Ann Inst Stat Math (2012) 64:881–910
DOI 10.1007/s10463-011-0336-7

An optimal modification of the LIML estimation for many instruments and persistent heteroscedasticity

Naoto Kunitomo

Received: 8 May 2009 / Revised: 13 April 2011 / Published online: 14 September 2011

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 the *estimating equations estimation* or a *reduced rank regression* in the statistical multivariate linear models when the number of restrictions or the dimension of estimating equations 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 of the estimating equations is large with persistent heteroscedasticity including a case of *many weak instruments*.

Keywords Estimation of structural equation · Estimating equation estimation · Reduced rank regression · Many instruments · Persistent heteroscedasticity · AOM-LIML · Asymptotic optimality

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 statistical inference problem of a structural equation in a system of simultaneous equations when the number of

N. Kunitomo (✉)
Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113-0033, Japan
e-mail: kunitomo@e.u-tokyo.ac.jp

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; 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. There has been a growing literature in econometrics on the problem of many instruments including Chao and Swanson (2005), Hansen et al. (2008), Anderson et al. (2005, 2008, 2010) 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 of explanatory variables 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. (2005, 2010) have shown that the LIML estimator has an asymptotic optimal property when K_2 and n are large under a set of conditions. On the other hand, the Jackknife Instrumental Variables Estimation (JIVE) method has been proposed and its properties have been investigated. (See for instance Angrist et al. 1999; Chao and Swanson 2004) Also, Hausman et al. (2007) proposed the jackknife version of the LIML estimator (called HLIM or JLIML) and the Fuller modification. They suggested that the HLIM estimator improves the bias property of the LIML estimator in case of *the persistent heteroscedasticity*, which we shall define precisely in Sect. 3.1.

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. Our motivation is to remove the main cause of possible bias of the LIML estimation when we have the persistent heteroscedasticity and many instruments, while we want to keep the information contained in observations in the form of sufficient statistics in the classical case when the number of instruments is fixed. As we shall explain in Sect. 3.2, the possible information loss caused by our modification can be asymptotically negligible in the first-order asymptotic sense. We shall show that the AOM-LIML estimator improves some properties of the LIML estimator. The AOM-LIML estimator has good asymptotic properties and it 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* under a set of assumptions as we shall state in Sect. 3. When the number of instruments grows with the sample size, we have incidental parameters and we cannot apply the standard arguments on the semi-parametric asymptotic optimality. The GMM estimator, which is semi-parametrically efficient in the standard asymptotic theory, is badly biased when there are many instruments, for example. (See Kunitomo and Matsushita (2009); Anderson et al. 2008.) In this paper, we also relate the AOM-LIML estimator to other estimation methods known and show that the HLIM estimator is asymptotically equivalent to the AOM-LIML estimator in the large-K_2 asymptotic theory. Hence both the AOM-LIML estimator and the HLIM estimator dominate many non-LIML type estimators, but they cannot be improved asymptotically in the class of estimators including the Jackknife Instrumental Variables Estimators. The results of this paper show 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 Sect. 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 Sect. 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 the HLIM estimators. For an illustration of our results in Sect. 3.3, we shall give some figures in Appendix. Finally, some concluding remarks will be given in Sect. 4. The proof of our theorems will be gathered in Sect. 5.

2 Alternative estimation methods of a structural equation with many instruments

Let a single linear structural equation be

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

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and a vector of G_2 endogenous variables, respectively, 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 (K_1 and G_2 are fixed integers), u_i are mutually independent disturbance terms with $\mathcal{E}(u_i|\mathbf{z}_i^{(n)}) = 0$ and $\mathcal{E}(u_i^2|\mathbf{z}_i^{(n)}) = \sigma_i^2$, and the $K_n \times 1$ vectors $\mathbf{z}_i^{(n)}$ $(i = 1, \ldots, n)$ are the instrumental variables. We assume that (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 observations with $\mathbf{y}_1^{(n)} = (y_{1i})$ and $\mathbf{Y}_2^{(n)} = (\mathbf{y}_{2i}')$.

In this paper, we shall consider the cases when (i) $(\mathbf{y}_i', \mathbf{z}_i^{(n)'})$ are a sequence of mutually independent random variables or (ii) $\mathbf{y}_i'$ are mutually independent random variables and $\mathbf{z}_i^{(n)'}$ are a sequence of non-random vectors, while u_i are conditionally heteroscedastic (independent) disturbances.

As a typical situation we have

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

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 $\mathbf{V} = (\mathbf{v}_1^{(n)}, \mathbf{V}_2^{(n)})$. Here $\mathbf{V} = (\mathbf{v}_i')$ is an $n \times (1 + G_2)$ matrix of disturbances (the ith row $\mathbf{v}_i'$ is a $1 \times (1 + G_2)$ vector) with $\mathcal{E}(\mathbf{v}_i|\mathbf{z}_i^{(n)}) = \mathbf{0}$ and

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

We further assume that the conditional covariance matrix $\mathbf{\Omega}_i$ and the conditional variance $\sigma_i^2 = \mathcal{E}(u_i^2|\mathbf{z}_i^{(n)})$ are bounded.

The formulation of (1) and (2) includes the statistical linear models as special cases. We write

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi}_n + \mathbf{V}, \tag{3}$$

$\mathbf{\Pi}_n$ is a $K_n \times (1 + G_2)$ matrix of coefficients and the $n \times K_n$ matrix $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_{2n}) = (\mathbf{z}_i^{(n)'})$ (the ith row $\mathbf{z}_i^{(n)'} = (\mathbf{z}_{1i}', \mathbf{z}_{2i}^{(n)'})$ is the vector of K_n $(= K_1 + K_{2n})$ instruments). The $K_{2n} \times 1$ $(K_{2n} = K_n - K_1)$ vector $\mathbf{z}_{2i}^{(n)}$ is the set of instruments excluded from (1), but they are included in (2).

When $\boldsymbol{\gamma}_1 = \mathbf{0}$ and $\boldsymbol{\beta}_2$ in (1) is uniquely identified ($K_n \geq G_2$), the rank of $\mathbf{\Pi}_n$ in (3) is G_2 and it corresponds to 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 $\mathbf{z}_i^{(n)}$ satisfies the orthogonal condition

$$\mathcal{E}[u_i z_i^{(n)}] = \mathbf{0} \quad (i = 1, \ldots, n), \tag{4}$$

the model of (1) and (2) is the same as *an estimation equation* problem well known in statistics literature, but it may be important to mention that 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 (1) and (2) [or (3)] gives $u_i = (1, -\boldsymbol{\beta}_2')\mathbf{v}_i$ and

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

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

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{\Omega}_i \xrightarrow{p} \mathbf{\Omega} \tag{5}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \xrightarrow{p} \sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta} > 0. \tag{6}$$

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

$$\mathbf{G} = \mathbf{Y}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y}, \tag{7}$$

and

$$\mathbf{H} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}, \tag{8}$$

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

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

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

$$\left(\frac{1}{n}\mathbf{G} - \frac{1}{q_n}\lambda_n\mathbf{H}\right)\hat{\boldsymbol{\beta}}_{\mathrm{LI}} = \mathbf{0}, \tag{9}$$

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

$$\left|\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}\right| = 0. \tag{10}$$

The solution to (9) and (10) gives the minimum of the variance ratio

$$R_n = \frac{[\sum_{i=1}^n \mathbf{z}_i^{(n)'}(y_{1i} - \boldsymbol{\gamma}_1'\mathbf{z}_{1i} - \boldsymbol{\beta}_2'\mathbf{y}_{2i})][\sum_{i=1}^n \mathbf{z}_i^{(n)}\mathbf{z}_i^{(n)'}]^{-1}[\sum_{i=1}^n \mathbf{z}_i^{(n)}(y_{1i} - \boldsymbol{\gamma}_1'\mathbf{z}_{1i} - \boldsymbol{\beta}_2'\mathbf{y}_{2i})]}{\sum_{i=1}^n (y_{1i} - \boldsymbol{\gamma}_1'\mathbf{z}_{1i} - \boldsymbol{\beta}_2'\mathbf{y}_{2i})^2}. \tag{11}$$

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

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

It minimizes the numerator of the variance ratio (11). 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 $\hat{\boldsymbol{\beta}}_{\mathrm{LI}}$ or $\hat{\boldsymbol{\beta}}_{\mathrm{TS}}$, respectively.

The estimation methods we shall consider in this paper are closely related to other methods, which have been commonly used in a large number of econometric analyses, and we shall explain some practical implications of the results reported in this paper. The GMM estimation (or the estimating equation method in statistical literatures) can be regarded as a semi-parametric extension of the TSLS estimator and they have similar finite sample properties. The MEL estimator and the LIML estimator have similar finite sample properties. as discussed by Anderson et al. (2008) and Kunitomo and Matsushita (2009). It has been known that the GMM estimator has a significant bias when K_n is large, while the MEL estimator often does not have such bias and has similar finite sample properties as long as K_{2n} is not large. However, the calculation of MEL becomes extremely difficult when K_n is large and its use has not been implemented in such a situation. See Anderson et al. (2005, 2008, 2010) and Kunitomo and Matsushita (2009) on the finite sample properties of the GMM, MEL, TSLS and LIML estimators in more detail.

3 An asymptotically optimal modification of LIML

3.1 Alternative modifications of the LIML estimator

Anderson et al. (2005, 2010) have considered a set of sufficient conditions 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

$$\frac{K_{2n}}{n} \longrightarrow c \quad (0 \le c < 1) \tag{\textbf{A-I}}$$

and

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

as $d_n \xrightarrow{p} \infty$ $(n \to \infty)$, where $\mathbf{\Phi}_{22.1}$ is a $G_2 \times G_2$ non-singular constant matrix and

$$d_n^2 = \mathrm{tr}(\mathbf{\Pi}_{2n}^{(z)'}\mathbf{Z}_{2.1}'\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\mathbf{\Pi}_{2n}^{(z)})$$

is the non-centrality parameter.

