'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}\ , \tag{2.9}$$

and

$$\mathbf{H} = \mathbf{Y}^{'}\left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}\right)\mathbf{Y}\ , \tag{2.10}$$

where $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}^{'}\mathbf{Z}_{2.1}$ and

$$\mathbf{A} = \begin{pmatrix} \mathbf{Z}_1^{'} \\ \mathbf{Z}_{2n}^{'} \end{pmatrix} (\mathbf{Z}_1, \mathbf{Z}_{2n}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{2.11}$$

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

$$(2.12) \qquad (\frac{1}{n}\mathbf{G} - \frac{1}{q_n}\lambda_n \mathbf{H})\hat{\boldsymbol{\beta}}_{LI} = \mathbf{0} \ ,$$

where $q_n = n - K_n$ $(q_n > G_2 + 1)$ and $\lambda_n$ is the smallest root of

$$(2.13) \qquad |\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}| = 0 \ .$$

The solution to (2.11) gives the minimum of the variance ratio

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

(2.14)

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

$$(2.15) \qquad \mathbf{Y}_2^{(n)'}\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}^{'}\mathbf{Y}\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{pmatrix} = \mathbf{0} \ .$$

It minimizes the numerator of the variance ratio (2.14). The LIML and the TSLS estimators of $\boldsymbol{\gamma}_1$ are $\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Y}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is $\hat{\boldsymbol{\beta}}_{LI}$ or $\hat{\boldsymbol{\beta}}_{TS}$, respectively.

The GMM estimation (or the estimating equation method in statistical literatures) can be regarded as a semi-parametric extension of the TSLS estimator. It has been known that the GMM estimator has a significant bias when $K_n$ is large. The MEL estimation can be regarded as a semi-parametric extension of the LIML estimator because the latter can be defined as the minimum variance ratio estimation. Since the calculation of MEL becomes extremely difficult, however, its use has not been implemented when $K_n$ is large. See Anderson et al. (2005, 2007, 2008), Kunitomo and Matsushita (2008) on the finite sample properties of the GMM, MEL, TSLS, and LIML estimators in the detail.

# 3 An Asymptotically Optimal Modification of LIML

## 3.1 Alternative Modifications of the LIML estimator

Anderson, Kunitomo and Matsushita (2007) have considered a set of sufficient conditions in details for an asymptotic optimality of the LIML estimator in a linear structural equation estimation with $\mathbf{\Pi}_{2n}^{(z)} = \mathbf{Z}_1\mathbf{\Pi}_{12} + \mathbf{Z}_2\mathbf{\Pi}_{22}^{(n)}$ ($\mathbf{\Pi}_{22}^{(n)}$ is a $K_{2n} \times G_2$ coefficient matrix) when there are many instruments and the disturbances are homoscedastic. The basic conditions are

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

and

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

as $d_n \xrightarrow{p} \infty$ $(n \to \infty)$, where $\mathbf{\Phi}_{22.1}$ is a nonsingular constant matrix and the noncentrality parameter $d_n^2 = \mathrm{tr}(\mathbf{\Pi}_{2n}^{(z)'}\mathbf{Z}_{2.1}'\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n}^{(z)})$. In the following analysis we shall mainly discuss the standard case when $d_n^2 = O_p(n)$. However, it is straightforward to extend the results to other cases including the case of *many weak instruments*, which we shall mention briefly.

Since the estimation of structural coefficients depends on $\mathbf{G}$ in (2.9), the projection matrix $\mathbf{P}_{2.1} = (p_{ij}^{(2.1)}) = \mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'$ has an important role for the small sample properties of estimators. In Anderson et al. (2007) the condition

$$(\mathbf{A}-\mathbf{VI}) \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(2.1)} - c\right]^2 = 0$$

plays a crucial role, where $p_{ii}^{(2.1)}$ are the diagonal elements of $\mathbf{P}_{2.1}$. The typical example of (A-VI) is the case when we have orthogonal dummy variables which have 1 or $-1$ in their all components so that $(1/n)\mathbf{A}_{22.1} = \mathbf{I}_{K_{2n}}$ and $p_{ii}^{(2.1)} = K_{2n}/n$ $(i = 1,\cdots,n)$. When both (2.7) and (A-VI) hold,

$$(\mathbf{WH}) \qquad \operatorname*{plim}_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n}p_{ii}^{(2.1)}\mathbf{\Omega}_i - c\mathbf{\Omega}\right] = \mathbf{O}$$

by applying the Cauchy-Schwartz inequality. We say the *Weak Heteroscedasticity* condition holds if we have (WH). If it is not satisfied, we say the *Persistent Heteroscedasticity* condition holds and denote (PH). Under (WH), the LIML estimator has some desirable asymptotic properties in the sense that it has the consistency, the asymptotic normality and it attains the lower bound of the asymptotic variance in a class of estimators as $d_n \xrightarrow{p} \infty$ $(n \to \infty)$ as stated in Section 4 of Anderson et al. (2007).

In the more general cases with (PH), however, the distribution of the LIML estimator could be significantly affected by the presence of (conditional) heteroscedasticity of disturbance terms with many instruments. It is mainly because the condition (WH) is not necessarily satisfied. In this respect, there can be several ways to improve the LIML estimation method. Since the projection matrix of instruments has a key role, it is useful to summarize its property.

**Lemma 1** : Let $\mathbf{P}_Z = (p_{ij}^{(n)}) = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{Q}_Z = (q_{ij}^{(n)}) = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. We assume that the rank of matrix $\mathbf{Z}$ is $K_n$ $(> G_2)$. Then $0 \leq p_{ii}^{(n)} < 1$ $(i = 1, \cdots, n)$ and $0 < q_{ii}^{(n)} \leq 1$ $(i = 1, \cdots, n)$. (A-I) implies

$$\bar{p}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} p_{ii}^{(n)} = \frac{K_n}{n} \longrightarrow c\ , \tag{3.1}$$

$$\bar{q}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} q_{ii}^{(n)} = 1 - \frac{K_n}{n} \longrightarrow 1 - c\ , \tag{3.2}$$

where $c_n = K_n/n \to c$ as $n \to \infty$.

The main reason why the LIML estimator does not necessarily have good properties when the disturbances are heteroscedastic with many instruments is the presence of the possible correlation between the conditional covariance $\mathbf{\Omega}_i$ and $p_{ii}^{(n)}$ $(i = 1, \cdots, n)$, which prevents from satisfying (WH). Then we could use this characterization of the diagonal elements of the projection matrix to improve the LIML estimation.

For $\mathbf{P}_Z = (p_{ij}^{(n)}) = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{Q}_Z = (q_{ij}^{(n)}) = \mathbf{I}_n - \mathbf{P}_Z$ and $\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'$, we utilize the relations $\mathbf{P}_{2.1} = (\mathbf{I}_n - \mathbf{P}_{Z_1})\mathbf{P}_Z(\mathbf{I}_n - \mathbf{P}_{Z_1})$ and $\mathbf{Q}_Z = (\mathbf{I}_n - \mathbf{P}_{Z_1})(\mathbf{I}_n - \mathbf{P}_Z)(\mathbf{I}_n - \mathbf{P}_{Z_1})$. We construct $\mathbf{P}_M = (p_{ij}^{(m)})$ and $\mathbf{Q}_M = (q_{ij}^{(m)}) = \mathbf{I}_n - \mathbf{P}_M$ such that

$p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i \neq j)$, $p_{ii}^{(m)} - K_{2n}/n \to 0$ $(i, j = 1, \cdots, n)$ and

$$\text{(3.3)} \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[ p_{ii}^{(m)} - c \right]^2 = 0 \ .$$

Then we define two $(K_1 + 1 + G_2) \times (K_1 + 1 + G_2)$ matrices by

$$\text{(3.4)} \qquad \mathbf{G}_M = \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{Y}' \end{pmatrix} \mathbf{P}_M \left( \mathbf{Z}_1, \mathbf{Y} \right) \ , \ \mathbf{H}_M = \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{Y}' \end{pmatrix} \mathbf{Q}_M \left( \mathbf{Z}_1, \mathbf{Y} \right) \ .$$

