J. Japan Statist. Soc.
Vol. 38 No. 2 2008 207–223

APPROXIMATE DISTRIBUTIONS OF THE LIKELIHOOD RATIO STATISTIC IN A STRUCTURAL EQUATION WITH MANY INSTRUMENTS

Yukitoshi Matsushita*

This paper studies properties of the likelihood ratio (LR) tests associated with the limited information maximum likelihood (LIML) estimators in a structural form estimation when the number of instrumental variables is large. Two types of asymptotic theories are developed to approximate the distribution of the likelihood ratio (LR) statistic under the null hypothesis $H_0 : \beta = \beta_0$: a (large sample) asymptotic expansion and a large-K_n asymptotic theory. Size comparisons of two modified LR tests based on these two asymptotics are made with Moreira's conditional likelihood ratio (CLR) test and the large K t-test.

Key words and phrases: Asymptotic expansion, large-K_n asymptotics, many instruments.

1. Introduction

Statistical inference procedures in structural equation models can be crucially affected by the quality and the number of the instrumental variables. It has been known that when instruments are only weakly correlated with the endogenous variables, classical normal and chi-square asymptotic approximations to the finite-sample distributions of estimators and statistics can be poor. See Nelson and Startz (1990a, b), Bound *et al.* (1995), and Staiger and Stock (1997), for instance. If the number of the instrumental variables is large, efficiency can be improved, but it makes the finite-sample properties of usual inference procedures poor. In addition, in recent microeconometric applications, some econometricians have used many instrumental variables in estimating an important structural equation. One empirical example of this kind often cited in econometric literature is Angrist and Krueger (1991), where they used 178 instruments in one of their specifications. Bound *et al.* (1995) shows that the properties of the two stage least squares (TSLS) estimator can be poor in the case of many weak instruments, even when the sample size is huge.

In order to overcome these problems, several new statistical procedures have been proposed recently. For inference on all the coefficients of endogenous parameters, the Anderson-Rubin (AR) test is a fundamental building block for developing reliable inference procedures with weak instruments; see Anderson and Rubin (1949). Kleibergen (2002) and Moreira (2001) proposed a score-type statistic, while Moreira (2003) proposed a conditional likelihood ratio (CLR) test,

Received November 4, 2006. Revised May 3, 2007. Accepted May 7, 2007.
*CIRJE, Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Email: yukitoshi_mjp@yahoo.co.jp

both of which are also shown to be robust to weak instruments. Among these testing procedures, the CLR test has been found to dominate the other tests in terms of power. Andrews *et al.* (2006) showed that the CLR test is quite close to being uniformly the most powerful invariant among a class of two-sided tests.

On the other hand, there has been another approach to provide better approximation using "large-K_n asymptotics", where the number of instruments (K) is allowed to increase with the number of observations (n). Kunitomo (1980) and Morimune (1983) were early developers of the large-K_n asymptotics, and they derived asymptotic expansions of the distributions of the k-class estimators including the TSLS and the limited information maximum likelihood (LIML) estimators in the case of two endogenous variables. Multivariate first order approximations to the distributions were derived by Bekker (1994) and Anderson *et al.* (2006). Bekker (1994) found that the large-K_n asymptotics provides better approximations than the one where K is fixed. Hansen *et al.* (2006) consider the same model and show that Bekker's (1994) standard error corrects the size problem. Matsushita (2006) has derived an asymptotic expansion of the distributions of the LIML estimator and (large K) t-ratio under H_0 under the large-K_n asymptotics.

The main purpose of this paper is to explore finite sample properties of the likelihood ratio (LR) test, on all the coefficients of endogenous variables in a structural equation model, when the number of the instrumetal variables is large. We develop two types of alternative asymptotic theories to approximate the null distribution of the LR statistic: a (large sample) asymptotic expansion (in the case of normal disturbances), and a large-K_n asymptotics (in the case of non-normal disturbances). We propose two types of modified LR tests from these asymptotics, and compare their finite sample properties with that of Moreira's conditional likelihood ratio (CLR) test using Monte Carlo experiments.

The model and several test statistics are explained in Section 2. An asymptotic expansion of the null distribution of the LR statistic is given in Section 3, while an approximate null distribution based on the large-K_n asymptotics is given in Section 4. Some Monte Carlo experiments are provided in Section 5, and conclusions are provided in Section 6.