In the following analysis, we shall mainly discuss the corresponding standard case when $d_n^2 \sim n$ and (**A-II**$'$) can be regarded as one type of the non-degeneracy condition for the limiting distribution of estimators when $G_2 > 1$. However, it is straightforward to extend the results to other cases including the case of *many weak instruments*, which we shall mention briefly in Sect. 3.3. The cases when $c > 0$ in (**A-I**) have been often called *many instruments* in recent econometrics. Since there can often be many instruments available and the sample size can be large in some micro-econometric applications, the large-K_2 asymptotic theory can be relevant to investigate the properties of alternative estimators. It may be interesting to note that it was exactly the same situation in the earlier developments by Kunitomo (1980, 1981, 1982, 1987), Morimune (1983) and Bekker (1994) with slightly different notations.

Since the estimation of structural coefficients depends on $\mathbf{G}$ in (7), 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. (2010) the condition

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

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 $p_{ii}^{(2.1)} = K_{2n}/n$ $(i = 1, \ldots, n)$. Since K_1 (the number of included instrumental variables in the structural equation of interest) is fixed and $K_n = K_1 + K_{2n}$, (**A-I**) is equivalent to the condition that $K_n/n = c_n \to c$ and then (**A-VI**) is equivalent to the condition that $(1/n)\sum_{i=1}^{n}[p_{ii}^{(n)} - c]^2 \xrightarrow{p} 0$ with the projection operator $\mathbf{P}_Z = (p_{ij}^{(n)}) = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

When both (5) [and (6)] and (**A-VI**) hold,

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

by applying the Cauchy–Schwarz 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 we shall denote this condition as (PH). The related problem has been investigated systematically by Hausman et al. (2007) in a slightly different, but a more general setting. Under (WH), the LIML estimator has some desirable asymptotic properties in the sense that it has consistency, 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. (2010).

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 as a proposition. The proof will be in Sect. 5.

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* $0 \le p_{ii}^{(n)} < 1$ $(i = 1, \ldots, n)$ *and* $0 < q_{ii}^{(n)} \le 1$ $(i = 1, \ldots, n)$. (**A-I**) *implies*

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

and

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

where $c_n = K_n/n \to c$ *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 $p_{ii}^{(n)}$ $(i = 1, \ldots, n)$, which prevents from satisfying (WH). Then we could use this characterization of the diagonal elements of the projection matrix to improve the LIML estimation. If we replace $p_{ii}^{(2.1)}$ $(i = 1, \ldots, n)$ by some quantities near K_{2n}/n, we automatically satisfy the crucial conditions of (**A-VI**) and then (WH). Hence. we would expect that the resulting modification improves the original LIML estimation while we do not destroy the basic structure and keep the desirable aspects. When we have many instruments under the standard homoscedasticity situation, we do not need to change the essential condition of the original 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 $p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i \neq j)$, $p_{ii}^{(m)} - K_{2n}/n \to 0$ $(i, j = 1, \ldots, n)$ and

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}[p_{ii}^{(m)} - c]^2 = 0. \tag{12}$$

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

$$\mathbf{G}_M = \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{Y}' \end{pmatrix} \mathbf{P}_M \left(\mathbf{Z}_1, \mathbf{Y}\right), \quad \mathbf{H}_M = \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{Y}' \end{pmatrix} \mathbf{Q}_M \left(\mathbf{Z}_1, \mathbf{Y}\right). \tag{13}$$

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

$$\left[\frac{1}{n}\mathbf{G}_M - \frac{1}{q_n}\lambda_n \mathbf{H}_M\right]\hat{\boldsymbol{\theta}}_{\mathrm{MLI}} = \mathbf{0}, \tag{14}$$

where $q_n = n - K_n$ (>0) and λ_n is the (non-negative) smallest root of

$$\left|\frac{1}{n}\mathbf{G}_M - l\frac{1}{q_n}\mathbf{H}_M\right| = 0. \tag{15}$$

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

When $p_{ii}^{(n)}\ (i = 1, \ldots, n)$ are close to c_n or c_n is small, the AOM-LIML estimator is very close to the LIML estimator for practical purposes. Hausman et al. (2007) have defined the HLIM estimator by setting $\mathbf{P}_H = (p_{ij}^*),\ p_{ii}^* = 0\ (i = 1, \ldots, n)$ and replacing $\mathbf{P}_M$ and $\mathbf{Q}_M$ by $\mathbf{P}_H$ and $\mathbf{Q}_H = \mathbf{I}_n - \mathbf{P}_H$ in (13), (14) and (15) but without (12). Then we find that it is not in the class of the AOM-LIML estimation with (12). Numerically, however, the AOM-LIML estimator can be close to the 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. The asymptotic property of the HLIM estimator and its relation to the AOM-LIML estimator shall be discussed at the end of Sect. 3.2.

Our construction of the AOM-LIML estimation includes a class of modified LIML estimators. For instance, if we perturb the latent root of (14) such as $\lambda_n^* = \lambda_n - a/n$ (a is a positive constant), we have the Fuller type modification (Fuller 1977), which is asymptotically equivalent to the AOM-LIML estimator in the first order asymptotic sense because $\sqrt{n}[\lambda_n - \lambda_n^*] \xrightarrow{p} 0$. (See the proof of Theorem 1 in Section 5.) It is also possible to define the corresponding modifications of the TSLS estimator and the GMM estimator as we have constructed the AOM-LIML estimator and the HLIM estimator. An estimation method called JIVE (Jackknife Instrumental Variables Estimator) has been proposed and its properties have been investigated by Chao and Swanson (2005), and also Chao et al. (2009), for instance.

We note that $\boldsymbol{G}_M$ with $\mathbf{P}_M$ should be positive definite (a.s.) 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 more stable than the HLIM estimator in some cases. In the actual computation, we should check if two matrices are in violation of the positive definiteness. Such situations rarely occur in our limited experiments as reported in Sect. 3.2, but we can modify them further in our method without any difficulty if it occurred.

3.2 Asymptotic optimality of AOM-LIML

We shall investigate the asymptotic properties of the AOM-LIML estimator when there are many instruments. One of the attractive features of the AOM-LIML estimator is that it satisfies (12) while it is very similar to the original LIML estimator in many cases. Since we have incidental parameters when the number of instruments grows with the sample size, we cannot apply the standard arguments on the semi-parametric asymptotic optimality and the asymptotic optimality of the AOM-LIML estimation in the proper sense is not obvious.

We have the consistency and the asymptotic normality of the AOM-LIML estimator when the disturbances are heteroscedastic with many instruments under a set of conditions. The proof will be given in Sect. 5.

Theorem 1 *Let* $\mathbf{v}_i\ (i = 1, 2, \ldots, n)$ *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* $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i'|\mathbf{z}_i^{(n)}) = \boldsymbol{\Omega}_i$ *(a.s.) is a bounded function of* $\mathbf{z}_i^{(n)}$*, say,* $\boldsymbol{\Omega}_i[n, \mathbf{z}_i^{(n)}]$*. We also assume that* $\mathcal{E}[\|\mathbf{v}_i\|^4]$ *are bounded. For* (1) *and* (2)*,* *suppose* (**A-I**), (5), (12),

$$\frac{1}{n}\max_{1\le i\le n}\|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})\|^2 \xrightarrow{p} 0, \tag{16}$$

$$\frac{1}{n}\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M - c_*\mathbf{Q}_M)\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}^*, \tag{17}$$

and $\boldsymbol{\Phi}^*$ *is a positive definite (constant) matrix as* $n \to \infty$, $K_n \to \infty$ *and* $q_n \to \infty$, *where* $\boldsymbol{\Pi}_{*n}^{(z)} = (\mathbf{Z}_1, \boldsymbol{\Pi}_{2n}^{(z)}) = (\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})')$. *We denote* $(1/n)\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{P}_M\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_1^*$, $(1/q_n)\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{Q}_M\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_2^*$ *and* $c_* = c/(1-c)$. *Then*

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

where

$$\boldsymbol{\Psi}^* = \boldsymbol{\Phi}^{*-1}\left[\boldsymbol{\Psi}_1^* + \boldsymbol{\Psi}_2^*\right]\boldsymbol{\Phi}^{*-1}, \tag{18}$$

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

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

provided that we have the convergence of $\boldsymbol{\Psi}_1^*$ *and* $\boldsymbol{\Psi}_2^*$ *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* $\mathbf{w}_{2i} = \mathbf{v}_{2i} - u_i(\mathbf{0}, \mathbf{I}_{G_2})\boldsymbol{\Omega}\boldsymbol{\beta}/\sigma^2$ $(i = 1, \ldots, n)$.

Condition (17) is slightly stronger than (**A-II$'$**), but they are equivalent when $c = 0$. The first term of (18) is due to the non-centrality 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 in Sect. 3.3. When $(1/n)\sum_{i=1}^{n}\mathbf{z}_i^{(n)}\mathbf{z}_i^{(n)'} \sim \mathbf{I}_{K_n}$ and $(1/K_n)\mathbf{z}_i^{(n)'}\mathbf{z}_j^{(n)} \sim 0$ for all $i \neq j$ $(= 1, \ldots, n)$, for instance, $\boldsymbol{\Phi}^*$ could be singular. Since these conditions are extreme and often unrealistic, the asymptotic distribution in Theorem 1 gives a reasonable approximation to the finite sample distribution of the AOM-LIML estimator as we shall indicate in Sect. 3.3.

When (2) is linear, we have (3) and (4), and then we partition the $(K_1 + K_{2n}) \times (1 + G_2)$ coefficient matrix as

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

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

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

and assume the condition (**A-VI**). Then by setting $p_{ij}^{(m)} = p_{ij}^{(n)}$ $(i, j = 1, \ldots, n)$, that is the LIML estimator, we find that $\boldsymbol{\Phi}_2^* = \mathbf{O}$ and

$$\boldsymbol{\Psi}_1^* = \sigma^2 \operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{\Pi}_{22}^{(n)'} \boldsymbol{A}_{22.1} \boldsymbol{\Pi}_{22}^{(n)} = \sigma^2 \boldsymbol{\Phi}_{22.1},$$

which is non-singular when we have (**A-II**$'$) and $\boldsymbol{A}_{22.1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$. In this case we have

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

and $[\,\cdot\,]_{22}$ is the $G_2 \times G_2$ right-lower corner of the corresponding $(1+G_2)\times(1+G_2)$ 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 $\sum_{i,j=1}^{n} p_{ij}^{(n)2} = \sum_{i=1}^{n} p_{ii}^{(n)} = K_n$. When the disturbance terms are homoscedastic, we have $\mathcal{E}[\mathbf{w}_{2i}u_i] = \mathbf{0}$ and then the second term of $\boldsymbol{\Psi}_2^*$ is zero. Hence the right-lower corner of $\boldsymbol{\Psi}_2^*$ is reduced to

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

Then $\boldsymbol{\Psi}^*$ in (18) corresponds to

$$\boldsymbol{\Psi}_A^* = \sigma^2\boldsymbol{\Phi}^{*-1} + c_*\boldsymbol{\Phi}^{*-1}[\mathbf{O}, \mathbf{I}_{G_2}]'[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega}]_{22}[\mathbf{O}, \mathbf{I}_{G_2}]\boldsymbol{\Phi}^{*-1}, \qquad (19)$$

where $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$ and $c_* = c/(1-c)$. We find that (19) reduces to (3.8) of Theorem 2 in Anderson et al. (2010).