By using $\mathbf{G}_M$ and $\mathbf{H}_M$, we define a class of modifications of the LIML estimator (we may call AOM-LIML) such that $\hat{\boldsymbol{\theta}}_{MLI}$ $(= (-\hat{\boldsymbol{\gamma}}_{1.MLI}', \hat{\boldsymbol{\beta}}_{MLI}')')$ and $\hat{\boldsymbol{\beta}}_{MLI}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.MLI}')')$ of $\boldsymbol{\theta} = (-\boldsymbol{\gamma}_1', 1, -\boldsymbol{\beta}_2')'$ is the solution of

$$\text{(3.5)} \qquad \left[ \frac{1}{n} \mathbf{G}_M - \frac{1}{q_n} \lambda_n \mathbf{H}_M \right] \hat{\boldsymbol{\theta}}_{MLI} = \mathbf{0} \ ,$$

where $q_n = n - K_n$ $(> 0)$ and $\lambda_n$ is the smallest root of

$$\text{(3.6)} \qquad \left| \frac{1}{n} \mathbf{G}_M - l \frac{1}{q_n} \mathbf{H}_M \right| = 0 \ .$$

As the simplest case, the AOM-LIML estimator is defined by using the deterministic sequences $p_{ii}^{(m)} = c_n$, $p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i \neq j; i, j = 1, \cdots, n)$.

When $p_{ii}^{(n)}$ $(i = 1, \cdots, n)$ are close to $c_n$ or $c_n$ is small, the AOM-LIML estimator is very close to the LIML estimator for practical purpose. Hausman et al. (2007) have defined the JLIML (or HLIM) estimator by setting $\mathbf{P}_H = (p_{ij}^*)$, $p_{ii}^* = 0$ $(i = 1, \cdots, n)$ and replacing $\mathbf{P}_M$ and $\mathbf{Q}_M$ by $\mathbf{P}_H$ and $\mathbf{Q}_H = \mathbf{I}_n - \mathbf{P}_H$ in (3.4), (3.5) and (3.6) but without (3.3). Then we find that it is not in the class of the AOM-LIML estimation with (3.6). Numerically, however, the AOM-LIML estimator can be close to the JLIML (or HLIM) estimator in some situation when $c_n$ is close to zero. When $c_n$ is not 0, however, there can be some differences in finite samples. It is also possible to define the corresponding modifications of the TSLS estimator and the GMM estimator. An estimation method called JIVE (Jackknife Instrumental Variables Estimators) has been proposed and its properties have been investigated by Chao and Swanson (2005), for instance.

We note that $\boldsymbol{G}_M$ with $\mathbf{P}_M$ should be positive definite (a.s.) in order to define the AOM-LIML estimation. This condition is weaker than the corresponding one with $\mathbf{P}_H$. Hence we expect that the AOM-LIML estimator may be stable than the JLIML estimator in some cases.

## 3.2 Asymptotic Optimality of AOM-MLIML

We shall investigate the asymptotic properties of the AOM-LIML estimator when there are many instruments. One of attractive features of the AOM-LIML estimator is that it satisfies (3.3) while we can utilize nearly full information of data.

We have the consistency and the asymptotic normality of the MLIML estimator when the disturbances are heteroscedastic with many instruments under a set of conditions. The proof will be given in Section 6.

**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 $\mathbf{v}_i$ $(i = 1, 2, \cdots, n)$ be a set of $(1 + G_2) \times 1$ independent random vectors such that $\mathcal{E}(\mathbf{v}_i|\mathbf{z}_i^{(n)}) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}|\mathbf{z}_i^{(n)}) = \mathbf{\Omega}_i$ (a.s.) is a function of $\mathbf{z}_i^{(n)}$, say, $\mathbf{\Omega}_i[n, \mathbf{z}_i^{(n)}]$. For (2.1) and (2.2), suppose (A-I), (2.7), (3.3),

$$\text{(3.7)} \qquad \frac{1}{n}\max_{1\le i\le n} \|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})\|^2 \xrightarrow{p} 0$$

and

$$\text{(3.8)} \qquad \frac{1}{n}\mathbf{\Pi}_{*n}^{(z)'}(\mathbf{P}_M - c_*\mathbf{Q}_M)\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{\Phi}^*$$

is a positive definite matrix as $n \to \infty$, $K_n \to \infty$ and $q_n \to \infty$, where $\mathbf{\Pi}_{*n}^{(z)} = \left(\mathbf{Z}_1, \mathbf{\Pi}_{2n}^{(z)}\right) = (\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)'}))$. We denote $(1/n)\mathbf{\Pi}_{*n}^{(z)'}\mathbf{P}_M\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{\Phi}_1^*$, $(1/q_n)\mathbf{\Pi}_{*n}^{(z)'}\mathbf{Q}_M\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{\Phi}_2^*$ and $c_* = c/(1-c)$. Also suppose $\mathcal{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathcal{E}[\|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)'})\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)'})$ are stochastic).
Then

$$\text{(3.9)} \qquad \sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.MLI} \\ \hat{\boldsymbol{\beta}}_{2.MLI} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^*)$$

where

$$\text{(3.10)} \qquad \mathbf{\Psi}^* = \mathbf{\Phi}^{*-1}\left[\mathbf{\Psi}_1^* + \mathbf{\Psi}_2^*\right]\mathbf{\Phi}^{*-1} ,$$

$$\mathbf{\Psi}_1^* = \operatorname{plim}\frac{1}{n}\sum_{i,j,k=1}^{n} \boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})[p_{ij}^{(m)} - c_*q_{ij}^{(m)}]\sigma_j^2[p_{jk}^{(m)} - c_*q_{jk}^{(m)}]\boldsymbol{\pi}_{*k}(\mathbf{z}_k^{(n)})' ,$$

$$\mathbf{\Psi}_2^* \;=\; \mathrm{plim}\frac{1}{n}\sum_{i,j=1}^{n}\left[\sigma_i^2\mathcal{E}(\mathbf{w}_{*j}\mathbf{w}_{*j}^{'}|\mathbf{z}_j^{(n)})+\mathcal{E}(\mathbf{w}_{*i}u_i|\mathbf{z}_i^{(n)})\mathcal{E}(\mathbf{w}_{*j}^{'}u_j|\mathbf{z}_j^{(n)})\right]\left[p_{ij}^{(m)}-c_*q_{ij}^{(m)}\right]^2 ,$$

provided that $\mathbf{\Psi}_1^*$ and $\mathbf{\Psi}_2^*$ converge in probability as $n\to\infty$, $\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})=(\mathbf{z}_{1i}^{'},\boldsymbol{\pi}_{2i}^{'}(\mathbf{z}_i^{(n)}))^{'}$, $\mathbf{w}_{*i}=(\mathbf{0}^{'},\mathbf{w}_{2i}^{'})^{'}$, and $\mathbf{w}_{2i}=\mathbf{v}_{2i}-u_i(\mathbf{0},\mathbf{I}_{G_2})\mathbf{\Omega}\boldsymbol{\beta}/\sigma^2$ $(i=1,\cdots,n)$.

The first term of (3.10) is due to the noncentrality parameter and the second term is due to the covariance estimation. We could interpret *many weak instruments* as the case when the first term is negligible as we shall discuss.