2. The model and test statistics

Let a single structural equation be

$$\boldsymbol{y}_1 = \boldsymbol{Y}_2\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{\gamma} + \boldsymbol{u}, \tag{2.1}$$

where $\boldsymbol{y}_1$ and $\boldsymbol{Y}_2$ are $n\times 1$ and $n\times G_1$ matrices, respectively, of observations of the endogenous variables, $\boldsymbol{Z}_1$ is an $n\times K_1$ matrix of observations of the K_1 exogenous variables, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are column vectors with G_1 and K_1 unknown parameters, and $\boldsymbol{u}$ is a column vector of n disturbances. We assume that (2.1) is the first equation in a simultaneous system of $G_1 + 1$ linear stochastic equations relating $G_1 + 1$ endogenous variables and $K(= K_1 + K_2)$ exogenous variables. The reduced form of $\boldsymbol{y} = (\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2)$ is defined as

$$\boldsymbol{Y} = \boldsymbol{Z\Pi} + \boldsymbol{V} = (\boldsymbol{Z}_1 \;\; \boldsymbol{Z}_2)\begin{pmatrix}\boldsymbol{\pi}_1\\ \boldsymbol{\Pi}_2\end{pmatrix} + (\boldsymbol{v}_1 \;\; \boldsymbol{V}_2), \tag{2.2}$$

where $\boldsymbol{Z}$ is an $n \times K$ matrix of instrumental variables, $\boldsymbol{\pi}_1 = (\boldsymbol{\pi}_{11} \;\; \boldsymbol{\Pi}_{12})$ and $\boldsymbol{\Pi}_2 = (\boldsymbol{\pi}_{21} \;\; \boldsymbol{\Pi}_{22})$ are $K_1 \times (1+G_1)$ and $K_2 \times (1+G_1)$ matrices, respectively, of the reduced form coefficients, and $(\boldsymbol{v}_1 \;\; \boldsymbol{V}_2)$ is an $n \times (1+G_1)$ matrix of disturbances. The rows of $\boldsymbol{V}$ are independently distributed, each row having mean 0 and (nonsingular) covariance matrix

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

In order to relate (2.1) and (2.2), we postmultiply (2.2) by $(1, \;\; -\boldsymbol{\beta}')'$, then $\boldsymbol{u} = \boldsymbol{v}_1 - \boldsymbol{V}_2\boldsymbol{\beta}$, $\gamma = \boldsymbol{\pi}_{11} - \boldsymbol{\Pi}_{12}\boldsymbol{\beta}$, and

$$\boldsymbol{\pi}_{21} = \boldsymbol{\Pi}_{22}\boldsymbol{\beta}. \tag{2.4}$$

The matrix $(\boldsymbol{\pi}_{21} \;\; \boldsymbol{\Pi}_{22})$ is of rank G_1 and so is $\boldsymbol{\Pi}_{22}$. The components of $\boldsymbol{u}$ are independently distributed with mean 0 and variance σ^2, which is defined to be $\omega_{11} - 2\boldsymbol{\beta}'\omega_{21} + \boldsymbol{\beta}'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}$.

The LIML estimator of $\boldsymbol{\beta}$, as originally developed by Anderson and Rubin (1949), is a maximum likelihood estimator when the disturbances are normally distributed. We define, for any full column matrix $\boldsymbol{F}$, $\boldsymbol{P}_F = \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'$ and $\bar{\boldsymbol{P}}_F = \boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'$. The LIML estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_{LI}$ satisfying

$$\left\{\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\end{pmatrix}(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})(\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2) - \hat{\lambda}\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\end{pmatrix}\bar{\boldsymbol{P}}_Z(\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2)\right\}\begin{pmatrix}1\\ -\hat{\boldsymbol{\beta}}_{LI}\end{pmatrix} = \boldsymbol{0}, \tag{2.5}$$

where $\hat{\lambda}$ is the smallest root of

$$\left|\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\end{pmatrix}(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})(\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2) - \lambda\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\end{pmatrix}\bar{\boldsymbol{P}}_Z(\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2)\right| = 0. \tag{2.6}$$

We note that the LIML estimator is the minimizer of the variance ratio

$$\lambda = \frac{\boldsymbol{b}'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Yb}}{\boldsymbol{b}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Yb}}, \tag{2.7}$$

where $\boldsymbol{b}' = (1, -\boldsymbol{\beta}')$.

The TSLS estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_{TS}$ satisfying

$$\boldsymbol{Y}_2'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})(\boldsymbol{y}_1 \;\; \boldsymbol{Y}_2)\begin{pmatrix}1\\ -\hat{\boldsymbol{\beta}}_{TS}\end{pmatrix} = \boldsymbol{0}. \tag{2.8}$$

It minimizes the numerator of the variance ratio (2.7). The LIML and the TSLS estimators of $\boldsymbol{\gamma}$ are

$$\hat{\gamma} = (\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1'\boldsymbol{Y}\hat{\boldsymbol{\beta}}, \tag{2.9}$$

where $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_{LI}$ or $\hat{\boldsymbol{\beta}}_{TS}$, respectively. See Anderson (2005) for the details of the LIML and TSLS estimators.