For the estimation of the parameters in the structural equation of interest $\boldsymbol{\theta}$ $(= (-\boldsymbol{\gamma}_1', 1, -\boldsymbol{\beta}_2')')$, it may be natural to investigate the estimation 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 heteroscedasticity) 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) for instance. 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 will give the proof in Sect. 5.

Theorem 2 *Assume that* (1) *and* (2) *hold and the regularity conditions of Theorem 1. Define the class of consistent estimators for* $(\boldsymbol{\gamma}_1', \boldsymbol{\beta}_2')'$ *by*

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \phi\left(\frac{1}{n}\mathbf{G}_M,\ \frac{1}{q_n}\mathbf{H}_M\right), \tag{20}$$

where ϕ *is continuously differentiable and its derivatives are bounded at the probability limits of random matrices in* (20) *as* $K_{2n} \to \infty$ *and* $n \to \infty$ *and* $0 \le c < 1$. *Then under the assumptions of Theorem* 1,

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

where

$$\boldsymbol{\Psi} \ge \boldsymbol{\Psi}^* \tag{21}$$

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

When the conditions (WH$'$) and (**A-VI**) are satisfied, the result of Theorem 2 corresponds to an extension of Theorem 4 of Anderson et al. (2010). When the equations of (2) are linear, the disturbances are normally distributed with the homoscedastic disturbances and the instrumental variables are deterministic,

$$\begin{aligned} \mathbf{I}_n(\boldsymbol{\theta}) &= \frac{1}{\sigma^2}\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{P}_Z\boldsymbol{\Pi}_{*n}^{(z)} \\ &\sim \frac{1}{\sigma^2}\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{P}_M\boldsymbol{\Pi}_{*n}^{(z)} + \frac{1}{\sigma^2}\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{D}_n - c_n\mathbf{I}_n)\boldsymbol{\Pi}_{*n}^{(z)}, \end{aligned} \tag{22}$$

which corresponds to the Fisher information and $\mathbf{D}_n = \mathrm{diag}(\mathbf{P}_Z)$.

Hence, the condition (**A-VI**) or the classical case when $c = 0$ in (**A-I**) in the linear models is the sufficient condition such that we do not lose the information amount essentially by modifying the LIML estimation, because the asymptotic loss of information is of higher order asymptotically. If they were not satisfied, the AOM-LIML estimator has possible information loss asymptotically although it is still consistent and it has the asymptotic normality. Theorems 1 and 2 imply that the AOM-LIML estimator has an asymptotic optimality even when the number of instruments is fixed (i.e. $c = 0$) because the information loss is quite small.

Also Anderson et al. (2010) 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 $\hat{\boldsymbol{\theta}}_{\text{HLI}}$ $(= (-\hat{\boldsymbol{\gamma}}'_{1.\text{HLI}}, 1, -\hat{\boldsymbol{\beta}}'_{2.\text{HLI}})')$ be the HLIM estimator defined by Hausman et al. (2007). Then it is possible to show that the HLIM estimator cannot be improved asymptotically further. Since Hausman et al. (2007) have investigated its asymptotic properties in detail, our derivation shall be brief and we focus on the asymptotic equivalence of two estimation methods.

Theorem 3 *Let* $\mathbf{P}_H = (p^*_{ij})$ *such that* $p^*_{ii} = 0$, $p^*_{ij} = p^{(n)}_{ij}$ $(i \neq j; i, j = 1, \cdots, n)$, $\mathbf{Q}_H = \mathbf{I}_n - \mathbf{P}_H$ *in* (13), (14) *and* (15) *instead of* $\boldsymbol{P}_M$ *and* $\mathbf{Q}_M$. *Suppose* (**A-I**), (7),

$$\frac{1}{n}\boldsymbol{\Pi}^{(z)\prime}_{*n}(\mathbf{P}_Z - \mathbf{D}_n)\boldsymbol{\Pi}^{(z)}_{*n} \xrightarrow{p} \boldsymbol{\Phi}^*_D \tag{23}$$

and $\boldsymbol{\Phi}^*_D$ *is a positive definite (constant) matrix as* $n \to \infty$ *and* $K_n \to \infty$, *where* $\mathbf{D}_n = \text{diag}(\mathbf{P}_Z)$. *Also suppose that* $\mathcal{E}[\|\mathbf{v}_i\|^2|\mathbf{z}^{(n)}_i]$ *and* $\mathcal{E}[\|\mathbf{v}_i\|^4]$ *are bounded. Then*

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

where $\boldsymbol{\Psi}^*$ *is given by* (18).

This result has some interesting aspects because the HLIM estimator does not use $\mathbf{P}_M$ and $\mathbf{Q}_M$ explicitly. As we shall show in Sect. 5 [for instance (36) and (37)] that the scaling factor (the denominator of the asymptotic variance) $\boldsymbol{\Phi}^*_D = (1-c)\boldsymbol{\Phi}^*$ is strictly smaller than $\boldsymbol{\Phi}^*$, which is the scaling factor of the AOM-LIML estimator, when $0 < c < 1$ while the effects on the asymptotic variance are cancelled out asymptotically at the end. (Since $c_* = c/(1-c)$, we have $(1 + c_*)^{-1} = (1 - c)$ in (37) of Section 5.) Theorem 3 together with Theorem 2 implies that the HLIM (or JLIML) estimation cannot be improved asymptotically in a class of estimators which depend on the functions of $\mathbf{G}_M$ and $\mathbf{Q}_M$ with some $\mathbf{P}_H$ and $\mathbf{Q}_H$.

Anderson et al. (2010) have shown that the LIML estimator dominates the bias-corrected TSLS estimator asymptotically, for instance, when $0 < c < 1$ under the condition of (WH). The Jackknife Instrumental Variables Estimators such as the Jackknife-TSLS estimator, except the LIML type, cannot be efficient asymptotically because they cannot attain the asymptotic bound under (PH). Since both AOM-LIML and HLIM estimators attain the same asymptotic bound, they cannot be improved in the class of estimators (14) with $\mathbf{G}_M$ and $\mathbf{H}_M$ ($\mathbf{P}_M$ and $\mathbf{Q}_M$ or $\mathbf{P}_H$ and $\mathbf{Q}_H$) in the sense of their asymptotic distributions in the large-K_2 asymptotics.

3.3 On alternative parameter sequences

It is possible to extend our results on the asymptotic optimality of the LIML estimation to alternative parameter sequences, which correspond to alternative different (practical) situations. For this purpose, we consider the linear model (1) and (3) and we define the non-centrality parameter

$$d_{*n}^2 = \mathrm{tr}[\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M - c_*\mathbf{Q}_M)\boldsymbol{\Pi}_{*n}^{(z)}].$$

Then we can consider alternative parameter sequences and the corresponding asymptotic theories. (In our notations d_{*n} is the analogous quantity to d_n at the beginning of Sect. 3.1.) When $c = 0$ and $d_{*n} \sim \sqrt{n}$ (or it is of greater order than $\sqrt{n}$), we may call *the classical asymptotic sequence*. When $c > 0$ and $d_{*n} \sim \sqrt{n}$, we have the asymptotic sequence of *many instruments*, which has been reported in Sect. 3.2. These cases in the linear models with θ_n and d_n in Sect. 3.1 have been investigated by Anderson et al. (2010) in a systematic way under the assumption of homoscedastic disturbances.

Although it may be possible to develop our analysis to alternative asymptotic sequences, we shall illustrate one important result in this subsection. We consider the linear model (1) and (3) when the normalizing factor $d_{*n} = o_p(n^{1/2})$ and $\sqrt{n}/d_{*n}^2 \xrightarrow{p} 0$, which may correspond to the case of *many weak instruments* with the persistently heteroscedastic disturbances. Then we still have the asymptotic optimality result even in this situation. Since the proof is quite similar to that of Theorem 5 in Anderson et al. (2010), we shall give a brief derivation in Sect. 5.

Theorem 4 *Consider the linear model of* (1) *and* (3). *Suppose* (**A-I**) *and* (5) *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*

$$\frac{1}{d_{*n}^2}\max_{1\le i\le n}\|\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})\|^2 \xrightarrow{p} 0,$$
$$\frac{1}{d_{*n}^2}\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M - c_*\mathbf{Q}_M)\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}^{**}$$

and $\boldsymbol{\Phi}^{**}$ is a positive definite (constant) matrix as $n \to \infty$, $K_n \to \infty$ and $q_n \to \infty$, where $\boldsymbol{\Pi}_{*n}^{(z)} = (\mathbf{Z}_1, \boldsymbol{\Pi}_{2n}^{(z)}) = (\boldsymbol{\pi}_{*i}(\mathbf{z}_i^{(n)})')$. Also suppose that $\mathcal{E}[\|\mathbf{v}_i\|^2|\mathbf{z}_i^{(n)}]$ and $\mathcal{E}[\|\mathbf{v}_i\|^4]$ are bounded. We use the notations $(1/d_{*n}^2)\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{P}_Z^*\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_1^*$, $(1/d_{*n}^2)$ $\boldsymbol{\Pi}_{*n}^{(z)'}\mathbf{Q}_Z^*\boldsymbol{\Pi}_{*n}^{(z)} \xrightarrow{p} \boldsymbol{\Phi}_2^*$ and $c_* = c/(1-c)$. As $n \to \infty$, for the AOM-LIML estimator,

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

and for any estimator $\hat{\boldsymbol{\theta}}$ in the class of (20),

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

where $\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^{**}$,

$$\boldsymbol{\Psi}^{**} = \boldsymbol{\Phi}^{*-1}\boldsymbol{\Psi}_2^{**}\boldsymbol{\Phi}^{*-1}, \tag{24}$$

and

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

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

The variance of the limiting distribution of the AOM-LIML estimator [i.e. (24)] is simpler than (18) because the effects of n dominate the first term of (18) in Theorem 1. This result illustrates that the AOM-LIML estimation has the asymptotic optimality in a wide range of alternative (such as the non-centrality and the normalizing factor) parameter sequences.