When (2.2) is linear, we have (2.4) and we partition the $(K_1+K_{2n})\times(1+G_2)$ coefficient matrix as

$$\mathbf{\Pi}_n=\begin{pmatrix}\boldsymbol{\pi}_{11} & \mathbf{\Pi}_{12}\\ \boldsymbol{\pi}_{21}^{(n)} & \mathbf{\Pi}_{22}^{(n)}\end{pmatrix} . \tag{3.11}$$

Suppose the disturbances have the homoscedasticity or weakly heteroscedastic in the sense

$$(\mathbf{WH})^{'} \qquad \max_{1\le i\le n}\|\mathbf{\Omega}_i-\mathbf{\Omega}\|\xrightarrow{p}0$$

and assume the condition (A-VI). Then by setting $p_{ij}^*=p_{ij}^{(n)}$ $(i,j=1,\cdots,n)$, $\mathbf{\Phi}_2^*=\mathbf{O}$ and

$$\mathbf{\Psi}_1^*=\sigma^2\mathrm{plim}\frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\boldsymbol{A}_{22.1}\mathbf{\Pi}_{22}^{(n)}=\sigma^2\mathbf{\Phi}_{22.1}\ . \tag{3.12}$$

In this case we have

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

where $\boldsymbol{A}_{22.1}=\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ and $[\ \cdot\ ]$ is the $G_2\times G_2$ lower left-corner of the matrix. We also use the relations $\sum_{i,j=1}^{n}p_{ij}^{(n)2}=\sum_{i=1}^{n}p_{ii}^{(n)}=K_n$, $\sum_{i,j=1}^{n}q_{ij}^{(n)2}=\sum_{i=1}^{n}q_{ii}^{(n)}=n-K_n$ and $\sum_{i,j=1}^{n}p_{ij}^{(n)2}=\sum_{i=1}^{n}p_{ii}^{(n)}=K_n$. Hence the right-lower corner of $\mathbf{\Psi}_2^*$ is reduced to

$$\begin{aligned}[\mathbf{\Psi}_2^*]_{22} &= \sigma^2\mathrm{plim}\frac{1}{n}\sum_{i,j=1}^{n}\left[p_{ij}^{(n)}-c_*q_{ij}^{(n)}\right]^2\mathcal{E}(\mathbf{w}_{2i}\mathbf{w}_{2i}^{'})\\ &= \left[\frac{c}{1-c}\right]\sigma^2\left[\mathbf{\Omega}-\frac{1}{\sigma^2}\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{\Omega}\right]_{22} .\end{aligned} \tag{3.13}$$

Then $\mathbf{\Psi}^*$ in (3.10) corresponds to

$$(3.14) \qquad \mathbf{\Psi}_A^* = \sigma^2 \mathbf{\Phi}^{*-1} + c_* \mathbf{\Phi}^{*-1} \left[\mathbf{O}, \mathbf{I}_{G_2}\right]' \left[\mathbf{\Omega}\sigma^2 - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]_{22} \left[\mathbf{O}, \mathbf{I}_{G_2}\right] \mathbf{\Phi}^{*-1}$$

where $\sigma^2 = \boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}$ and $c_* = c/(1-c)$. We find that (3.13) reduces to (3.8) of *Theorem 2* in Anderson et al. (2007).

For the estimation of the vector of structural parameters $\boldsymbol{\theta}$, it may be natural to investigate the procedures based on two $(K_1+1+G_2) \times (K_1+1+G_2)$ matrices $\mathbf{G}_M$ and $\mathbf{H}_M$ (by modifying $\mathbf{G}$ and $\mathbf{H}$ for the persistent heteroecedasticity) and hence we consider a class of estimators which are functions of these matrices. Typical examples of this class are the modified versions of the OLS estimator, the TSLS estimator, and the LIML estimator including the one proposed by Fuller (1977). (It also includes other estimators which are asymptotically equivalent to these estimators.) Then we have a new result on the asymptotic optimality of the AOM-LIML estimator in a class of estimators. We give the proof in Section 6.

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

$$(3.15) \qquad \hat{\boldsymbol{\theta}} = \phi(\frac{1}{n}\mathbf{G}_M, \frac{1}{q_n}\mathbf{H}_M) ,$$

where $\phi$ is continuously differentiable and its derivatives are bounded at the probability limits of random matrices in (3.4) as $K_{2n} \to \infty$ and $n \to \infty$ and $0 \le c < 1$. Then under the assumptions of *Theorem 1*,

$$(3.16) \qquad \sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}) ,$$

where

$$(3.17) \qquad \mathbf{\Psi} \ge \mathbf{\Psi}^*$$

and $\mathbf{\Psi}^*$ is given in Theorem 1.

When the conditions ($\mathbf{WH}'$) and (A-VI) are satisfied, the result of Theorem 2 corresponds to an extension of Theorem 4 of Anderson et al. (2007). When the equations of (2.2) are linear and the disturbances are normally distributed with the homoscedastic disturbances,

$$\mathbf{I}(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\mathbf{\Pi}_{*n}^{(z)'}\mathbf{P}_Z\mathbf{\Pi}_{*n}^{(z)} \tag{3.18}$$

corresponds to the Fisher information. Hence $(\mathbf{WH}^{'})$ and $c = 0$ in (A-I) in the linear models are the sufficient condition that we do not loose the information amount by modifying the LIML estimation asymptotically. If they were not satisfied, the AOM-LIML estimator has some information loss asymptotically although it is still consistent and it has the asymptotic normality.

Also Anderson et al. (2007) have investigated an asymptotic optimality of alternative estimators in three possible cases on the sequences of $d_n$ and $n$ when both $d_n$ and $n$ go to infinity under homoscedasticity assumption. From our construction of the AOM-LIML method, it is straightforward to obtain the corresponding asymptotic results for alternative parameter sequences when the disturbances are heteroscedastic and there are many instruments at the same time.

Let $\hat{\boldsymbol{\theta}}_{HLI}$ $(= (-\hat{\boldsymbol{\gamma}}_{1.HLI}^{'}, 1, -\hat{\boldsymbol{\beta}}_{2.HLI}^{'})')$ be the JLIML (or HLIM) estimator defined by Hausman et al. (2007). Then as a Corollary to *Theorem 1*, it is possible to show that the JLIML (or HLIM) estimator cannot be improved asymptotically further.

**Theorem 3** : We take $\mathbf{P}_H = (p_{ij}^*)$ such that $p_{ii}^* = 0$, $p_{ij}^* = p_{ij}^{(n)}$ $(i \neq j; i, j = 1, \cdots, n)$, $\mathbf{Q}_H = \mathbf{I}_n - \mathbf{P}_H$ in (3.4), (3.5) and (3.6) instead of $\boldsymbol{P}_M$ and $\mathbf{Q}_M$. Suppose (A-I), (2.7), and

$$\frac{1}{n}\mathbf{\Pi}_{*n}^{(z)'}(\mathbf{P}_n - \mathbf{D}_n)\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{\Phi}_D^* \tag{3.19}$$

is a positive definite matrix as $n \to \infty$ and $K_n \to \infty$, where $\mathbf{D}_n = \mathrm{diag}(\mathbf{P}_Z)$. Also suppose $\mathcal{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathcal{E}[\|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)'})\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)'})$ are stochastic). Then

$$\sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.HLI} \\ \hat{\boldsymbol{\beta}}_{2.HLI} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^*) \tag{3.20}$$

where $\mathbf{\Psi}^*$ is given by (3.10).

Theorem 3 together with Theorem 2 implies that the JLIML (or HLIM) estimation cannot be improved asymptotically in a class of estimators which depend on functions of $\mathbf{G}_M$ and $\mathbf{Q}_M$ with some $\mathbf{P}_H$ and $\mathbf{Q}_H$. The condition (3.19) is equivalent to (3.8) because of (5.29).

Next, we consider the linear model (2.1) and (2.4) when the noncentrality parameter $d_n = o_p(n^{1/2})$ and $\sqrt{n}/d_n^2 \xrightarrow{p} 0$, which may correspond to the case of *many weak instruments*. We have the asymptotic optimality result in this situation. Since the proof is similar to that of Theorem 5 in Anderson et al. (2007), we omit the detail. It is possible to extend the result further with an additional assumption and complication. The variance of the limiting distribution of the AOM-LIML estimator ((3.24) below) is simpler than (3.10) because the effects of $n$ dominate the first term of (3.10) in *Theorem 1*.