The likelihood ratio (LR) statistic for the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ can be defined as

$$l = (n - K)[\lambda_0 - \hat{\lambda}], \tag{2.10}$$

where

$$\lambda_0 = \frac{\boldsymbol{b}_0' \boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1}) \boldsymbol{Y} \boldsymbol{b}_0}{\boldsymbol{b}_0' \boldsymbol{Y}' \bar{\boldsymbol{P}}_Z \boldsymbol{Y} \boldsymbol{b}_0}, \tag{2.11}$$

$$\hat{\lambda} = \min_{\boldsymbol{b}} \frac{\boldsymbol{b}' \boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1}) \boldsymbol{Y} \boldsymbol{b}}{\boldsymbol{b}' \boldsymbol{Y}' \bar{\boldsymbol{P}}_Z \boldsymbol{Y} \boldsymbol{b}}, \tag{2.12}$$

and $\boldsymbol{b}_0' = (1, -\boldsymbol{\beta}_0')$. Morimune and Tsukuda (1984) discussed several test statistics, including the LR statistic, to test significance of a subset of coefficients.

Recently it has been discovered that usual first order asymptotic approximations can be quite poor in several cases. One is the case of *so-called* weak instruments, in which the instruments are weakly correlated to the included endogenous variables. In order to overcome this problem, several new statistical procedures robust to weak instruments have been proposed: the Anderson-Rubin (AR) test, a score-type test by Kleibergen (2002) and Moreira (2001), and a conditional likelihood ratio (CLR) test by Moreira (2003).

- **Anderson-Rubin (AR) Test**
  Anderson and Rubin (1949) proposed testing the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ to use the statistic

$$AR = \frac{(1, -\boldsymbol{\beta}_0{}') \boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1}) \boldsymbol{Y} (1, -\boldsymbol{\beta}_0{}')'}{(1, -\boldsymbol{\beta}_0{}') \boldsymbol{Y}' \bar{\boldsymbol{P}}_Z \boldsymbol{Y} (1, -\boldsymbol{\beta}_0{}')'/(n - K)}. \tag{2.13}$$

  In case of normal disturbances, under the null hypothesis, since the quadratic forms in the numerator and denominator of (2.13) are independent χ^2 random variables, the AR statistic has an exact $F_{K_2, T-K}$ null distribution. Under either the standard large sample theory or the more general conditions of weak-instrument asymptotics, the AR statistic is asymptotically distributed as a $\chi^2(K_2)$ distribution under the null hypothesis. See Staiger and Stock (1997), for instance. Thus the AR test is one of the testing procedures which are robust to weak instruments.

- **Score-type Test**
  Define the statistics

$$\boldsymbol{S} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1}) \boldsymbol{Y} \boldsymbol{b}_0 (\boldsymbol{b}_0{}' \boldsymbol{\Omega} \boldsymbol{b}_0)^{-1/2} \tag{2.14}$$

  and

$$\boldsymbol{T} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1}) \boldsymbol{Y} \boldsymbol{\Omega}^{-1} \begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix} \left[(\boldsymbol{\beta}_0, \boldsymbol{I}_{G_1}) \boldsymbol{\Omega}^{-1} \begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix} \right]^{-1/2}, \tag{2.15}$$

and $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{T}}$ denote $\boldsymbol{S}$ and $\boldsymbol{T}$ evaluated with $\hat{\boldsymbol{\Omega}} = \boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}/(n-K)$ replacing $\boldsymbol{\Omega}$, where $\boldsymbol{b}_0 = (1, -\boldsymbol{\beta}_0{}')'$. Kleibergen (2002) proposed the statistic

$$\mathcal{K} = \hat{\boldsymbol{S}}'\hat{\boldsymbol{T}}(\hat{\boldsymbol{T}}'\hat{\boldsymbol{T}})^{-1}\hat{\boldsymbol{T}}'\hat{\boldsymbol{S}}. \tag{2.16}$$

Kleibergen showed that under either the standard large sample asymptotics or weak-instrument asymptotics, the limiting distribution of the $\mathcal{K}$ statistic under the null hypothesis is $\chi^2(G_1)$, i.e. robust to the weak instruments.