3.4 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 et al. (2005, 2008) in a systematic way. As an example, we present only three figures (Figs. 1, 2, 3) in Appendix when we have the linear structural equation model with (3), (4) and $G_2 = 1$ for simplicity. We take a typical case of many instruments when K_2 (or K_{2n}) is relatively large. Here we have used the method of numerical evaluation for the cumulative distribution function (cdf) of the LIML estimator based on the simulation. We have enough numerical accuracy in most cases by using the same simulation setting except for the factor of heteroscedasticity of disturbances, which has been explained by Anderson et al. (1982) and Anderson et al. (2005, 2008) in more detail.

We first generate a sequence of independent random vectors $\mathbf{v}_i$ and $\mathbf{z}_i^{(n)}$ $(i = 1, \ldots, n)$, and then generate a large number of empirical distributions for the normalized forms of alternative estimators by fixing the several *key parameters* in the structural equation of interest with (1) and (3). It is important to mention that the normalization makes the numerical comparison of the finite sample distributions of alternative estimators quite accurate even if the finite moments of some estimators do not necessarily exist. (The standard use of MSE of estimators in simulations is not necessarily meaningful in such cases.) We take the key parameters under the homoscedastic disturbances ($\boldsymbol{\Omega} = \boldsymbol{\Omega}_i$) and the linear model of (1) and (3) as the benchmark case, and we control the effects of the key parameters K_2 (or K_{2n}), $n - K$ (or $n - K_n$), $\alpha = [\omega_{22}/|\boldsymbol{\Omega}|^{1/2}](\beta_2 - \omega_{12}/\omega_{22})$ (the standardized coefficient) with ($\boldsymbol{\Omega} = (\omega_{ij})$) and $\delta^2 = \boldsymbol{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}\boldsymbol{\Pi}_{22}^{(n)}/\omega_{22}$ (the non-centrality) when $G_2 = 1$. See Anderson et al. (1982, 2005, 2008) for the details of the computation method of simulations including the various cases of non-normal disturbances and the precise parameterization of key quantities in figures.

As a simple example of the LIML modification, we set $G_2 = K_1 = 1$, $\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{I}_K)$ $(i = 1 \cdots, n)$, $\mathbf{v}_i|\mathbf{z}_i^{(n)} \sim N(\mathbf{0}, \boldsymbol{\Omega}_i)$ $(i = 1, \ldots, n)$. In order to make comparison easier, we concentrate the effects of $\mathbf{Z}_1$ in (14) and (15), and use the representation

$$\left[\frac{1}{n}\mathbf{Y}'\mathbf{P}_{2.1}^{(m)}\mathbf{Y} - \lambda_n \frac{1}{q_n}\mathbf{Y}'\mathbf{Q}_Z^{(m)}\mathbf{Y}\right]\hat{\boldsymbol{\beta}} = \mathbf{0}\,, \tag{25}$$

where $\hat{\boldsymbol{\beta}} = (1, -\hat{\beta}_2)'$, $(\mathbf{P}_{2.1}^{(m)})_{ij} = (\mathbf{P}_{2.1})_{ij}\ (i \neq j)$, $(\mathbf{Q}_Z^{(m)})_{ij} = (\mathbf{Q}_Z)_{ij}\ (i \neq j)$,

$$(\mathbf{P}_{2.1}^{(m)})_{ii} = 1 - (\mathbf{Q}_Z^{(m)})_{ii} = \frac{K_n}{n} + \left[\frac{z_{1i}^2}{\sum_{j=1}^n z_{1j}^2} - \frac{1}{n}\right]$$

and z_{1i} $(i = 1, \ldots, n)$ is the included exogenous variable in the structural equation of interest. We have taken this setting mainly because $p_{ii}^{(m)}$ are not exactly the same as K/n (or K_n/n) and we allow some variability. Then we have investigated the finite sample properties of $\hat{\beta}_2$ for the coefficient β_2 by the simulated $\hat{\beta}_2$ with the LIML, HLIM and AOM-LIML estimators. The number of replications of the simulations are 10,000.

Figures 1 and 2 correspond to the homoscedastic disturbance cases, while Fig. 3 corresponds to the case of persistent heteroscedasticity which is quite similar to the one reported by Hausman et al. (2007). (We have tried to replicate their example.) Three figures in Appendix show the estimated cdf of estimators in the standard form, i.e.

$$\sqrt{n}[\boldsymbol{\Psi}]_{22}^{*-1/2}(\hat{\beta}_2 - \beta_2), \tag{26}$$

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 in the form of (26) is $N(0, 1)$ in the large-K_2 asymptotics and it is denoted by "o".

From these three 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 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 AOM-LIML and HLIM estimators considerably as well as the standard normal distribution, because the effects of correlation between $p_{ii}(\mathbf{Z})$ and $\boldsymbol{\Omega}_i$ $(i = 1, \ldots, n)$ do not decrease as K_n and n increase. In this case, the AOM-LIML estimator with (14) and (15) improves both the LIML and HLIM estimators in the finite samples. These observations on the finite sample properties of alternative estimators agree with our theoretical results of Sect. 3.2, although we have covered only some cases among many possibilities.

4 Concluding remarks

In this paper, we have introduced a class of modifications of the LIML estimation methods and investigated their statistical properties. We have excluded the GMM type estimation methods in our consideration because they are badly biased when there are many instruments. (See Anderson et al. 2005, 2008, 2010; Kunitomo and Matsushita 2009.) When there are many instruments and the disturbances have heteroscedasticity, it might be argued that the LIML estimator does lose good asymptotic properties in the extremely heteroscedastic cases. However, as we have shown, a simple modification of the LIML estimation, called the AOM-LIML estimator, gives 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 have also shown that the AOM-LIML estimator improves the LIML estimator, and the HLIM (or JLIML) estimator is asymptotically equivalent to a simple case of the AOM-LIML estimator when there are many instruments and the persistent heteroscedasticity in disturbances exists at the same time. Although these estimators are asymptotically the same in the sense of the first-order large$-K_2$ asymptotics, there are some differences in the finite samples.

Several important issues still remain for further investigations. For the more general non-linear estimating equation models with (4), the non-linear 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 $u_i(\boldsymbol{\theta}) = y_{1i} - \boldsymbol{\gamma}_1' \mathbf{z}_{1i} - \boldsymbol{\beta}_2' \mathbf{y}_{2i}$ $(i = 1, \ldots, n)$ and minimizing the variance ratio in (11), where $f_i(\,\cdot\,)$ is a known function and $\boldsymbol{\theta}$ is the vector of unknown (structural) parameters. Then our method would be extended to non-linear structural equations with some notational complications. When the number of restrictions or the dimension becomes large with the sample size, however, semi-parametric methods such as the GMM and the maximum empirical likelihood (MEL) estimation may have some difficulty in theory as well as in practical computation.

Although we have incidental parameters when the number of instruments grows with the sample size and we cannot apply the standard arguments on the semi-parametric asymptotic optimality, we have shown that the AOM-LIML estimator has the asymptotic optimality in a class of estimators. There is an interesting topic on the higher-order efficiency of estimation, which is under current investigation.

Finally, a more practical question is the relevance of *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. (2010), which are often straightforward.

Proof of Lemma 1 Let $\mathbf{Z} = (\mathbf{z}_i^{(n)'})$ and $\mathbf{A}_{n(i)} = \sum_{j=1, j\neq i}^{n} \mathbf{z}_j^{(n)} \mathbf{z}_j^{(n)'}$, where $\mathbf{z}_i^{(n)'}$ are $K_n \times 1$ vectors. The assumption implies that $\mathbf{A}_{n(i)}$ is non-singular. Then

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

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

Proof of Theorem 1 The proof consists of two steps.
Step I This step develops a convenient representation of the normalized estimator, which is asymptotically equivalent to the one of the AOM-LIML estimator. (The basic arguments are quite similar to the ones in Anderson et al. (2010).)

From (1) and (2) we write $\mathbf{Y} = \mathbf{\Pi}_n^{(z)} + \mathbf{V}$, $\mathbf{\Pi}_n^{(z)} = (\mathbf{\Pi}_{1n}^{(z)}, \mathbf{\Pi}_{2n}^{(z)})$ 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

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

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

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

where $\bar{\mathbf{\Omega}} = (1/n)\sum_{i=1}^{n} \mathbf{\Omega}_i$. By using (17) and the fact that $\mathbf{\Phi}_1^*$ is positive definite and then as $n \longrightarrow \infty$

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

and

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

Then as $n \longrightarrow \infty$,

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

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

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

By using (5) and (17),

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

and

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

Then (17) implies

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

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

$$l_n = \frac{\boldsymbol{\theta}'\frac{1}{n}\mathbf{G}_M\boldsymbol{\theta}}{\boldsymbol{\theta}'\frac{1}{n}\mathbf{H}_M\boldsymbol{\theta}} \xrightarrow{p} \frac{\boldsymbol{\theta}'\mathbf{G}_0\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{H}_0\boldsymbol{\theta}}$$

and the minimum of the right-hand side is c under the condition (17), hence we have that $\text{plim}\lambda_n = c$ is the unique solution and $\hat{\boldsymbol{\theta}}_{\text{MLI}} \xrightarrow{p} \boldsymbol{\theta}$ as $n \to \infty$ because of (14) and (15).

