Journal of Econometrics 157 (2010) 191–204

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

On the asymptotic optimality of the LIML estimator with possibly many instruments[☆]

T.W. Anderson[a,b], Naoto Kunitomo[c,*], Yukitoshi Matsushita[c]

[a] *Department of Statistics, Stanford University, United States*

[b] *Department of Economics, Stanford University, United States*

[c] *Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1 Tokyo, Japan*

ARTICLE INFO

Article history:
Received 21 December 2007
Received in revised form
1 December 2009
Accepted 4 December 2009
Available online 24 December 2009

JEL classification:
C13
C30

Keywords:
Structural equation
Simultaneous equations system
Many instruments
Many weak instruments
Limited information maximum likelihood
Asymptotic optimality

ABSTRACT

We consider the estimation of the coefficients of a linear structural equation in a simultaneous equation system when there are many instrumental variables. We derive some asymptotic properties of the limited information maximum likelihood (LIML) estimator when the number of instruments is large; some of these results are new as well as *old*, and we relate them to results in some recent studies. We have found that the variance of the limiting distribution of the LIML estimator and its modifications often attain the asymptotic lower bound when the number of instruments is large and the disturbance terms are not necessarily normally distributed, that is, for the micro-econometric models of some cases recently called *many instruments* and *many weak instruments*.

1. Introduction

Over the past three decades there has been increasing interest and research on the estimation of a single structural equation in a system of simultaneous equations when the number of instruments (the number of exogenous variables excluded from the structural equation), say K_2, is large relative to the sample size, say n. The relevance of such models is due to the collection of large data sets and the development of computational equipment capable of analysis of such data sets. One empirical example of this kind often cited in econometric literature is Angrist and Krueger (1991); there has been some discussion by Bound et al. (1995) since then. Asymptotic distributions of estimators and test criteria are developed on the basis that 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.

Bekker (1994) has written "To my knowledge a first mention of such a parameter sequence was made, with respect to the linear functional relationship model, in Anderson (1976, p. 34). This work was extended to simultaneous equations by Kunitomo (1980, 1982) and Morimune (1983), who gave asymptotic expansions for the case of a single explanatory endogenous variable". Following Bekker there have been many studies of the behavior of estimators of the coefficients of a single equation when K_2 and n are large.

The main purpose of the present paper is to show that one estimator, the Limited Information Maximum Likelihood (LIML) estimator, has some optimum properties when K_2 and n are large. As background we state and derive some asymptotic distributions of the LIML and Two-Stage Least Squares (TSLS) estimators as $K_2 \to \infty$ and $n \to \infty$. Some of these results are improvements on Kunitomo (1981, 1982), Morimune (1983) and Bekker (1994), Chao and Swanson (2005), van Hasselt (2006), Hansen et al. (2008), and they are presented in a unified notation.

In addition to the LIML and TSLS estimators there are other instrumental variables (IV) methods. See Anderson et al. (1982) on the earlier studies of the finite sample properties, for instance.

☆ AKM09-11-23-2. This is a revised first part of Discussion Paper CIRJE-F-321 under the title "A New Light from Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models" (Graduate School of Economics, University of Tokyo, February 2005) which was presented at the Econometric Society World Congress 2005 at London (August 2005). We thank Yoichi Arai, the co-editor and the referees of this journal for some comments to the earlier versions.

* Corresponding author. Tel.: +81 3 5841 5614.
E-mail address: kunitomo@e.u-tokyo.ac.jp (N. Kunitomo).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.12.001

Several semiparametric estimation methods have been developed including the generalized method of moments (GMM) estimation and the empirical likelihood (EL) method. (See Hayashi, 2000 for instance.) However, it has been recently recognized that the classical methods have some advantages in microeconometric situations with many instruments.

In this paper we shall give the results on the asymptotic properties of the LIML estimator when the number of instruments is large and we develop *the large-K_2 asymptotic theory* or *the many instruments asymptotics* including the so-called case of *many weak instruments*. The TSLS and the GMM estimators are badly biased and they lose even consistency in some of these situations. Our results on the asymptotic properties and optimality of the LIML estimator and its variants give new interpretations of the numerical information of the finite sample properties and some guidance on the use of alternative estimation methods in simultaneous equations and micro-econometric models with *many weak instruments*. There is a growing amount of literature on the problem of many instruments in econometric models. We shall try to relate our results to some recent studies, including Donald and Newey (2001), Hahn (2002), Stock and Yogo (2005), Chao and Swanson (2005, 2006), van Hasselt (2006), van der Ploeg and Bekker (unpublished), Bekker and van der Ploeg (2005), Chioda and Jansson (unpublished), Hansen et al. (2008), and Anderson et al. (forthcoming).

In Section 2 we state the formulation of a linear structural model and the alternative estimation methods of unknown parameters with possibly many instruments. In Section 3 we develop the large-K_2 asymptotics (or *many instruments* asymptotics) and give some results on the asymptotic normality of the LIML estimator when n and K_2 are large. Then we shall present some results on the asymptotic optimality of the LIML estimator in the sense that it attains the lower bound of the asymptotic variance in a class of consistent estimators with *many instruments* under reasonable assumptions. Also we discuss a more general formulation of the models and the related problems. In Section 4 we show that the asymptotic results in Section 3 agree with the finite sample properties of estimators. Then brief concluding remarks will be given in Section 5. The proof of our theorems will be given in Section 6.