- **Conditional Likelihood Ratio (CLR) Test**
  The likelihood ratio (LR) statistic for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, when $\boldsymbol{\Omega}$ is known, is given by

$$LR = \frac{\boldsymbol{b}_0'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0}{\boldsymbol{b}_0'\boldsymbol{\Omega}\boldsymbol{b}_0} - \min_{\boldsymbol{b}} \frac{\boldsymbol{b}'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{\Omega}\boldsymbol{b}}. \tag{2.17}$$

  Moreira (2003) showed that the LR statistic is a function of $\boldsymbol{S}$ and $\boldsymbol{T}$ defined in (2.14) and (2.15), and that, in the fixed-instruments and normal-distubances model with known $\boldsymbol{\Omega}$, if its critical value is computed from the conditional distribution given $\boldsymbol{T}$, this conditional likelihood ratio (CLR) test is similar (i.e. fully robust to weak instruments). Moreira (2003) and Andrews *et al.* (2006) suggested computing the null distribution by Monte Carlo simulation or numerical integration. In parctice, $\boldsymbol{\Omega}$ is unknown. However, $\boldsymbol{\Omega}$ can be consistently estimated by $\hat{\boldsymbol{\Omega}} = \boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}/(n-K)$ under the weak-instrument asymptotics, and the conditional likelihood ratio (CLR) test based on the plug-in value of $\boldsymbol{\Omega}$ can be shown to be asymptotically robust to weak instruments under general conditions (stochastic instruments and nonormal disturbances).

3. Asymptotic expansion of the null distribution of the LR statistic

In this section and the next, we will develop two types of alternative asymptotic theories to approximate the null distribution of the LR statistic: a (large sample) asymptotic expansion (Section 3) and a large-K_n asymptotics (Section 4) in order to explore finite sample properties of the likelihood ratio (LR) test when the number of the instrumetal variables is large.

First, we consider a modification of the likelihood ratio test based on an asymptotic expansion of the distribution of the LR statistic under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. The following notations are used throughout this chapter:

$$\boldsymbol{q}_2' = \frac{1}{\sigma^2}(\boldsymbol{\omega}_{12} - \beta'\boldsymbol{\Omega}_{22}, \boldsymbol{0}) : 1 \times p, \tag{3.1}$$

$$\boldsymbol{C}_1 = \boldsymbol{q}_2\boldsymbol{q}_2' : p \times p, \tag{3.2}$$

$$\boldsymbol{C}_2 = \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{C}_1 : p \times p, \tag{3.3}$$

$$\boldsymbol{X} = \boldsymbol{Z}\begin{pmatrix} \boldsymbol{\Pi}_{12} & \boldsymbol{I}_{K_1} \\ \boldsymbol{\Pi}_{22} & \boldsymbol{0} \end{pmatrix} : n \times p, \tag{3.4}$$

and

$$\tilde{\boldsymbol{Q}} = \boldsymbol{X}'\boldsymbol{X} : p \times p. \tag{3.5}$$

We give a large sample asymptotic expansion of the distribution of the LR statistic (2.10) under H_0 in the case of normal disturbances, which is similar to Theorem 1 of Morimune and Tsukuda (1984). The derivation is provided in Appendix A.

THEOREM 1. *Assume there exists a positive definite matrix* $\boldsymbol{Q} = p\lim_{n\to\infty} n^{-1}\tilde{\boldsymbol{Q}}$ *such that* $\boldsymbol{Q} = n^{-1}\tilde{\boldsymbol{Q}} + O_p(n^{-1})$. *When the disturbances are normally distributed, the following asymptotic expansion corresponds to the sample size going to infinity*:

$$\begin{aligned} P(l \leq \xi) = {} & H_{G_1}(\xi) \\ & - \frac{\xi}{n}\left\{\frac{1}{G_1}\sigma^2 \operatorname{tr}(\boldsymbol{Q}^{-1}\boldsymbol{C}_2)L - \frac{1}{2}[G_1 - 2 - \xi]\right\} h_{G_1}(\xi) \\ & + O(n^{-3/2}), \end{aligned} \tag{3.6}$$

where H_{G_1} *and* h_{G_1} *are the* χ^2 *distribution function and* χ^2 *density function with* G_1 *degrees of freedom*, *respectively*, *and* $L = K_2 - G_1$.