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

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

$\mathbf{H}_1 = \sqrt{q_n}(\frac{1}{q_n}\mathbf{H} - \mathbf{H}_0)$, $\lambda_{1n} = \sqrt{n}(\lambda_n - c)$ and $\mathbf{b}_1 = \sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{MLI}} - \boldsymbol{\theta})$. From (14) and (15), we have

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

Since $(\mathbf{G}_0 - c^*\,\mathbf{H}_0)\boldsymbol{\theta} = \mathbf{0}$, the above equation gives

$$\mathbf{B}'\boldsymbol{\Phi}^*\sqrt{n}\left[\begin{pmatrix}\hat{\boldsymbol{\gamma}}_{1.\mathrm{MLI}}\\ \hat{\boldsymbol{\beta}}_{2.\mathrm{MLI}}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right] = \left(\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1\right)\boldsymbol{\theta} + o_p(1). \quad (27)$$

Multiplication of (27) 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 (27) 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 the substitution for λ_{1n} in (27) yields

$$\begin{aligned}
&\sqrt{n}\left[\begin{pmatrix}\hat{\boldsymbol{\gamma}}_{1.\mathrm{MLI}}\\ \hat{\boldsymbol{\beta}}_{2.\mathrm{MLI}}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right]\\
&= \boldsymbol{\Phi}^{*-1}\mathbf{J}'(\mathbf{G}_1 - \lambda_{1n}\mathbf{H}_0 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\theta} + o_p(1)\\
&= \boldsymbol{\Phi}^{*-1}\mathbf{J}'\left[\mathbf{I}_{K_1+G_2+1} - \frac{\mathbf{H}_0\boldsymbol{\theta}\boldsymbol{\theta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\right](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\theta} + o_p(1). \quad (28)
\end{aligned}$$

By using the relation $\mathbf{V}\boldsymbol{\beta} = \mathbf{u}$ and $\mathbf{H}_0\boldsymbol{\theta} = \boldsymbol{\Omega}^*\boldsymbol{\theta}$, the vector of $\boldsymbol{\Phi}^*$ times the last term of (28) can be written as

$$\begin{aligned}
&\mathbf{J}'\left[\mathbf{I}_{K_1+G_2+1} - \frac{\mathbf{H}_0\boldsymbol{\theta}\boldsymbol{\theta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\right](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\theta}\\
&= \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M - c_*\mathbf{Q}_M)\mathbf{u} + \sqrt{c}\frac{1}{\sqrt{K_n}}\mathbf{J}'\left[\begin{pmatrix}\mathbf{O}\\ \mathbf{V}'\end{pmatrix}\mathbf{P}_M\mathbf{u} - K_n\begin{pmatrix}\mathbf{O}\\ \mathbf{I}_{G_2}\end{pmatrix}\bar{\boldsymbol{\Omega}}\boldsymbol{\beta}\right]\\
&\quad -\sqrt{cc_*}\frac{1}{\sqrt{q_n}}\mathbf{J}'\left[\begin{pmatrix}\mathbf{O}\\ \mathbf{V}'\end{pmatrix}\mathbf{Q}_M\mathbf{u} - q_n\begin{pmatrix}\mathbf{O}\\ \mathbf{I}_{G_2}\end{pmatrix}\bar{\boldsymbol{\Omega}}\boldsymbol{\beta}\right],
\end{aligned}$$

where $K_n + q_n = n$. We define a $(K_1 + G_2) \times n$ matrix

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

and

$$\mathbf{w}_{*i} = [\mathbf{0}, \mathbf{I}_{G_2}]\left[\mathbf{I}_{K_1+G_2} - \begin{pmatrix}\mathbf{O}\\ \mathbf{I}_{G_2}\end{pmatrix}(\mathbf{0}, \mathbf{I}_{G_2})\frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}\right]\mathbf{v}_i.$$

Then (28) is further rewritten as

$$\sqrt{n}\left[\begin{pmatrix}\hat{\boldsymbol{\gamma}}_{1.\mathrm{MLI}}\\ \hat{\boldsymbol{\beta}}_{2.\mathrm{MLI}}\end{pmatrix}-\begin{pmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right]=\boldsymbol{\Phi}^{*-1}\frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M-c_*\mathbf{Q}_M)\mathbf{u}$$
$$+\boldsymbol{\Phi}^{*-1}\frac{1}{\sqrt{n}}[\mathbf{W}'(\mathbf{P}_M-c_*\mathbf{Q}_M)\mathbf{u}]+o_p(1). \quad (29)$$

(The last representation has a convenient form in the sense that the first term is a linear combination of u_i $(i=1,\ldots,n)$ and the second term is a cross quadratic form of u_i and $\mathbf{w}_i$ $(i=1,\ldots,n)$. If $\mathbf{v}_i$ $(i=1,\ldots,n)$ are homoscedastic and normally distributed, u_i and $\mathbf{w}_i$ are the sequence of independent random variables.)

Step II 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 and Lemmas in Anderson et al. (2010). The most important step is to apply an appropriate martingale central limit theorem (MCLT) to each terms of (29) in *Step I*.

By following *Lemma 3* of Anderson et al. (2010), for any (conformable) vector $\mathbf{a}$ we set $t_{1i}^{(n)}=(\mathbf{a}'\boldsymbol{\Pi}_{*n}^{(z)'}(\mathbf{P}_M-c_*\mathbf{Q}_M))_i$ (ith element) and $t_{2i}^{(n)}=\mathbf{a}'\mathbf{w}_{*i}$ $(i=1,\ldots,n)$. Also for the $n\times n$ matrix $\mathbf{B}=\mathbf{P}_M-c_*\mathbf{Q}_M$ $(=(b_{ij}))$ we set $s_{1i}^{(n)}=(1/\sqrt{n})t_{1i}^{(n)}u_i$, $s_{2i}^{(n)}=(1/\sqrt{n})t_{2i}^{(n)}u_ib_{ii}$, $s_{3i}^{(n)}=(1/\sqrt{n})u_i\sum_{j=1}^{i-1}t_{2j}^{(n)}b_{ji}$, $s_{4i}^{(n)}=(1/\sqrt{n})t_{2i}^{(n)}\sum_{j=1}^{i-1}u_jb_{ij}$ and $b_{ij}=p_{ij}^{(m)}-c_*(\delta_i^j-q_{ij}^{(m)})$ $(i,j=1,\ldots,n)$. Let $\mathcal{F}_{n,i}$ be the σ-field generated by the random variables $u_j,\mathbf{v}_j$ $(1\le j\le i, 1\le i\le n)$ and $\mathcal{F}_{n,0}$ be the initial σ-field. Because a sequence of $\mathbf{z}_i^{(n)}$ are the instrumental variables and we consider the case of $(\mathbf{y}_i,\mathbf{z}_i^{(n)})$ being independent with respect to i, we only need to investigate the situation when $\mathbf{z}_i^{(n)}\in\mathcal{F}_{n,0}$. By construction we have $\mathcal{F}_{n,i-1}\subseteq\mathcal{F}_{n,i}$ for $i=1,\ldots,n$. Then $T_n=\sum_{i=1}^n X_{ni}$ can be decomposed $X_{ni}=s_{1i}^{(n)}+s_{2i}^{(n)}+s_{3i}^{(n)}+s_{4i}^{(n)}$ and $\mathcal{E}[X_{ni}|\mathcal{F}_{n,i-1}]=0$ $(i=1,\ldots,n)$. Since each term X_{ni} $(i=1,\ldots,n)$ are martingale difference sequences, by direct calculations we find

$$\mathcal{E}[X_{ni}^2|\mathcal{F}_{n,i-1}]=\frac{1}{n}[t_{1i}^{(n)}]^2\sigma_i^2+\frac{1}{n}b_{ii}^2\mathcal{E}_{i-1}[t_{2i}^{(n)}u_i]^2$$
$$+\frac{1}{n}\sigma_i^2\left[\sum_{j=1}^{i-1}b_{ij}t_{2j}\right]^2$$
$$+\frac{1}{n}\mathcal{E}_{i-1}(t_{2i}^2)\left[\sum_{j=1}^{i-1}b_{ij}u_j\right]^2+\frac{2}{n}t_{1i}^{(n)}b_{ii}\mathcal{E}_{i-1}[t_{2i}^{(n)}u_i^2]$$
$$+\frac{2}{n}\sigma^2t_{1i}^{(n)}\sum_{j=1}^{i-1}b_{ij}t_{2j}+\frac{2}{n}\mathcal{E}_{i-1}(u_i^2t_{2i})b_{ii}\sum_{j=1}^{i-1}b_{ij}t_{2j}$$

$$+\frac{2}{n}\mathcal{E}_{i-1}(u_i t_{2i}^2)b_{ii}\sum_{j=1}^{i-1}b_{ij}u_j + \frac{2}{n}\mathcal{E}_{i-1}(u_i t_{2i})\left(\sum_{j=1}^{i-1}b_{ij}u_j\right)$$
$$\times\left(\sum_{j'=1}^{i-1}b_{ij'}t_{2,j'}\right),$$

which can be further simplified by using the relations $b_{ii} = o_p(1)$ $(i = 1, \ldots, n)$ in the present case, where we use the notation $\mathcal{E}_{i-1}[t_{2i}^2] = \mathcal{E}[t_{2i}^2|\mathcal{F}_{n,i-1}]$, for instance. Because of the conditional heteroscedasticities and $\mathcal{E}_{i-1}[u_i t_{2i}]$ are not necessarily zeros, we need to evaluate the last term. As a result, we can derive the second term of $\boldsymbol{\Psi}_2^*$, which was absent in the homoscedasticity case in Theorem 1 of Anderson et al. (2010). (It is zero in the homoscedastic case because u_i and $\mathbf{w}_i$ are independent random variables.)