**Theorem 4** : Consider the linear model of (2.1) and (2.4). Suppose (A-I) and (2.7) hold, and let $d_n = o_p(n^{1/2})$ and $\sqrt{n}/d_n^2 \xrightarrow{p} 0$ as $n \to \infty$ and $K_n \to \infty$. Assume

$$\text{(3.21)} \qquad \frac{1}{d_n^2} \max_{1 \le i \le n} \|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})\|^2 \xrightarrow{p} 0$$

and

$$\text{(3.22)} \qquad \frac{1}{d_n^2} \boldsymbol{\Pi}_{*n}^{(z)\prime} (\mathbf{P}_M - c_* \mathbf{Q}_M) \boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}^{**}$$

is a positive definite matrix as $n \to \infty$, $K_n \to \infty$ and $q_n \to \infty$, where $\boldsymbol{\Pi}_{*n}^{(z)} = \left(\mathbf{Z}_1, \boldsymbol{\Pi}_{2n}^{(z)}\right) = (\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)\prime}))$. Also suppose $\mathcal{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathcal{E}[\|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)\prime})\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)\prime})$ are stochastic). We denote $(1/d_n^2)\boldsymbol{\Pi}_{*n}^{(z)\prime}\mathbf{P}_Z^*\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_1^*$, $(1/d_n^2)\boldsymbol{\Pi}_{*n}^{(z)\prime}\mathbf{Q}_Z^*\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_2^*$ and $c_* = c/(1-c)$. For the AOM-LIML estimator,

$$\text{(3.23)} \qquad \left[\frac{d_n^2}{\sqrt{n}}\right] \left[ \begin{pmatrix} \hat{\boldsymbol{\gamma}}_{1.MLI} \\ \hat{\boldsymbol{\beta}}_{2.MLI} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{**})$$

and for any estimator $\hat{\boldsymbol{\theta}}$ in the class of (3.15),

$$\text{(3.24)} \qquad \left[\frac{d_n^2}{\sqrt{n}}\right] \left[ \begin{pmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}) ,$$

where

$$\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^{**} , \tag{3.25}$$

and

$$\boldsymbol{\Psi}^{**} = \boldsymbol{\Phi}^{*-1}\boldsymbol{\Psi}_2^{**}\boldsymbol{\Phi}^{*-1} , \tag{3.26}$$

$$\boldsymbol{\Psi}_2^{**} \;=\; \text{plim}\frac{1}{n}\sum_{i,j=1}^{n}\left[\sigma_i^2\mathcal{E}(\mathbf{w}_{*j}\mathbf{w}_{*j}^{'}|\mathbf{z}_j^{(n)}) + \mathcal{E}(\mathbf{w}_{*i}u_i|\mathbf{z}_i^{(n)})\mathcal{E}(\mathbf{w}_{*j}^{'}u_j|\mathbf{z}_j^{(n)})\right]\left[p_{ij}^{(m)} - c_* q_{ij}^{(m)}\right]^2 ,$$

provided that $\boldsymbol{\Psi}_2^{**}$ converge in probability as $n \to \infty$.

## 3.3 On Finite Sample Distributions of LIML and AOM-LIML

The finite sample properties of the LIML estimator and semi-parametric estimators including the GMM and MEL estimators have been investigated by Anderson, Kunitomo and Matsushita (2005, 2008) in a systematic way. As an example we present only three figures (Figures 1A-3A) in Appendix when we have the linear model with (2.4) and $G_2 = 1$ for the simplicity. (We took a typical case when $K_2$ is relatively large.) We have used the numerical evaluation of the cumulative distribution function (cdf) of the LIML estimator based on the simulation and we have enough numerical accuracy in most cases. See Anderson et al. (2005, 2008) for the detail of the numerical computation method. The key parameters in figures are $K_2$ (or $K_{2n}$), $n-K$ (or $n-K_n$), $\alpha = [\omega_{22}/|\boldsymbol{\Omega}|^{1/2}](\beta_2 - \omega_{12}/\omega_{22})$ ($\boldsymbol{\Omega} = (\omega_{ij})$) and $\delta^2 = \boldsymbol{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}^{(n)}/\omega_{22}$ when $G_2 = 1$ and the disturbances are homoscedastic.

As a simple example of the LIML modification, we consider the case when $K_1 = 1$, $\mathbf{z}_i \sim N(0, \mathbf{I}_K)$ $(i = 1\cdots, n)$ and we take

$$p_{ii}^{(m)} = 1 - q_{ii}^{(m)} = \frac{K_n}{n} + \left[\frac{z_{1i}^2}{\sum_{j=1}^{n} z_{1j}^2} - \frac{1}{n}\right] \tag{3.27}$$

and $p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i \neq j; i,j = 1,\cdots,n)$. Figures 1A and 2A correspond to the homoscedastic disturbance case while Figure 3A corresponds to the case of persistent heteroscedasticity which is similar to the one reported by Hausman et al. (2007).

Three figures in Appendix show the estimated cdf of estimators in the standard form, that is,

$$\sqrt{n}\boldsymbol{\Psi}_{22}^{*-1/2}\left(\hat{\beta}_2-\beta_2\right)\ , \tag{3.28}$$

where $\boldsymbol{\Psi}_{22}^{*}$ is the right-lower corner of $\boldsymbol{\Psi}^{*}$, which is given by Theorem 1. The limiting distribution of the AOM-LIML estimator is $N(0,1)$ in the large-$K_2$ asymptotics and it is denoted by "o".

From these figures we have found that the distribution function of the AOM-LIML estimator is very similar to that of the LIML estimator in the homoscedastic disturbance cases. At the same time we also have found that the distribution function of the AOM-LIML estimator is very similar to that of the JLIML (or HLIM) estimator in the particular heteroscedastic disturbance case treated by Hausman et al. (2007). In that case the finite sample distribution of the LIML estimator is different from the MLIM and JLIML estimators considerably as well as the standard normal distribution because the effects of correlation between $\|\mathbf{z}_i^{(n)}\|/n$ and $\boldsymbol{\Omega}_i$ do not decrease as $K_n$ and n increase. In this case the AOM-LIML estimator with (3.4) improves both the LIML and JLIML estimators in the finite samples. These observations agree with our theoretical results of Section 3.2.

## 4. Concluding Remarks

In this paper, we have introduced a class of modifications of the LIML estimation method. When there are many instruments and the disturbances have heteroscedasticity, it might be argued that the LIML estimator does loose good asymptotic properties in the extremely heteroscedastic cases. However, as we have shown that a simple modification of the LIML estimation, called the AOM-LIML estimator, gives the consistency, the asymptotic normality and an asymptotic optimality under a set of assumptions. The AOM-LIML estimator is close to the LIML estimator when the disturbances are homoscedastic or weakly heteroscedastic while it can be different when the disturbances have persistent heteroscedasticity. We also have shown that the AOM-LIML estimator improves the LIML estimator and the JLIML (or HLIM) estimator is asymptotically equivalent to a simple case of the AOM-LIML estimator when there are many instruments and the persistent heteroscedasiticity exists at the same time. There are some differences in the finite samples.

There are several important issues still remained for further investigations. For the more general non-linear estimating equation model (2.5), the nonlinear LIML and TSLS estimators can be defined by substituting $u_i(\boldsymbol{\theta}) = y_{1i} - f_i(\mathbf{z}_{1i}, \mathbf{y}_{2i}, \boldsymbol{\theta})$ 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 (2.13), where $f_i(\,\cdot\,)$ is a known function and $\boldsymbol{\theta}$ is the vector of unknown (structural) parameters. Then our method can be extended to such cases with some notational complications. When the number of restrictions or the dimension becomes large with the sample size, however, the semi-parametric methods such as the GMM and the maximum empirical likelihood (MEL) estimations may have some difficulty in theory as well as in practical computation.

Finally, a more practical question is the relevance *Persistent Heteroscedasticity* in real applications. A more systematic investigation of the finite sample properties of alternative semi-parametric estimation methods would be needed.

# 5 Proof of Theorems

In this section we give the proofs of *Theorems*. The methods of proofs are basically some modifications of Section 6 of Anderson et al. (2007), which are often straightforward.