The Cornish-Fisher type expansion gives the approximate percentile of the distribution of l as a simple function of the χ^2 percentile. The α percentile of l is

$$u_\alpha + \frac{u_\alpha}{n}\left\{\frac{1}{G_1}\operatorname{tr}(\boldsymbol{Q}^{-1}\boldsymbol{C}_2)\sigma^2 L - \frac{1}{2}(G_1 - 2 - u_\alpha)\right\}, \tag{3.7}$$

where u_α is the α percentile of the χ^2 distribution with G_1 degrees of freedom. The unknown parameters $tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_2)$ can be estimated by the consistent estimator of $\boldsymbol{Q}$ and $\boldsymbol{C}_2$, which are

$$\hat{\boldsymbol{Q}}^{-1} = n\begin{pmatrix} \boldsymbol{Y}_2'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_2 - \hat{\lambda}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2 & \boldsymbol{Y}_2'\boldsymbol{Z}_1 \\ \boldsymbol{Z}_1'\boldsymbol{Y}_2 & \boldsymbol{Z}_1'\boldsymbol{Z}_1 \end{pmatrix}^{-1} \tag{3.8}$$

and

$$\hat{\boldsymbol{C}}_2 = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2/q - \frac{1}{\hat{\sigma}^4}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2/q^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \tag{3.9}$$

where we use the notations that $\hat{\sigma}^2 = \hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}/q$ and $\hat{\boldsymbol{b}} = (1, -\hat{\boldsymbol{\beta}}_{LI}')'$. We propose a modified LR test (LR_{m1}) using the critical value

$$u_\alpha + \frac{u_\alpha}{n}\left\{\frac{1}{G_1}\operatorname{tr}(\hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{C}}_2)\hat{\sigma}^2 L - \frac{1}{2}(G_1 - 2 - u_\alpha)\right\}, \tag{3.10}$$

instead of u_α.

4. Large-K_n asymptotic approximation of the null distribution of the LR statistic

In this section, we develop an alternative approximation using "large-K_n asymptotics" in the case of non-normal disturbances. We consider the sequence which allows the number of the (excluded) instruments (K_2) to grow with the number of observations (n). We assume that

$$\begin{aligned} & n \to \infty, \\ & K/n = c_1 + O(n^{-1}), \qquad (0 \le c_1 < 1) \\ & K/q = c_2 + O(n^{-1}), \qquad (0 \le c_2 < \infty) \end{aligned} \tag{4.1}$$

where we defined $q = n - K$.

Under the sequences (4.1), the next theorem follows. The derivation is provided in Appendix B.

THEOREM 2. *Assume that* $E[\|\boldsymbol{v}_i\|^6]$ *are bounded, and that there exists a constant positive definite matrix* $\boldsymbol{Q} = p\lim_{n\to\infty} n^{-1}\tilde{\boldsymbol{Q}}$ *such that* $\boldsymbol{Q} = n^{-1}\tilde{\boldsymbol{Q}} + O_p(n^{-1})$. *Then, under* H_0, *under the sequences* (4.1),

$$l \xrightarrow{d} \frac{1}{\sigma^2}\boldsymbol{U}'\boldsymbol{Q}\boldsymbol{U}, \tag{4.2}$$

where $\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Psi})$, *and*

$$\begin{aligned} \boldsymbol{\Psi} = {} & \sigma^2\boldsymbol{Q}^{-1} + c_1(1+c_2)\boldsymbol{Q}^{-1}\left[\begin{pmatrix} \boldsymbol{\Omega}_{22}\sigma^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1} \\ & + \boldsymbol{Q}^{-1}[(\boldsymbol{\Xi}_3 + \boldsymbol{\Xi}_3') + \eta\boldsymbol{\Gamma}_4]\boldsymbol{Q}^{-1}. \end{aligned}$$

The limit distribution can also be expressed as $r_1\chi^2_{1,1} + \cdots + r_p\chi^2_{1,G_1}$, *where the* $\chi^2_{1,j}$*s are independent* χ^2 *variables with one degree of freedom and the weights* $r_1, \ldots, r_{G_1}$ *are the* G_1 *eigenvalues of* $\boldsymbol{Q}\boldsymbol{\Psi}/\sigma^2$. *Here we have used the notations that* $\boldsymbol{\Xi}_3 = p\lim_{n\to\infty} \boldsymbol{D}_2'\frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i[(1+c_2)a_{ii}^{(n)} - c_2]E[u_i^2\boldsymbol{w}_{2i}']$, $\eta = (1+c_2)^2 p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n a_{ii}^{(n)2} - c_2^2$, $a_{ii}^{(n)} = \boldsymbol{z}_i'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_i$, $\boldsymbol{\Gamma}_4 = E(u_i^2\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}') - \sigma^2E[\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}']$, *and* $\boldsymbol{w}_{2i} = (\boldsymbol{v}_{2i}' \;\; \boldsymbol{0}')' - u_i\boldsymbol{q}_2$.