We evaluate each of the terms of X_{ni} and $\mathcal{E}_{i-1}(X_{ni})$, and then apply an MCLT to $T_n = \sum_{i=1}^n X_{ni}$. As an illustration, an important step is to show

$$\frac{1}{n}\sum_{i=2}^{n}\left[\left(\sum_{j=1}^{i-1}b_{ij}t_{2j}\right)^2 - \mathcal{E}\left(\sum_{j=1}^{i-1}b_{ij}t_{2j}\right)^2\right]\xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=2}^{n}\left[\mathcal{E}_{i-1}(t_{2i}^2)\left(\sum_{j=1}^{i-1}b_{ij}u_j\right)^2 - \mathcal{E}_{i-1}(t_{2i}^2)\mathcal{E}\left(\sum_{j=1}^{i-1}b_{ij}u_j\right)^2\right]\xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=2}^{n}\left[\mathcal{E}_{i-1}(u_i t_{2i})\left(\sum_{j=1}^{i-1}b_{ij}u_j\right)\left(\sum_{j'=1}^{i-1}b_{ij'}t_{2,j'}\right), -\mathcal{E}_{i-1}(u_i t_{2i})\sum_{j=1}^{i-1}b_{ij}^2\mathcal{E}_{j-1}(u_j t_{2,j})\right]$$
$$\xrightarrow{p} 0.$$

Under the assumptions in Theorem 1, it is straightforward but quite tedious (due to many terms involved) to show the required conditions for the MCLT. This is possible mainly due to the fact that we have assumed enough moment conditions, and the original $\mathbf{P}_Z$ and $\mathbf{Q}_Z$ are projection matrices such that the components of b_{ij} have some properties as discussed in Lemmas 1 and 2 of Anderson et al. (2010).

We set $V_n = \sum_{i=1}^n \mathcal{E}[X_{ni}^2|\mathcal{F}_{n,i-1}]$. Also, we note that by utilizing that for any $\xi > 0$ and some $\nu > 0$, $\sum_{i=1}^n \mathcal{E}[(X_{ni})^2 I(|X_{ni}| \geq \xi)] \leq (1/\xi)^\nu \sum_{i=1}^n \mathcal{E}[X_{ni}^{2+\nu}]$, and then we apply Theorem 3.5 of Hall and Heyde (1980) as the relevant MCLT. The only remaining task is to show their conditions (i) $\max_{1\leq i\leq n}\mathcal{E}[X_{ni}^2|\mathcal{F}_{n,i-1}] \xrightarrow{p} 0$ and (ii) for any $\xi > 0$ $\sum_{i=1}^n \mathcal{E}[X_{ni}^2 I(|X_{ni}| \geq \xi)] \to 0$ (as $n \to \infty$) under the assumptions of Theorem 1. We note that in the original version of Theorem 3.5, the martingale $T_n = \sum_{i=1}^n X_{ni}$ have been normalized by the unconditional variance $\mathcal{E}[T_n^2]$, but it does not change the required conditions because the conditional variance V_n converges to a (non-degenerate) positive constant in probability by using (17) in the present situation.

We have omitted the details of the derivations here because we need to use quite similar arguments repeatedly as in Anderson et al. (2010), but we take care of the conditional heteroscedasticities of the associated disturbances. Since we can utilize the boundedness of the conditional covariance and enough moment conditions on the disturbances, they are straightforward, but can be quite lengthy. We have some simplifications because we can use $\mathbf{P}_M$ and $\mathbf{Q}_M$, instead of $\mathbf{P}_Z$ and $\mathbf{Q}_Z$, with (14) and (15) to derive the asymptotic properties of the AOM-LIML estimator. By utilizing the construction of the diagonal parts of $\mathbf{P}_M$ and $\mathbf{Q}_M$, we have the results. □

The next proof of Theorem 2 is the modification of the proof of Theorem 4 of Anderson et al. (2010) and we shall use their notations and arguments. For the sake of completeness we give its proof.

Proof of Theorem 2 We first derive the necessary conditions for the consistency of the class of estimators. Then we shall develop its linearized representation with the restrictions of consistency. Finally, we shall derive the conditions to minimize the asymptotic covariance matrix of the class of linearized estimators.

We set $p = K_1 + 1 + G_2$ and the vector of true parameters $\boldsymbol{\theta}' = (-\boldsymbol{\gamma}_1', 1, -\boldsymbol{\beta}_2') = (-\theta_1, \ldots, -\theta_{K_1}, 1, -\theta_{K_1+1}, \ldots, -\theta_{K_1+G_2})$ and $\boldsymbol{\theta}_2' = (\theta_1, \ldots, \theta_{K_1+G_2})$. The estimator of the vector of coefficients $\boldsymbol{\theta}$ can be written as

$$\hat{\theta}_k = \phi_k\left(\frac{1}{n}\mathbf{G}_M, \frac{1}{q_n}\mathbf{H}_M\right) \quad (k = 1, \ldots, p-1). \tag{30}$$

with $p - 1 = K_1 + G_2$.

We use a $(p-1) \times p$ matrix $\mathbf{B}$ and a $p \times p$ matrix $\boldsymbol{\Omega}^* = (\omega_{ij}^*)$ in the proof of Theorem 1. For the estimator to be consistent, we need the condition

$$\theta_k = \phi_k[\mathbf{B}'\boldsymbol{\Phi}_1^*\mathbf{B} + c\ \boldsymbol{\Omega}^*, \mathbf{B}'\boldsymbol{\Phi}_2^*\mathbf{B} +\ \boldsymbol{\Omega}^*] \tag{31}$$

for $k = 1, \ldots, p-1$ as the identities with respect to each of the components of $\boldsymbol{\gamma}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\Phi}_g^* = (\varphi_{ij}^{(g)})$ $(g = 1, 2)$ and $\boldsymbol{\Omega}^*$. We set a $p \times p$ matrix

$$\mathbf{T}^{(k)} = \left(\frac{\partial \phi_k}{\partial g_{ij}}\right) = (\tau_{ij}^{(k)}) \quad (k = 1, \ldots, p-1;\ i, j = 1, \ldots, p)$$

evaluated at the probability limits. (They are bounded by assumptions.) We define $p \times p$ matrices Θ_g $(= (\theta_{ij}^{(g)}))$ by $\Theta_g = \mathbf{B}'\boldsymbol{\Phi}_g^*\mathbf{B}$ $(g = 1, 2)$.

Next, we consider the role of the second matrix in (30). By differentiating (31) with respect to ω_{ij}^* $(i, j = 1, \ldots, p)$, we have the condition

$$c\frac{\partial \phi_k}{\partial g_{ij}} = -\frac{\partial \phi_k}{\partial h_{ij}} \ (k = 1, \ldots, p-1; i, j = 1, \ldots, p) \tag{32}$$

evaluated at the probability limit. (We use the notation that the partial derivatives are taken with respect to each of the elements of two matrices in (30) and (31).) By differentiating each of the components of ϕ_k $(k = 1, \ldots, p-1)$ with respect to θ_i $(i = 1, \ldots, p-1)$, we have

$$\frac{\partial \phi_k}{\partial \theta_i} = \sum_{g,h=1}^{p} \left[\frac{\partial \phi_k}{\partial g_{gh}} \frac{\partial g_{gh}}{\partial \theta_i} + \frac{\partial \phi_k}{\partial h_{gh}} \frac{\partial h_{gh}}{\partial \theta_i} \right]$$
$$= \sum_{g,h=1}^{p} \frac{\partial \phi_k}{\partial g_{gh}} \left[\frac{\partial g_{gh}}{\partial \theta_i} - c \frac{\partial h_{gh}}{\partial \theta_i} \right].$$

For the notational convenience, we set $\tau_{**}^{(k)} = \tau_{K_1+1,K_1+1}^{(k)}$ $(k = 1, \ldots, p-1)$. Then by differentiating each of the terms of (31) with respect to θ_i $(i = 1, \ldots, p-1)$ [as (6.47) and (6.48) in Anderson et al. 2010] and rearranging the resulting terms, we can express

$$\begin{aligned} &\operatorname{tr}\left[\mathbf{T}^{(k)}\left(\frac{\partial \mathbf{\Theta}_1}{\partial \theta_i} - c\frac{\partial \mathbf{\Theta}_2}{\partial \theta_i}\right)\right] \\ &= 2\tau_{**}^{(k)} \sum_{j=1}^{p-1} (\varphi_{ij}^{(1)} - c\varphi_{ij}^{(2)})\theta_j + 2\sum_{j=1}^{p-1} (\varphi_{ij}^{(1)} - c\varphi_{ij}^{(2)})\tau_{ij}^{(k)} \\ &= \delta_i^k, \end{aligned} \tag{33}$$

where we define $\delta_k^k = 1$, $\delta_i^k = 0$ $(k \neq i)$ and hence we have $\frac{\partial \phi_k}{\partial \theta_i} = \delta_i^k$ in (31).

Define a $p \times p$ $(p = K_1 + 1 + G_2)$ partitioned matrix

$$\mathbf{T}^{(k)} = \begin{bmatrix} \mathbf{T}_{11}^{(k)} & \boldsymbol{\tau}_1^{(k)} & \mathbf{T}_{12}^{(k)} \\ \boldsymbol{\tau}_1^{(k)\prime} & \tau_{**}^{(k)} & \boldsymbol{\tau}_2^{(k)\prime} \\ \mathbf{T}_{12}^{(k)\prime} & \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix}.$$

($\mathbf{T}_{11}^{(k)}, \mathbf{T}_{12}^{(k)}, \mathbf{T}_{22}^{(k)}$ are $K_1 \times K_1$, $K_1 \times G_2$, $G_2 \times G_2$ matrices, respectively, $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors, respectively, and s_{**} is a scalar random variable.)

By differentiating each of the elements of $\mathbf{\Theta}_g$ with respect to $\varphi_{ij}^{(g)}$, we find the relation for $g = 1, 2$ that

$$\begin{aligned} \operatorname{tr}\left(\mathbf{T}^{(k)}\frac{\partial \mathbf{\Theta}_g}{\partial \varphi_{ij}^{(g)}}\right) &= \operatorname{tr}\begin{bmatrix} \mathbf{T}_{11}^{(k)} & \boldsymbol{\tau}_1^{(k)} & \mathbf{T}_{12}^{(k)} \\ \boldsymbol{\tau}_1^{(k)\prime} & \tau_{**}^{(k)} & \boldsymbol{\tau}_2^{(k)\prime} \\ \mathbf{T}_{12}^{(k)\prime} & \boldsymbol{\tau}_2 & \mathbf{T}_{22}^{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{K_1} & \mathbf{O} \\ \boldsymbol{\gamma}_1' & \boldsymbol{\beta}_2' \\ \mathbf{O} & \mathbf{I}_{G_2} \end{bmatrix} \left[\mathbf{e}_i\mathbf{e}_j'\right] \begin{bmatrix} \mathbf{I}_{K_1} & \boldsymbol{\gamma}_1 & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{bmatrix} \\ &= \tau_{**}^{(k)} \left[\boldsymbol{\gamma}_1', \boldsymbol{\beta}_2'\right] \left[\mathbf{e}_i\mathbf{e}_j'\right] \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \\ &\quad + 2\left[\boldsymbol{\gamma}_1', \boldsymbol{\beta}_2'\right] \left[\mathbf{e}_i\mathbf{e}_j'\right] \begin{bmatrix} \boldsymbol{\tau}_1^{(k)} \\ \boldsymbol{\tau}_2^{(k)} \end{bmatrix} + \operatorname{tr}\left[\mathbf{e}_i\mathbf{e}_j'\right] \begin{bmatrix} \mathbf{T}_{11}^{(k)} & \mathbf{T}_{12}^{(k)} \\ \mathbf{T}_{12}^{(k)\prime} & \mathbf{T}_{22}^{(k)} \end{bmatrix}, \end{aligned}$$

where $\mathbf{e}_i' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the ith place and zeros in other elements.