**Proof of Lemma 1** : Let $\mathbf{Z}_{2.1} = (\mathbf{z}_i^{*'})$ ($\mathbf{z}_i^{*'}$ are $K_{2n} \times 1$ vectors) and $\mathbf{A}_{n(i)} = \sum_{j=1, j\neq i}^{n} \mathbf{z}_j^* \mathbf{z}_j^{*'}$. Then

$$
\text{(5.1)} \qquad p_{ii}^{(n)} = \mathbf{z}_i^{*'} \left[ \mathbf{z}_i^* \mathbf{z}_i^{*'} + \mathbf{A}_{n(i)} \right]^{-1} \mathbf{z}_i^* = \frac{\mathbf{z}_i^{*'} \mathbf{A}_{n(i)}^{-1} \mathbf{z}_i^*}{1 + \mathbf{z}_i^{*'} \mathbf{A}_{n(i)}^{-1} \mathbf{z}_i^*}
$$

and $0 \leq p_{ii}^{(n)} < 1$. For $\mathbf{Q}_n$ we apply the same argument to $\mathbf{I}_n - \mathbf{Q}_n$ and we find that $0 < q_{ii}^{(n)} \leq 1$. **Q.E.D.**

**Proof of Theorem 1** : From (2.1) and (2.2) we write $\mathbf{Y} = \mathbf{\Pi}_n^{(z)} + \mathbf{V}$, $\mathbf{\Pi}_n^{(z)} = \left(\mathbf{\Pi}_{1n}^{(z)}, \mathbf{\Pi}_{2n}^{(z)}\right)$ and $\mathbf{\Pi}_{1n}^{(z)} = \mathbf{\Pi}_{2n}^{(z)}\boldsymbol{\beta}_2 + \mathbf{Z}_1\boldsymbol{\gamma}_1$. By substituting this relation into $\mathbf{G}_M$ yields

$$
\begin{aligned}
\mathbf{G}_M &= \left[\begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix} + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix}\right] \mathbf{P}_M \left[\left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right) + (\mathbf{O}, \mathbf{V})\right] \\
&= \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix} \mathbf{P}_M \left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right) + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix} \mathbf{P}_M (\mathbf{O}, \mathbf{V}) \\
&\quad + \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix} \mathbf{P}_M (\mathbf{O}, \mathbf{V}) + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix} \mathbf{P}_M \left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right) ,
\end{aligned}
$$

where $\mathbf{P}_M$ is given in Section 3 and we define an $n \times (K_1 + 1 + G_2)$ matrix $\mathbf{V}_* = (\mathbf{O}, \mathbf{V})$. Then

$$
\begin{aligned}
\mathbf{G}_M &- \left[\begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix} \mathbf{P}_M \left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right)' + K_n \begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2+1} \end{pmatrix} \bar{\mathbf{\Omega}} (\mathbf{O}, \mathbf{I}_{G_2+1})\right] \\
&= \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix} \mathbf{P}_M (\mathbf{O}, \mathbf{V}) + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix} \mathbf{P}_M \left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right) \\
&\quad + \left[\begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix} \mathbf{P}_M (\mathbf{O}, \mathbf{V}) - K_n \begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2+1} \end{pmatrix} \bar{\mathbf{\Omega}} (\mathbf{O}, \mathbf{I}_{G_2+1})\right] ,
\end{aligned}
$$

where $\bar{\mathbf{\Omega}} = (1/n)\sum_{i=1}^n \mathbf{\Omega}_i$. By using $(\mathrm{A} - \mathrm{II}')$ and (3.8), we find that as $n \longrightarrow \infty$

$$
\frac{1}{n}\mathbf{\Pi}_n^{(z)'}\mathbf{P}_M\mathbf{V} \xrightarrow{p} \mathbf{O} , \tag{5.2}
$$

and

$$
\frac{1}{n}\left[\begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix} \mathbf{P}_M (\mathbf{O}, \mathbf{V}) - K_n \begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2+1} \end{pmatrix} \bar{\mathbf{\Omega}} (\mathbf{O}, \mathbf{I}_{G_2+1})\right] \xrightarrow{p} \mathbf{O} . \tag{5.3}
$$

Then as $n \longrightarrow \infty$,

$$
\frac{1}{n}\mathbf{G}_M \xrightarrow{p} \mathbf{G}_0 = \mathbf{B}'\mathbf{\Phi}_1^*\mathbf{B} + c\,\mathbf{\Omega}^* , \tag{5.4}
$$

where a $(K_1 + G_2) \times [K_1 + (1 + G_2)]$ matrix

$$
\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2) = \left[\begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2} \end{pmatrix}\right]
$$

and a $(K_1+1+G_2)\times(K_1+1+G_2)$ matrix

$$\mathbf{\Omega}^* = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{\Omega} \end{bmatrix} .$$

By using (2.7) and (3.8),

$$\text{(5.5)} \quad \frac{1}{q_n}\mathbf{H}_M = \frac{1}{q_n}\left[\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{\Pi}_n^{(z)} \end{pmatrix} + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix}\right]\mathbf{Q}_M\left[\left(\mathbf{Z}_1, \mathbf{\Pi}_n^{(z)}\right) + (\mathbf{O}, \mathbf{V})\right] \xrightarrow{p} \mathbf{H}_0$$

and

$$\mathbf{H}_0 = \mathbf{B}'\mathbf{\Phi}_2^*\mathbf{B} + \mathbf{\Omega}^* .$$

Then (3.6) implies

$$\text{(5.6)} \qquad \left|\mathbf{B}'\left[\mathbf{\Phi}_1^* - (\text{plim}\ \lambda_n)\mathbf{\Phi}_2^*\right]\mathbf{B} - \left[(\text{plim}\ \lambda_n) - c\right]\mathbf{\Omega}^*\right| = 0$$

and we find that $\text{plim}\lambda_n = c$ is a solution. Because $\lambda_n$ is the minimum of

$$\text{(5.7)} \qquad l_n = \frac{\boldsymbol{\theta}'\frac{1}{n}\mathbf{G}_M\boldsymbol{\theta}}{\boldsymbol{\theta}'\frac{1}{n}\mathbf{H}_M\boldsymbol{\theta}} \xrightarrow{p} \frac{\boldsymbol{\theta}'\mathbf{G}_0\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{H}_0\boldsymbol{\theta}}$$

and the minimum of the right-hand side is $c$ under the condition (3.8), Then

$$\text{(5.8)} \qquad \text{plim}\lambda_n = c$$

is the unique solution and $\hat{\boldsymbol{\theta}}_{MLI} \xrightarrow{p} \boldsymbol{\theta}$ as $n \to \infty$ because of (3.5) and (3.6).

Define $\mathbf{G}_1$, $\mathbf{H}_1$, $\lambda_{1n}$, and $\mathbf{b}_1$ by

$$\begin{aligned} \mathbf{G}_1 &= \frac{1}{\sqrt{n}}\left[\begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{\Pi}_n^{(z)'} \end{pmatrix}\mathbf{P}_M(\mathbf{O},\mathbf{V}) + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix}\mathbf{P}_M\left(\mathbf{Z}_1, \mathbf{\Pi}_{*n}^{(z)}\right)\right. \\ &\quad \left. + \begin{pmatrix} \mathbf{O} \\ \mathbf{V}' \end{pmatrix}\mathbf{P}_M(\mathbf{O},\mathbf{V}) - K_n\begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2+1} \end{pmatrix}\bar{\mathbf{\Omega}}\left(\mathbf{O}, \mathbf{I}_{G_2+1}\right)\right] , \end{aligned}$$

$\mathbf{H}_1 = \sqrt{q_n}(\frac{1}{q_n}\mathbf{H} - \mathbf{H}_0)$, $\lambda_{1n} = \sqrt{n}(\lambda_n - c)$ and $\mathbf{b}_1 = \sqrt{n}(\hat{\boldsymbol{\theta}}_{MLI} - \boldsymbol{\theta})$. From (3.5), we have