We can estimate the weights $r_1, \ldots, r_{G_1}$ using consistent estimators $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{\Psi}}$. In the case of the normal disturbances, $\boldsymbol{\Psi}$ is identical to the Bekker (1994) variance, and $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{\Psi}}$ can be defined by (3.8) and

$$\begin{aligned} \hat{\boldsymbol{\Psi}} = {} & \hat{\sigma}^2\hat{\boldsymbol{Q}}^{-1} \\ & + \frac{K}{n}(1+\hat{\lambda})\hat{\boldsymbol{Q}}^{-1} \\ & \times \begin{pmatrix} \frac{1}{q}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2\hat{\sigma}^2 - \frac{1}{q^2}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\hat{\boldsymbol{Q}}^{-1}, \end{aligned} \tag{4.3}$$

where $\hat{\sigma}^2 = \frac{1}{q}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{b}} = (1, -\hat{\boldsymbol{\beta}}_{LI}')'$, respectively.

In case of non-normality, $\mathbf{\Psi}$ has additional terms depending on the third and fourth order moments of the disturbances, which makes it complicated. However, Anderson *et al.* (2006) and Matsushita (2006) investigated the effects of these terms and found that they have little effects even when the distributions of the disturbances deviate from the normal. We also investigate the effects of the third and fourth order moments using Monte Carlo experiments in the next section.

We call the LR test, of which the critical value is computed from the asymptotic null distribution derived in Theorem 2, the large-K LR test (LR_{largeK}).

5. Size comparison with the CLR statistic

5.1. The case of normal disturbances

We conduct size comparisons of the two types of modified LR tests, LR_{m1} and LR_{largeK} with the CLR test by Moreira (2003) and the large K t-test (Bekker (1994), Matsushita (2006), for instance).

We considered models with two endogenous variables, i.e., $G_1 = 1$. In this case, the distributions of all the statistics considered here depend only on the key parameters used by Anderson *et al.* (1982), which are K_2, the number of excluded exogenous variables; $n - K$, the number of degrees of freedom in $\hat{\mathbf{\Omega}}$;

$$\delta^2 = \frac{\mathbf{\Pi}_{22}' \boldsymbol{A}_{22.1} \mathbf{\Pi}_{22}}{\omega_{22}}, \tag{5.1}$$

the noncentrality parameter associated with (2.1); and

$$\alpha = \frac{\omega_{22}\beta - \omega_{21}}{|\mathbf{\Omega}|^{1/2}} = -\frac{\rho}{(1-\rho^2)^{1/2}}, \tag{5.2}$$

where ρ is the correlation coefficient between $\boldsymbol{u}$ and $\boldsymbol{v}_2$. The numerator of the noncentrality parameter δ^2 represents the additional explanatory power due to $\boldsymbol{y}_{2i}$ over $\boldsymbol{z}_{1i}$ in the structural equation, and its denominator is the error variance of $\boldsymbol{y}_{2i}$. Hence, the noncentrality parameter δ^2 determines how well the equation is defined in the simultaneous equations system.

We generate a set of random numbers by using the two-equation system

$$\boldsymbol{y}_1 = \boldsymbol{y}_2 \beta^{(0)} + \boldsymbol{Z}_1 \gamma^{(0)} + \boldsymbol{u}, \tag{5.3}$$

and

$$\boldsymbol{y}_2 = \boldsymbol{Z} \mathbf{\Pi}_2^{(0)} + \boldsymbol{V}_2, \tag{5.4}$$

where $K_1 = 1$, $\boldsymbol{Z} \sim N(\mathbf{0}, I_K \otimes I_n)$, $(\boldsymbol{u}, \boldsymbol{V}_2) \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes I_n)$, $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and the true values of parameters $\beta^{(0)} = \gamma^{(0)} = 0$. We have controlled the values of δ^2 by choosing a real value of c and setting $(1 + K_2) \times 1$ vector $\mathbf{\Pi}_2^{(0)} = c(1, \ldots, 1)'$.

Tables 1–4 contain empirical sizes of the statistics at 10, 5, and 1% levels for various values of δ^2, K_2, and α. The number of repetitions is 10,000 in each experiment. We also use 5,000 realizations each of $\chi^2(1)$ and $\chi^2(K_2 - 1)$ random variables to simulate the critical values of Moreira's CLR statistic.

Table 1. Empirical sizes of statistics that test $H_0 : \beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 5$.

		$\alpha = 0.3$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	12.7	10.8	11.5	5.6	8.1
	5%	7.5	5.5	6.6	2.2	4.4
	1%	1.8	1.0	1.4	0.3	0.5
$K_2 = 5$	10%	14.5	11.6	12.3	7.5	9.8
	5%	8.8	6.1	7.1	3.9	6.7
	1%	2.4	1.3	1.7	0.7	1.3
$K_2 = 30$	10%	18.2	14.3	14.5	10.6	11.2
	5%	10.3	7.6	8.0	5.1	6.7
	1%	3.8	2.4	2.6	1.1	1.4

Table 2. Empirical sizes of statistics that test $H_0 : \beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 5$.