Furthermore, (33) can be represented as

$$2\tau_{**}^{(k)}(\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*)\boldsymbol{\theta}_2 + 2(\mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*)\begin{bmatrix}\boldsymbol{\tau}_1^{(k)}\\ \boldsymbol{\tau}_2^{(k)}\end{bmatrix} = \mathbf{e}_k, \tag{34}$$

where we have set $\boldsymbol{\theta}_2' = (\boldsymbol{\gamma}_1', \boldsymbol{\beta}_2')$. Since we have assumed that $\mathbf{\Phi}^*$ $(= \mathbf{\Phi}_1^* - c\mathbf{\Phi}_2^*)$ is positive definite, we solve (34) as

$$\begin{bmatrix}\boldsymbol{\tau}_1^{(k)}\\ \boldsymbol{\tau}_2^{(k)}\end{bmatrix} = \frac{1}{2}\mathbf{\Phi}^{*-1}\mathbf{e}_k - \tau_{**}^{(k)}\boldsymbol{\theta}_2.$$

Also as the conditions with respect to $\varphi_{ij}^{(k)}$ [by using the similar arguments as (6.54) and (6.55) of Anderson et al. (2010)], we have the representation

$$\tau_{**}^{(k)}\boldsymbol{\theta}_2\boldsymbol{\theta}_2' + \begin{bmatrix}\boldsymbol{\tau}_1^{(k)}\\ \boldsymbol{\tau}_2^{(k)}\end{bmatrix}\boldsymbol{\theta}_2' + \boldsymbol{\theta}_2(\boldsymbol{\tau}_1^{(k)'}, \boldsymbol{\tau}_2^{(k)'}) + \begin{bmatrix}\mathbf{T}_{11}^{(k)} & \mathbf{T}_{12}^{(k)}\\ \mathbf{T}_{12}^{(k)'} & \mathbf{T}_{22}^{(k)}\end{bmatrix} = \mathbf{O}\,.$$

Then we have the representation

$$\begin{aligned}\begin{bmatrix}\mathbf{T}_{11}^{(k)} & \mathbf{T}_{12}^{(k)}\\ \mathbf{T}_{12}^{(k)'} & \mathbf{T}_{22}^{(k)}\end{bmatrix} &= -\tau_{**}^{(k)}\boldsymbol{\theta}_2\boldsymbol{\theta}_2' - \begin{bmatrix}\boldsymbol{\tau}_1^{(k)}\\ \boldsymbol{\tau}_2^{(k)}\end{bmatrix}\boldsymbol{\theta}_2' - \boldsymbol{\theta}_2(\boldsymbol{\tau}_1^{(k)'}, \boldsymbol{\tau}_2^{(k)'})\\ &= \tau_{**}^{(k)}\boldsymbol{\theta}_2\boldsymbol{\theta}_2' - \frac{1}{2}\left[\mathbf{\Phi}^{*-1}\mathbf{e}_k\boldsymbol{\theta}_2' + \boldsymbol{\theta}_2\mathbf{e}_k'\mathbf{\Phi}^{*-1}\right].\end{aligned} \tag{35}$$

Let a $p \times p$ $(p = K_1 + 1 + G_2)$ matrix $\mathbf{S}$ $(= (s_{ij}))$ be partitioned as

$$\mathbf{S} = \mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1 = \begin{bmatrix}\mathbf{S}_{11} & \mathbf{s}_1 & \mathbf{S}_{12}\\ \mathbf{s}_1' & s_{**} & \mathbf{s}_2'\\ \mathbf{S}_{12}' & \mathbf{s}_2 & \mathbf{S}_{22}\end{bmatrix},$$

where $\mathbf{G}_1$ and $\mathbf{H}_1$ are defined as in the proof of Theorem 1 and $\mathbf{S}$ is partitioned into $(K_1+1+G_2)\times(K_1+1+G_2)$ elements. ($\mathbf{S}_{11}, \mathbf{S}_{12}, \mathbf{S}_{22}$ are $K_1 \times K_1,\ K_1 \times G_2,\ G_2 \times G_2$ matrices, respectively, and $\mathbf{s}_1$ and $\mathbf{s}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors, respectively.) We set the scalar random variable $s_{K_1+1,K_1+1} = s_{**}$ for the notational convenience.

Since we have assumed that $\phi_k(\cdot)$ is differentiable and its first derivatives are bounded at the true parameters, we shall consider the asymptotic distribution of the normalized estimator $\sqrt{n}[\hat{\theta}_k - \theta_k]$ $(k = 1, \ldots, p-1)$ with (30) and (31). Then by using the restrictions of (34) and (35), the linearized estimator of θ_k for the class of the modified LIML estimators, which are consistent, can be represented as

$$\sum_{g,h=1}^{p} \tau_{gh}^{(k)} s_{gh} = \tau_{**}^{(k)} s_{**} + 2(\boldsymbol{\tau}_1^{(k)\prime}, \boldsymbol{\tau}_2^{(k)\prime}) \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} + \mathrm{tr} \begin{bmatrix} \mathbf{T}_{11}^{(k)} & \mathbf{T}_{12}^{(k)} \\ \mathbf{T}_{12}^{(k)\prime} & \mathbf{T}_{22}^{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{12}' & \mathbf{S}_{22} \end{bmatrix}$$
$$= \tau_{**}^{(k)} \boldsymbol{\theta}' \mathbf{S} \boldsymbol{\theta} + \mathbf{e}_k' \boldsymbol{\Phi}^{*-1} \begin{bmatrix} \mathbf{S}_{11} & \mathbf{s}_1 & \mathbf{S}_{12} \\ \mathbf{S}_{12}' & \mathbf{s}_2 & \mathbf{S}_{22} \end{bmatrix} \boldsymbol{\theta} \ .$$

[We have used the fact that $c\sqrt{n/(n-q_n)} \sim \sqrt{cc_*}$ and (32) for each of the elements of the second matrix in (30) and (31).] Let

$$\boldsymbol{\tau}_{**} = \begin{bmatrix} \tau_{**}^{(1)} \\ \vdots \\ \tau_{**}^{(p-1)} \end{bmatrix}$$

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

$$\hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{**} \boldsymbol{\theta}' + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \right] \mathbf{S} \boldsymbol{\theta},$$

where we use $\mathbf{J}'$ in the proof of Theorem 1.

The asymptotic normality of the class of the modified LIML estimators can be established by using the similar arguments as the proof of Theorem 1. Since the asymptotic variance–covariance matrix of $\mathbf{S}\boldsymbol{\theta}$ has been obtained by the proof of Theorem 1, we have

$$\mathcal{E}\left[\hat{\mathbf{e}}\, \hat{\mathbf{e}}'\right] = \left[(\boldsymbol{\tau}_{**} + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \boldsymbol{\Omega}^* \boldsymbol{\theta}) \boldsymbol{\theta}' + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \left(\mathbf{I}_p - \frac{\boldsymbol{\Omega}^* \boldsymbol{\theta} \boldsymbol{\theta}'}{\boldsymbol{\theta}' \boldsymbol{\Omega}^* \boldsymbol{\theta}} \right) \right]$$
$$\times \mathcal{E}[\mathbf{S} \boldsymbol{\theta} \boldsymbol{\theta}' \mathbf{S}] \times \left[(\boldsymbol{\tau}_{**} + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \boldsymbol{\Omega}^* \boldsymbol{\theta}) \boldsymbol{\theta}' + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \left(\mathbf{I}_p - \frac{\boldsymbol{\Omega}^* \boldsymbol{\theta} \boldsymbol{\theta}'}{\boldsymbol{\theta}' \boldsymbol{\Omega}^* \boldsymbol{\theta}} \right) \right]'$$
$$= \boldsymbol{\Psi}^* + \mathcal{E}[(\boldsymbol{\theta}' \mathbf{S} \boldsymbol{\theta})^2] \left[\boldsymbol{\tau}_{**} + \boldsymbol{\Phi}^{*-1} \mathbf{J}' \boldsymbol{\Omega}^* \boldsymbol{\theta} \right] \left[\boldsymbol{\tau}_{**}' + \boldsymbol{\theta}' \boldsymbol{\Omega}^* \mathbf{J} \boldsymbol{\Phi}^{*-1} \right] + o(1) \ ,$$

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

$$\boldsymbol{\Phi}^* \boldsymbol{\tau}_{**} + \mathbf{J}' \boldsymbol{\Omega}^* \boldsymbol{\theta} = \mathbf{0} \ .$$

□

A Brief Derivation of Theorem 3 The full proof of Theorem 3 can be given by using similar arguments as the one in Theorem 1. Since many parts are quite similar and lengthy, we give the most important steps.

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

$$\mathbf{P}_M - c_* \mathbf{Q}_M = [\mathbf{P}_Z - \mathbf{D}_n + c_n \mathbf{I}_n] - c_*[\mathbf{I}_n - (\mathbf{P}_Z - \mathbf{D}_n + c_n \mathbf{I}_n)]$$
$$= (1 + c_*)(\mathbf{P}_Z - \mathbf{D}_n) + [c_n - c_*(1 - c_n)] \mathbf{I}_n .$$

The assumptions in Theorem 1 implies

$$\frac{1}{n}\mathbf{\Pi}_{*n}^{(z)'}[\mathbf{P}_M - c^*\mathbf{Q}_M]\mathbf{\Pi}_{*n}^{(z)} - (1+c_*)\frac{1}{n}\mathbf{\Pi}_{*n}^{(z)'}\left[\mathbf{P}_Z - \mathbf{D}_n\right]\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{O}. \tag{36}$$

Under the assumption of (23) that

$$\frac{1}{n}\mathbf{\Pi}_{*n}^{(z)'}(\mathbf{P}_Z - \mathbf{D}_n)\mathbf{\Pi}_{*n}^{(z)} \xrightarrow{p} \mathbf{\Phi}_D^* = (1+c_*)^{-1}\mathbf{\Phi}^* \tag{37}$$

is a positive definite matrix as $n \to \infty$.