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

Since $(\mathbf{G}_0 - c\,\mathbf{H}_0)\boldsymbol{\theta} = \mathbf{0}$, (3.5) gives

$$\text{(5.9)}\quad \mathbf{B}^{'}\boldsymbol{\Phi}^{*}\sqrt{n}\left[\begin{pmatrix} \hat{\gamma}_{1.LI} \\ \hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] = (\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1)\,\boldsymbol{\theta} + o_p(1)\ .$$

Multiplication of (5.9) from the left by $\boldsymbol{\theta}^{'} = (-\boldsymbol{\gamma}_1^{'}, 1, -\boldsymbol{\beta}_2^{'})$ yields

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

Also the multiplication of (5.9) on the left by a $(K_1 + G_2) \times (K_1 + 1 + G_2)$ choice matrix

$$\mathbf{J}^{'} = \begin{bmatrix} \mathbf{I}_{K_1}, \mathbf{0},\ \mathbf{O} \\ \mathbf{O}\ \ , \mathbf{0},\ \mathbf{I}_{G_2} \end{bmatrix}$$

and substitution for $\lambda_{1n}$ from (5.9) yield

$$\begin{aligned} \text{(5.10)}\qquad & \sqrt{n}\left[\begin{pmatrix} \hat{\gamma}_{1.LI} \\ \hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} - \begin{pmatrix} \gamma_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] \\ &= \boldsymbol{\Phi}^{*-1}\mathbf{J}^{'}\,(\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1)\,\boldsymbol{\theta} + o_p(1) \\ &= \boldsymbol{\Phi}^{*-1}\mathbf{J}^{'}\left[\mathbf{I}_{K_1+G_2+1} - \frac{\mathbf{H}_0\boldsymbol{\theta}\boldsymbol{\theta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\right](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\,\boldsymbol{\theta} + o_p(1)\ . \end{aligned}$$

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

$$\begin{aligned} \text{(5.11)}\quad & \mathbf{J}^{'}\,(\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\,\boldsymbol{\theta} \\ =\ & \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{*n}^{(z)'}\,(\mathbf{P}_M - c_*\mathbf{Q}_M)\,\mathbf{u} + \sqrt{c}\frac{1}{\sqrt{K_n}}\mathbf{J}^{'}\left[\begin{pmatrix} \mathbf{O} \\ \mathbf{V}^{'} \end{pmatrix}\mathbf{P}_M\mathbf{u} - K_n\begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2} \end{pmatrix}\bar{\boldsymbol{\Omega}}\boldsymbol{\beta}\right] \\ & -\sqrt{cc_*}\frac{1}{\sqrt{q_n}}\mathbf{J}^{'}\left[\begin{pmatrix} \mathbf{O} \\ \mathbf{V}^{'} \end{pmatrix}\mathbf{Q}_M\mathbf{u} - q_n\begin{pmatrix} \mathbf{O} \\ \mathbf{I}_{G_2} \end{pmatrix}\bar{\boldsymbol{\Omega}}\boldsymbol{\beta}\right]\ , \end{aligned}$$

where $K_n + q_n = n$. Then by defining a $(1 + K_1 + G_2) \times n$ matrix

$$\mathbf{W}^{'} = \mathbf{J}^{'}\left[\mathbf{I}_{K_1+G_2+1} - \frac{\boldsymbol{\Omega}^{*}\boldsymbol{\theta}\boldsymbol{\theta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\right]\begin{pmatrix} \mathbf{O} \\ \mathbf{V}^{'} \end{pmatrix}\ ,$$

(5.10) is rewritten as

$$\begin{aligned} \text{(5.12)}\quad & \sqrt{n}\left[\begin{pmatrix} \hat{\gamma}_{1.LI} \\ \hat{\boldsymbol{\beta}}_{2.LI} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right] \\ =\ & \boldsymbol{\Phi}^{*-1}\frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{*n}^{(z)'}\,(\mathbf{P}_M - c_*\mathbf{Q}_M)\,\mathbf{u} + \boldsymbol{\Phi}^{*-1}\frac{1}{\sqrt{n}}\left[\mathbf{W}^{'}\,(\mathbf{P}_M - c_*\mathbf{Q}_M)\,\mathbf{u}\right] + o_p(1)\ . \end{aligned}$$

Then the rest of the proof (i.e. for the asymptotic normality of the AOM-LIML estimator) is essentially the same as the proof of *Theorem 1*, *Theorem 2* and *Lemma 3* in Anderson et al. (2007). (We omit the detail because we need to use a martingale CLT for quadratic forms and it is straightforward, but quite lengthy.) Some care should be taken because we have to use $\mathbf{P}_M$ and $\mathbf{Q}_M$ instead of $\mathbf{P}_n$ and $\mathbf{Q}_n$, and (3.5) and (3.6) to derive the asymptotic properties of the AOM-LIML estimator. Because the construction of the diagonal parts of $\mathbf{P}_M$ and $\mathbf{Q}_M$, we have the results. **Q.E.D.**

The next proof of Theorem 2 is a simple modification of the proof of Theorem 4 of Anderson et al. (2007) and we shall use their arguments. For the sake of completeness we give the proof for the simple case.

**Proof of Theorem 2** : Without loss of generality, we assume $K_1 = 0$ and $K_{2n} = K_n$. (The notation becomes simple slightly and the essential arguments are clearer than otherwise. See the proof of Theorem 4 of Anderson et al. (2007).) We set the vector of true parameters $\boldsymbol{\theta}' = \boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2') = (1, -\beta_2, \cdots, -\beta_{1+G_2})$. An estimator of the vector $\boldsymbol{\beta}_2$ is composed of

$$\hat{\beta}_k = \phi_k(\frac{1}{n}\mathbf{G}_M, \frac{1}{q_n}\mathbf{H}_M) \ \ (k = 2, \cdots, 1+G_2) \ . \tag{5.13}$$

For the estimator to be consistent we need

$$\beta_k = \phi_k \left[ \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_1^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + c\,\boldsymbol{\Omega}, \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_2^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + \boldsymbol{\Omega} \right] \tag{5.14}$$

for $k = 2, \cdots, 1+G_2$ as the identities with respect to $\boldsymbol{\beta}_2$, $\boldsymbol{\Phi}_k^* = (\psi_{ij}^{(k)})$ $(k = 1, 2)$ and $\boldsymbol{\Omega} = (\omega_{ij})$. We set a $(1+G_2) \times (1+G_2)$ matrix

$$\mathbf{T}^{(k)} = \left( \frac{\partial \phi_k}{\partial g_{ij}} \right) = (\tau_{ij}^{(k)}) \ \ (k = 2, \cdots, 1+G_2; i, j = 1, \cdots, 1+G_2) \tag{5.15}$$

evaluated at the probability limits and then write a $(1+G_2) \times (1+G_2)$ matrix $\boldsymbol{\Theta}_k$ $(= (\theta_{ij}^{(k)}))$ as

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_1^* \left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) = \begin{bmatrix} \boldsymbol{\beta}_2'\boldsymbol{\Phi}_1^*\boldsymbol{\beta}_2 & \boldsymbol{\beta}_2'\boldsymbol{\Phi}_1^* \\ \boldsymbol{\Phi}_1^*\boldsymbol{\beta}_2 & \boldsymbol{\Phi}_1^* \end{bmatrix} \tag{5.16}$$

and $\mathbf{\Theta}_2$ is defined similarly.

Next we consider the role of the second matrix in (5.13). By differentiating (5.14) with respect to $\omega_{ij}$ $(i,j=1,\cdots,1+G_2)$, we have the condition

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

evaluated at the probability limit.