		$\alpha = 1$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	12.2	10.7	11.3	9.0	9.1
	5%	7.0	5.8	6.6	6.3	5.1
	1%	1.8	1.2	1.5	2.3	1.1
$K_2 = 5$	10%	13.0	11.0	11.5	8.9	10.7
	5%	7.0	5.2	5.9	5.4	4.9
	1%	1.9	1.0	1.4	2.0	1.3
$K_2 = 30$	10%	15.3	13.0	13.3	10.9	10.7
	5%	8.3	6.5	7.0	5.4	6.5
	1%	2.7	1.8	2.0	1.3	1.5

Table 3. Empirical sizes of statistics that test $H_0 : \beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 1$.

		$\alpha = 0.3$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	20.0	14.5	14.4	1.9	10.1
	5%	11.0	7.2	6.1	0.8	6.6
	1%	3.6	1.8	2.1	0.1	1.8
$K_2 = 5$	10%	27.1	17.3	16.0	4.5	13.1
	5%	18.0	10.2	10.8	1.7	7.6
	1%	6.2	2.9	2.1	0.2	2.9
$K_2 = 30$	10%	36.1	22.5	22.5	9.0	14.0
	5%	27.3	14.6	17.5	4.8	7.8
	1%	16.9	7.2	6.5	1.3	2.6

From the tables, when δ^2/K_2 is larger than five, all tests have reliable size properties. The LR_{m1} test improves upon the LR test, which is prone to reject H_0 more than it should, in all cases. When δ^2/K_2 is small, the size properties of the LR test become quite poor. (Tables 3–4) The observed size of the LR test at

Table 4. Empirical sizes of statistics that test $H_0 : \beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 1$.

		$\alpha = 1$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	18.1	13.5	11.4	9.9	9.8
	5%	9.5	6.7	6.1	6.8	5.1
	1%	2.7	1.5	1.9	2.4	1.8
$K_2 = 5$	10%	20.3	13.7	11.7	10.4	10.3
	5%	13.2	7.8	6.9	7.2	6.5
	1%	4.9	2.1	2.3	3.2	2.0
$K_2 = 30$	10%	25.8	17.1	19.2	9.3	10.4
	5%	19.1	11.3	11.9	6.0	7.5
	1%	9.1	4.3	3.8	2.7	2.2

Table 5. Empirical sizes of statistics that test H_0 (The Cases of Non-normal Disturbances): $\beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 1$.

		$u_i = (\chi^2(3) - 3)/\sqrt{6}$, $\alpha = 1$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	16.0	12.0	11.1	10.5	9.3
	5%	9.1	6.7	6.6	7.0	5.2
	1%	2.6	1.4	1.7	2.8	1.5
$K_2 = 5$	10%	21.0	14.1	13.2	10.8	12.3
	5%	13.7	8.3	7.7	7.5	7.2
	1%	5.1	2.5	2.7	3.1	2.3
$K_2 = 30$	10%	25.6	16.9	17.4	8.8	11.9
	5%	18.2	10.7	11.3	5.9	7.2
	1%	8.9	4.2	4.9	2.6	2.3

Table 6. Empirical sizes of statistics that test H_0 (The Cases of Non-normal Disturbances): $\beta = \beta_0$ with $n - K = 30$, $\delta^2/K_2 = 1$.

		$u_i = t(3)$, $\alpha = 1$				
		LR	LR_{m1}	CLR	t_{largeK}	LR_{largeK}
$K_2 = 2$	10%	16.7	12.9	12.2	9.7	10.3
	5%	10.0	7.2	6.9	6.5	5.5
	1%	2.8	1.7	1.9	2.4	1.5
$K_2 = 5$	10%	20.6	13.9	13.0	10.6	12.1
	5%	13.5	8.2	7.7	7.1	6.8
	1%	5.0	2.4	2.4	2.9	2.2
$K_2 = 30$	10%	25.2	17.0	17.5	8.9	11.9
	5%	18.1	10.8	11.4	5.6	6.8
	1%	8.9	4.0	4.7	2.1	2.2

the 5% asymptotic critical value can be over 20% when K_2 is thirty, for instance. One interesting finding is that the size properties of the CLR test is also poor when the number of the instruments is large. Since the CLR test is known to be robust to weak instruments and has good power properties, this finding seems to have some importance. When the number of the instruments is small (less than five), the LR_{m1} test and CLR test have reasonable size properties. However, as the number of the instruments increases, the LR_{m1} test as well as the CLR is size-distorted. The LR_{largeK} test has the best size properties when the number of the instruments is larger than five, while it is size-distorted when the degrees of overidentifiability is less than two.