2. Alternative estimation methods in structural equation models with possibly many instruments

We first consider the estimation problem of a structural equation in the classical linear simultaneous equations framework.[1] Let a single linear structural equation in an econometric model be

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

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and a vector of G_2 endogenous variables, $\boldsymbol{z}_{1i}$ is a vector of K_1 (included) exogenous variables in (2.1), $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters, and $u_1, \ldots, u_n$ are independent disturbance terms with $\mathcal{E}(u_i) = 0$ and $\mathcal{E}(u_i^2) = \sigma^2$ $(i = 1, \ldots, n)$. We assume that (2.1) is one equation in a system of $1 + G_2$ equations in $1 + G_2$ endogenous variables $\boldsymbol{y}_i' = (y_{1i}, \boldsymbol{y}_{2i}')'$. The reduced form of the model is

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

where $\mathbf{Y} = (\mathbf{y}_i')$ is the $n \times (1 + G_2)$ matrix of endogenous variables, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_{2n}) = (\mathbf{z}_i^{(n)'})$ is the $n \times K_n$ matrix of $K_1 + K_{2n}$ instrumental vectors $\mathbf{z}_i^{(n)} = (\mathbf{z}_{1i}', \mathbf{z}_{2i}^{(n)'})'$, $\mathbf{V} = (\mathbf{v}_i')$ is the $n \times (1 + G_2)$ matrix of disturbances,

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

is the $(K_1 + K_{2n}) \times (1 + G_2)$ matrix of coefficients, and

$$\mathcal{E}(\mathbf{v}_i \mathbf{v}_i') = \boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_2' \\ \boldsymbol{\omega}_2 & \boldsymbol{\Omega}_{22} \end{bmatrix}$$

is a positive definite matrix. The vector of K_n $(= K_1 + K_{2n}, n > 2)$ instrumental variables $\mathbf{z}_i^{(n)}$ satisfies the orthogonality condition $\mathcal{E}[u_i \boldsymbol{z}_i^{(n)}] = \mathbf{0}$ $(i = 1, \ldots, n)$. The relation between (2.1) and (2.2) gives

$$\begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21}^{(n)} & \boldsymbol{\Pi}_{22}^{(n)} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{pmatrix}, \tag{2.3}$$

$u_i = (1, -\boldsymbol{\beta}_2')\mathbf{v}_i = \boldsymbol{\beta}'\mathbf{v}_i$ and $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$ with $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2')$.

Let $\boldsymbol{\Pi}_{2n} = (\boldsymbol{\pi}_{21}^{(n)}, \boldsymbol{\Pi}_{22}^{(n)})$ be a $K_{2n} \times (1 + G_2)$ matrix of coefficients. Define the $(1 + G_2) \times (1 + G_2)$ matrices by

$$\mathbf{G} = \mathbf{Y}'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y} = \mathbf{P}_2'\mathbf{A}_{22.1}\mathbf{P}_2, \tag{2.4}$$

and

$$\mathbf{H} = \mathbf{Y}' \left(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \right) \mathbf{Y}, \tag{2.5}$$

where $\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}'\mathbf{Z}_{2.1}$, $\mathbf{Z}_{2.1} = \mathbf{Z}_{2n} - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $\mathbf{P}_2 = \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y}$,

$$\mathbf{Z}_1 = \begin{pmatrix} \mathbf{z}_{11}' \\ \vdots \\ \mathbf{z}_{1n}' \end{pmatrix}, \qquad \mathbf{Z}_{2n} = \begin{pmatrix} \mathbf{z}_{21}^{(n)'} \\ \vdots \\ \mathbf{z}_{2n}^{(n)'} \end{pmatrix}, \tag{2.6}$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_{2n}' \end{pmatrix} (\mathbf{Z}_1, \mathbf{Z}_{2n}) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{2.7}$$

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

$$\left(\frac{1}{n}\mathbf{G} - \frac{1}{q_n}\lambda_n \mathbf{H} \right) \hat{\boldsymbol{\beta}}_{LI} = \mathbf{0}, \tag{2.8}$$

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

$$\left| \frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H} \right| = 0. \tag{2.9}$$

The solution to (2.8) minimizes the variance ratio given in Box I. The TSLS estimator $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}_{2.TS}')')$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$ is given by

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

where $\mathbf{Y}_2 = (\mathbf{y}_i')$ is an $n \times G_2$ matrix. The TSLS estimator minimizes the numerator of the variance ratio Box I. The LIML and the TSLS estimators of $\boldsymbol{\gamma}_1$ are

$$\hat{\boldsymbol{\gamma}}_1 = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Y}\hat{\boldsymbol{\beta}}, \tag{2.11}$$

where $\hat{\boldsymbol{\beta}}$ is $\hat{\boldsymbol{\beta}}_{LI}$ or $\hat{\boldsymbol{\beta}}_{TS}$, respectively. In this paper we shall discuss the asymptotic properties of $\hat{\boldsymbol{\beta}}_2$ because of its simplicity, although it is straightforward to extend to treat $\sqrt{n}[\hat{\boldsymbol{\beta}}_2' - \boldsymbol{\beta}_2', \hat{\boldsymbol{\gamma}}_1' - \boldsymbol{\gamma}_1']$ with some additional notations. The LIML and TSLS estimators and their properties in the general case were originally developed by Anderson and Rubin (1949, 1950). See also Anderson (2005).

[1] We intentionally consider the standard classic situation and state our results mainly because they are clear. Nonetheless a generalization of the formulation and the corresponding results will be discussed in Section 3.3.