By differentiating each components of $\phi_k$ $(k=1,\cdots,G_2)$ with respect to $\beta_i$ $(i=1,\cdots,G_2)$, we have

$$\begin{aligned} \frac{\partial \phi_k}{\partial \beta_i} &= \sum_{g,h=1}^{G_2+1}\left[\frac{\partial \phi_k}{\partial g_{gh}}\frac{\partial g_{gh}}{\partial \beta_i} + \frac{\partial \phi_k}{\partial h_{gh}}\frac{\partial h_{gh}}{\partial \beta_i}\right] \\ &= \sum_{g,h=1}^{G_2+1}\frac{\partial \phi_k}{\partial g_{gh}}\left[\frac{\partial g_{gh}}{\partial \beta_i} - c\frac{\partial h_{gh}}{\partial \beta_i}\right] \end{aligned} \tag{5.17}$$

and we have

$$\mathbf{tr}\left[\mathbf{T}^{(k)}\left(\frac{\partial \mathbf{\Theta}_1}{\partial \beta_j} - c\frac{\partial \mathbf{\Theta}_2}{\partial \beta_j}\right)\right] = 2\tau_{11}^{(k)}\sum_{i=2}^{1+G_2}(\psi_{ji}^{(1)} - c\psi_{ji}^{(2)})\beta_i + 2\sum_{i=2}^{1+G_2}(\psi_{ji}^{(1)} - c\psi_{ji}^{(2)})\tau_{ji}^{(k)} = \delta_j^k \ , \tag{5.18}$$

where we define $\delta_k^k = 1$ and $\delta_j^k = 0$ $(k \neq j)$.

By defining a $(1+G_2)\times(1+G_2)$ partitioned matrix

$$\mathbf{T}^{(k)} = \begin{bmatrix} \tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)\prime} \\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix} , \tag{5.19}$$

(6.18) is represented as

$$2\tau_{11}^{(k)}\left(\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*\right)\boldsymbol{\beta} + 2\left(\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*\right)\ \boldsymbol{\tau}_2^{(k)} = \boldsymbol{\epsilon}_k \ , \tag{5.20}$$

where $\boldsymbol{\epsilon}_k^{'} = (0,\cdots,0,1,0,\cdots,0)$ with 1 in the k-th place and zeros in other elements. Since we assumed $\mathbf{\Phi}^*$ $(=\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*)$ is positive definite, we solve (5.20) as

$$\boldsymbol{\tau}_2^{(k)} = \frac{1}{2}\mathbf{\Phi}^{*-1}\boldsymbol{\epsilon}_k - \tau_{11}^{(k)}\boldsymbol{\beta}_2 \ . \tag{5.21}$$

Further by differentiating $\mathbf{\Theta}_g$ $(g=1,2)$ with respect to $\psi_{ij}^{(h)}$, we have the representation

$$\mathrm{tr}\left(\mathbf{T}^{(k)}\frac{\partial \mathbf{\Theta}_g}{\partial \psi_{ij}^{(1)}}\right) = \begin{cases} \beta_i^2\tau_{11}^{(k)} + 2\tau_{1i}^{(k)}\beta_i + \tau_{ii}^{(k)} & (i=j) \\ 2\beta_i\beta_j\tau_{11}^{(k)} + 2\tau_{1j}^{(k)}\beta_i + 2\tau_{1i}^{(k)}\beta_j + 2\tau_{ij}^{(k)} & (i \neq j) \end{cases} \tag{5.22}$$

and in the matrix form

$$\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} + \boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'} + \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'} + \mathbf{T}_{22}^{(k)} = \mathbf{O} \quad . \tag{5.23}$$

Then we have the representation

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

Let

$$\mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1 = \begin{bmatrix} s_{11} & \mathbf{s}_2^{'} \\ \mathbf{s}_2 & \mathbf{S}_{22} \end{bmatrix} \ , \tag{5.24}$$

where $\mathbf{G}_1$ and $\mathbf{H}_1$ are defined as in the proof of Theorem 1.
Since $\phi(\,\cdot\,)$ is differentiable and its first derivatives are bounded at the true parameters by assumption, the linearized estimator of $\beta_k$ in the class of our concern can be represented as

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

Let

$$\boldsymbol{\tau}_{11} = \begin{bmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{bmatrix} \tag{5.25}$$

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

$$\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\right]\mathbf{S}\boldsymbol{\beta} \ . \tag{5.26}$$

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

$$\begin{aligned} &\mathcal{E}\left[\hat{\mathbf{e}}\,\hat{\mathbf{e}}^{'}\right] \\ = &\left[(\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}})\right] \end{aligned}$$

$$\times\mathcal{E}[\mathbf{S}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{S}] \times \left[(\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}})\right]^{'}$$

$$= \boldsymbol{\Psi}^* + \mathcal{E}\left[(\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta})^2\right]\left[\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta}\right]\left[\boldsymbol{\tau}_{11}^{'} + \boldsymbol{\beta}^{'}\boldsymbol{\Omega}(\mathbf{0}, \boldsymbol{\Phi}^{*-1})^{'}\right] + o(1) \ ,$$

where $\boldsymbol{\Psi}^*$ has been given in Theorem 1 with $K_1 = 0$. This covariance matrix is the sum of a positive semi-definite matrix of rank 1 and a positive definite matrix. It has a minimum if

$$\boldsymbol{\tau}_{11} = -(\mathbf{0}, \boldsymbol{\Phi}^{*-1})\boldsymbol{\Omega}\boldsymbol{\beta} \ . \tag{5.27}$$

**Q.E.D.**

**Proof of Theorem 3** : We take $p_{ii}^{(m)} = 1 - q_{ii}^{(m)} = c_n$, $p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i \neq j; i, j = 1, \cdots, n)$ and $\mathbf{Q}_M = \mathbf{I}_n - \mathbf{P}_M$ in the AOM-LIML estimation. We use the fact that $\mathbf{P}_M = \mathbf{P}_n - \mathbf{D}_n + c_n\mathbf{I}_n$, $\mathbf{P}_H = \mathbf{P}_n - \mathbf{D}_n$ and $\mathbf{D}_n = \mathrm{diag}(\mathbf{P}_n)$. Then

$$\begin{aligned} \mathbf{P}_M - c_*\mathbf{Q}_M &= [\mathbf{P}_n - \mathbf{D}_n + c_n\mathbf{I}_n] - c_*[\mathbf{I}_n - (\mathbf{P}_n - \mathbf{D}_n + c_n\mathbf{I}_n)] \\ &= (1 + c_*)(\mathbf{P}_n - \mathbf{D}_n) + [c_n - c_*(1 - c_n)]\,\mathbf{I}_n \ . \end{aligned} \tag{5.28}$$

The assumptions in Theorem 1 implies

$$\frac{1}{n}\boldsymbol{\Pi}_{*n}^{(z)'}\left[\mathbf{P}_M - c^*\mathbf{Q}_n^*\right]\boldsymbol{\Pi}_{*n}^{(z)} - (1 + c_*)\frac{1}{n}\boldsymbol{\Pi}_{*n}^{(z)'}\left[\mathbf{P}_n - \mathbf{D}_n\right]\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{O} \ . \tag{5.29}$$

By using the same arguments for $\boldsymbol{\Psi}_i^*$ $(i = 1, 2)$ in (3.10), and we find that the corresponding terms of $\boldsymbol{\Psi}_i^*$ $(i = 1, 2)$ become

$$\begin{aligned} \boldsymbol{\Psi}_1^{**} &= (1 + c_*)^2 \mathrm{plim}\frac{1}{n}\sum_{i,j,k=1}^{n}\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})[p_{ij}^{(n)}(1 - \delta_i^j)]\sigma_j^2[p_{jk}^{(n)}(1 - \delta_j^k)]\boldsymbol{\pi}_{*k}(\mathbf{z}_k^{(n)})^{'} \ , \\ \boldsymbol{\Psi}_2^{**} &= (1 + c_*)^2 \mathrm{plim}\frac{1}{n}\sum_{i,j=1}^{n}\left[\sigma_i^2\mathcal{E}(\mathbf{w}_{*j}\mathbf{w}_{*j}^{'}|\mathbf{z}_j^{(n)}) + \mathcal{E}(\mathbf{w}_{*i}u_i|\mathbf{z}_i^{(n)})\mathcal{E}(\mathbf{w}_{*j}^{'}u_j|\mathbf{z}_j^{(n)})\right] \\ &\qquad \times\left[p_{ij}^{(n)}(1 - \delta_i^j)\right]^2 \ , \end{aligned}$$

where $\delta_i^j = 0$ $(i = j), 0$ $(i \neq j)$. Hence the factors $(1 + c_*)^2$ in $\boldsymbol{\Phi}^{**}$ and $\boldsymbol{\Psi}^*$ are cancelled out.