Let λ_{nH} be the smallest root of (14) and (15) in the HLIM estimation by using $\mathbf{P}_H$ and $\mathbf{Q}_H$ instead of $\mathbf{P}_M$ and $\mathbf{Q}_M$. Then we have plim $\lambda_{nH} = 0$ because $p_{ii}^* = 0$ $(i = 1, \ldots, n)$ and $(1/n)\mathbf{V}'\mathbf{P}_H\mathbf{V} \xrightarrow{p} \mathbf{O}$. By using the same arguments for $\mathbf{\Psi}_i^*$ $(i = 1, 2)$ in (18), we have the corresponding representation of (29) for the HLIM estimator as

$$\begin{aligned}
&\mathbf{\Phi}_D^*\sqrt{n}\left[\begin{pmatrix}\hat{\boldsymbol{\gamma}}_{1.\mathrm{HLI}}\\ \hat{\boldsymbol{\beta}}_{2.\mathrm{HLI}}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right]\\
&\quad = \frac{1}{\sqrt{n}}\mathbf{\Pi}_{*n}^{(z)'}\left(\mathbf{P}_Z - \mathbf{D}_n\right)\mathbf{u} + \frac{1}{\sqrt{n}}[\mathbf{W}'(\mathbf{P}_Z - \mathbf{D}_n)\mathbf{u}] + o_p(1).
\end{aligned} \tag{38}$$

Then, after some calculations, we can find that the corresponding terms of $\mathbf{\Psi}_i^*$ $(i = 1, 2)$ (i.e. (18) for the AOM-LIML estimator) become

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

$$\begin{aligned}
\mathbf{\Psi}_2^{**} &= (1+c_*)^2 \mathrm{plim}\frac{1}{n}\sum_{i,j=1}^{n}[\sigma_i^2\mathcal{E}(\mathbf{w}_{*j}\mathbf{w}_{*j}'|\mathbf{z}_j^{(n)}) + \mathcal{E}(\mathbf{w}_{*i}u_i|\mathbf{z}_i^{(n)})\mathcal{E}(\mathbf{w}_{*j}'u_j|\mathbf{z}_j^{(n)})]\\
&\quad \times[p_{ij}^{(n)}(1-\delta_i^j)]^2,
\end{aligned}$$

where $\delta_i^j = 0$ $(i = j)$, $\delta_i^j = 0$ $(i \neq j)$. Hence the factors $(1+c_*)^2$ in $\mathbf{\Phi}^{**}$ and $\mathbf{\Psi}^*$ are cancelled out. By using (36) and (37), the covariance matrix of the asymptotic distribution has the same form in Theorem 1.

The asymptotic normality can be shown by using similar arguments to *Step II* for the proof of Theorem 1, which have been omitted because these are straightforward but quite lengthy. □

Appendix

In Figs. 1, 2 and 3 the distribution functions of the LIML, HLIM and AOM-LIML estimators are shown with the large-K_2 normalization. (In brief, we use MLIML for the AOM-LIML estimator in figures.) 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 α stands for the normalized coefficient of an endogenous variable and δ^2 is the non-centrality parameter. The details of numerical computation method of this paper are given in Anderson et al. (2005, 2008) and Kunitomo and Matsushita (2009).

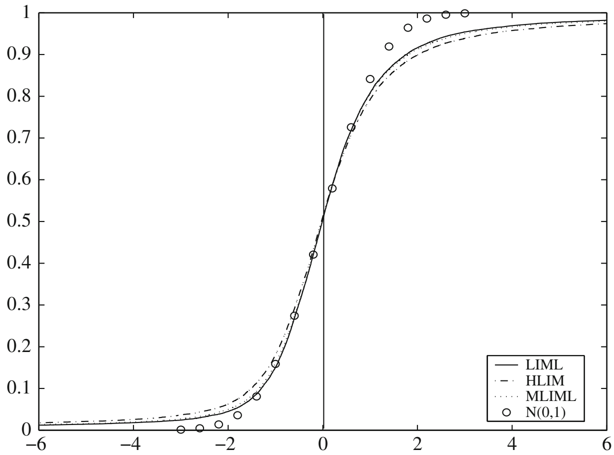

Fig. 1 CDF of standardized estimators: $n - K = 20$, $K_2 = 30$, $\alpha = 0.5$, $\delta^2 = 30$, $u_i = N(0, 1)$

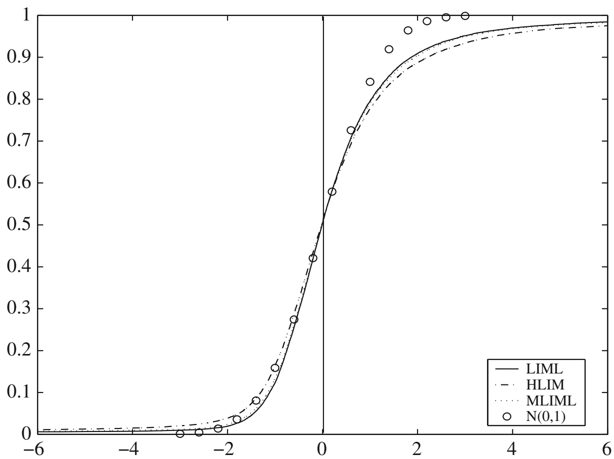

Fig. 2 CDF of standardized estimators: $n - K = 20$, $K_2 = 30$, $\alpha = 1$, $\delta^2 = 30$, $u_i = N(0, 1)$

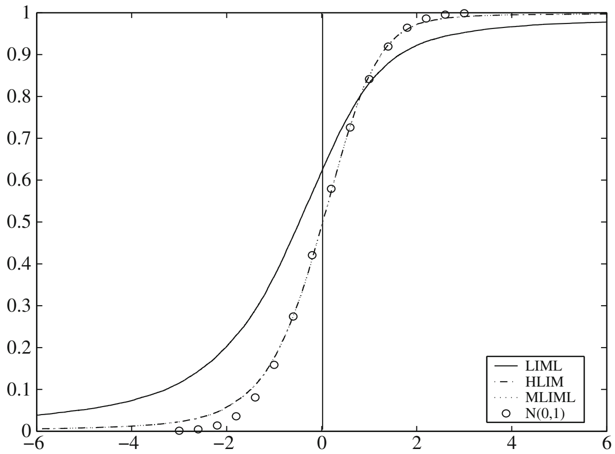

Fig. 3 CDF of standardized estimators: Heteroscedastic disturbances in Hausman et al. (2007), $n = 100$, $K = 10$, $\delta^2 = 30$

Acknowledgments This is a revised version of the 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 the useful discussions and for providing some figures. I also thank Theodore W. Anderson, Cheng Hsiao, the editor and referees of AISM for their helpful comments.

References

Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics, 12*, 1–45.

Anderson, T. W., Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics, 20*, 46–63.

Anderson, T. W., Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equation. *Annals of Mathematical Statistics, 21*, 570–582.

Anderson, T. W., Kunitomo, N., Sawa, T. (1982). Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica, 50*, 1009–1027.

Anderson, T. W., Kunitomo, N., Matsushita, Y. (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. http://www.e.u-tokyo.ac.jp/cirje/research/dp/2005.

Anderson, T. W., Kunitomo, N., Matsushita, Y. (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. http://www.e.u-tokyo.ac.jp/cirje/research/dp/2008 (forthcoming in Journal of Econometrics).

Anderson, T. W., Kunitomo, N., Matsushita, Y. (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments. *Journal of Econometrics, 157*, 191–204.

Angrist, J. D., Imbens, G. W., Krueger, A. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics, 14*, 57–67.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variables estimators. *Econometrica, 63*, 657–681.

Chao, J., Swanson, N. (2004). Asymptotic distributions of JIVE in a heteroscedastic IV regression with many instruments (working paper).

Chao, J., Swanson, N. (2005). Consistent estimation with a large number of weak instruments. *Econometrica 73*, 1673–1692.

Chao, J., Swanson, N., Hausman, J., Newey W., Woutersen T. (2009). Asymptotic distribution of JIVE in a heteroscedastic IV regression with many instruments (unpublished manuscript).

Fuller, W. (1977). Some properties of a modification of the limited information estimator. *Econometrica, 45*, 939–953.

Hall, P., Heyde, C. (1980). *Martingale limit theory and its application*. London: Academic Press.

Hansen, C., Hausman, J., Newey, W. K. (2008). Estimation with many instrumental variables. *Journal of Business and Economic Statistics, 26*, 398–422

Hausman, J., Newey, W., Woutersen, T., Chao, J., Swanson, N. (2007). Instrumental variables estimation with heteroscedasticity and many instruments (unpublished manuscript).

Hayashi, F. (2000). *Econometrics*. Princeton, NJ: Princeton University Press.

Kunitomo, N. (1980). Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association, 75*, 693–700.

Kunitomo, N. (1981). Asymptotic optimality of the limited information maximum likelihood estimator in large econometric models. *The Economic Studies Quarterly, XXXII*(3), 247–266.

Kunitomo, N. (1982). Asymptotic efficiency and higher order efficiency of the limited information maximum likelihood estimator in large econometric models. (Technical Report No. 365). Palo-Alto, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

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, 39*, 575–591.

Kunitomo, N., Matsushita, Y. (2009). Asymptotic expansions and higher order properties of semi-parametric estimators in a system of simultaneous equations. *Journal of Multivariate Analysis, 100*, 1727–1751.

Morimune, K. (1983). Approximate distributions of k-class estimators when the degree of overidentification is large compared with sample size. *Econometrica, 51*(3), 821–841.

Owen, A. B. (2001). *Empirical Likelihood*. New York: Chapman and Hall.

Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics, 22*, 300–325