Let $\lambda_H$ be the smallest root of (3.6) in the JLIML estimation by using $\mathbf{P}_H$ and $\mathbf{Q}_H$ instead of $\mathbf{P}_M$ and $\mathbf{Q}_M$. Then we have plim $\lambda_H^* = 0$ because $p_{ii}^* = 0$ $(i = 1, \cdots, n)$.

Then the asymptotic normality of the JLIML estimator can be established under the assumption of (3.19) that

$$\frac{1}{n}\mathbf{\Pi}_{*n}^{(z)\prime}(\mathbf{P}_n - \mathbf{D}_n)\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} (1+c_*)^{-1}\mathbf{\Phi}^* \tag{5.30}$$

is a positive definite matrix as $n \to \infty$. Hence the covariance matrix of the asymptotic distribution has the same form in Theorem 1. **Q.E.D.**

## References


[1] Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, Vol. 12, 1-45.

[2] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Methods of Simultaneous Equations with Possibly Many Instruments," Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo.

[3] 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, Revised (December 2008).

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

[5] 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.

[6] 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.

[7] Angrist, G.W. Imbens and A. Krueger (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, Vol. 14, 57-67.

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

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

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

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

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

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

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

[15] 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.

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

[17] Kunitomo, N. (1982), "Asymptotic Efficiency and Higher Order Efficiency of the Limited Information Maximum Likelihood Estimator in Large Econometric

Models," Technical Report No. 365, Institute for Mathematical Studies in the Social Sciences, Stanford University.

[18] Kunitomo, N. (1987), "A Third Order Optimum Property of the ML Estimator in a Linear Functional Relationship Model and Simultaneous Equation System in Econometrics," *Annals of the Institute of Statistical Mathematics*, Vol.39, 575-591.

[19] Kunitomo, N. and Matsushita, Y. (2008), "Asymptotic Expansions and Higher Order Properties of Semi-Parametric Estimators in a Linear Simultaneous Equations," Discussion Paper CIRJE-F-237, Graduate School of Economics, University of Tokyo, forthcoming in *Journal of Multivariate Analysis.*

[20] 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.

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

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

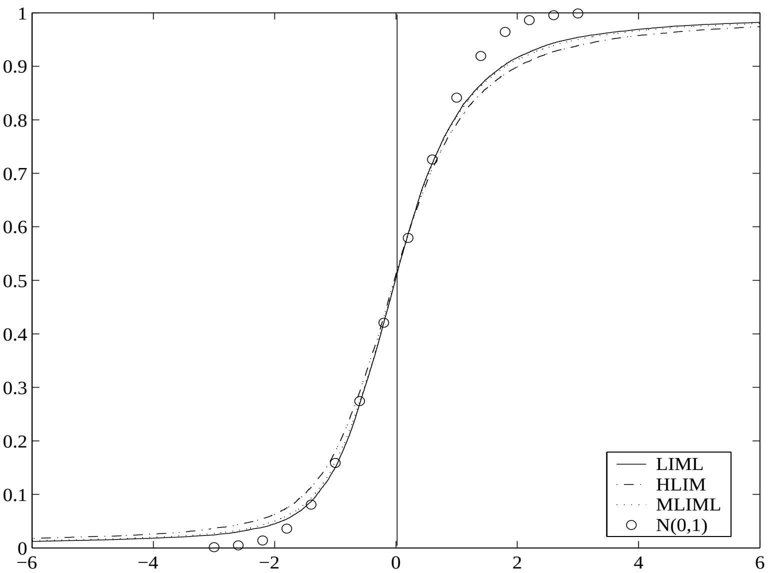


Figure 1A: CDF of Standardized estimators: $n - K = 20, K_2 = 30, \alpha = 0.5, \delta^2 = 30, u_i = N(0,1)$

## APPENDIX : FIGURES

In Figures 1A-3A the distribution functions of the LIML, the HLIM (or JLIML) and the MLIML estimators are shown with the large-$K_2$ normalization. The limiting distributions for the efficient estimators in the large-$K_2$ asymptotics are $N(0,1)$ as $n \to \infty$ and $K_{2n} \to \infty$ which are denoted as "o". The parameter $\alpha$ stands for the normalized coefficient of an endogenous variable and $\delta^2$ is the noncentrality parameter. The details of numerical computation method of this paper are given in Anderson et al. (2005, 2008).

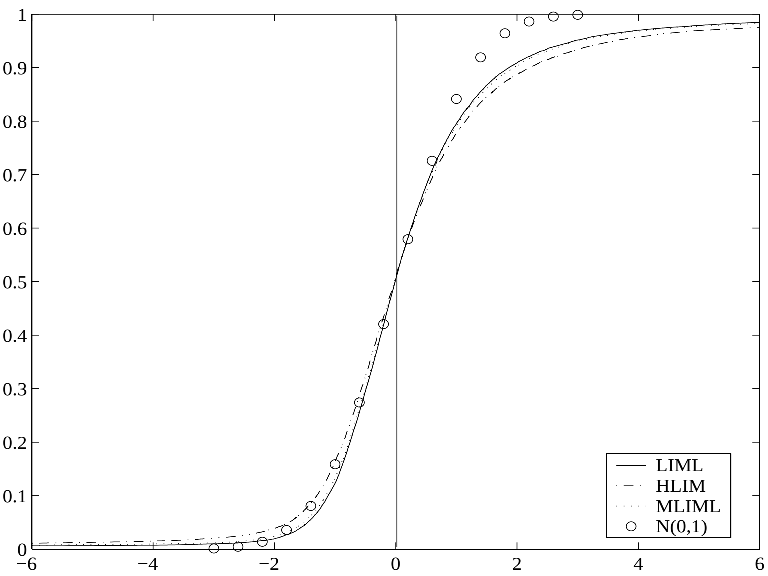


Figure 2A: CDF of Standardized estimators: $n - K = 20, K_2 = 30, \alpha = 1, \delta^2 = 30, u_i = N(0,1)$

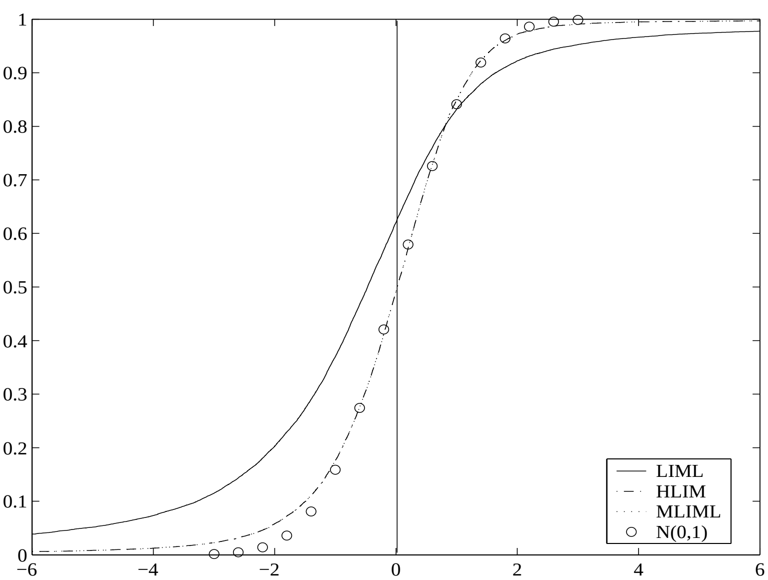


Figure 3A: CDF of Standardized estimators: Heteroscedastic disturbances in Hausman et.al (2007), $n = 100, K = 10, \delta^2 = 30$