5.2. The case of non-normal disturbances

Since the distributions of the LR statistics depend on the distributions of the disturbances, we have investigated the effects of the non-normality of disturbances. We calculated a large number of cases in which the distributions of disturbances are skewed ($\chi^2(3)$) and have long tails ($t(3)$). We have chosen the case of $n - K = 30$, $\alpha = 1$, and $\delta^2/K_2 = 1$ and reported the observed sizes at the 10%, 5% and 1% asymptotic critical values of LR, LR_{m1}, CLR, t_{largeK} and LR_{largeK} in Tables 5 and 6. We calculated the critical values of the LR_{m1}, t_{largeK} and LR_{largeK} using the asymptotic variance assuming normal disturbances. From these experiments, we see that size properties of all these statistics, which are derived under the assumption of normal disturbances, are approximately valid even if the distributions of disturbances deviate from normal.

6. Conclusions

In this paper, we have made two types of asymptotic approximations of the distribution of the likelihood ratio statistics under the null hypothesis, and propose modifications of the LR test. The Monte Carlo experiments show that, when the instruments are weak, the size properties of the LR test become quite poor, and the LR_{m1} test (based on the asymptotic expansion) improves upon the LR test when the number of the instruments is small and δ^2/K_2 is more than one. However, the LR_{m1} test can be size distorted when the number of the instruments is large. One finding is that the size properties of the CLR test can also be poor when the number of the instruments is large. The LR_{largeK} test (based on large-K_n asymptotics) has the best size properties when the number of the instruments is large and δ^2/K_2 is more than one.

Appendix A
Derivation of Theorem 1

We make use of the results of Kunitomo *et al.* (1983) and Morimune and Tsukuda (1984). The variance ratio $\hat{\lambda}$ defined by (2.12) is stochastically expanded as

$$(n-K)\hat{\lambda} = \hat{\lambda}^{(0)} + \frac{1}{\sqrt{n}}\hat{\lambda}^{(1)} + \frac{1}{n}\hat{\lambda}^{(2)} + O_p(n^{-3/2}), \tag{A.1}$$

where

$$\hat{\lambda}^{(0)} = \boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}/\sigma^2,$$

$$\hat{\lambda}^{(1)} = -\frac{1}{\sigma^2}\{2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}(x - 2\boldsymbol{q}_2'\boldsymbol{e}^{(0)})\},$$

$$\begin{aligned}\hat{\lambda}^{(2)} = \frac{1}{\sigma^2}\Bigg\{ & \left[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}\right]'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\left[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}\right] \\ & - 2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})[\boldsymbol{e}^{(1)} - (x - 2\boldsymbol{q}_2'\boldsymbol{e}^{(0)})\boldsymbol{e}^{(0)}] \\ & + \frac{1}{\sigma^2}\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}\Big[2(\boldsymbol{w}_{12} - \boldsymbol{\beta}'\boldsymbol{W}_{22}, \boldsymbol{0})\boldsymbol{e}^{(0)} - \frac{\sigma^2}{3}(x^2 - 2) \\ & + \sigma^2(x - 2\boldsymbol{q}_2'\boldsymbol{e}^{(0)})^2 + 2\sigma^2\boldsymbol{q}_2'\boldsymbol{e}^{(1)} - \sigma^2\boldsymbol{e}^{(0)\prime}(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{e}^{(0)}\Big]\Bigg\},\end{aligned}$$

where

$$\boldsymbol{e}^{(0)} = \boldsymbol{Q}^{-1}\boldsymbol{X}'\boldsymbol{u}/\sqrt{n},$$

and

$$\boldsymbol{e}^{(1)} = \boldsymbol{Q}^{-1}\{(\boldsymbol{V}_2, \boldsymbol{0})'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u} - \boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}\boldsymbol{q}_2 - \boldsymbol{X}'(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)}/\sqrt{n}\},$$

defining $\boldsymbol{w}_{12} = \sqrt{n}[\frac{1}{n}\boldsymbol{v}_1'\bar{\boldsymbol{P}}_Z\boldsymbol{V}_2 - \boldsymbol{\omega}_{12}]$, $\boldsymbol{W}_{22} = \sqrt{n}[\frac{1}{n}\boldsymbol{V}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{V}_2 - \boldsymbol{\Omega}_{22}]$, and $x = (1, -\boldsymbol{\beta}')\sqrt{n}[\frac{1}{n}\boldsymbol{V}\bar{\boldsymbol{P}}_Z\boldsymbol{V} - \boldsymbol{\Omega}](1, -\boldsymbol{\beta}')'$ which is distributed with mean zero and variance two.

Similarly λ_0 defined by (2.11) is expanded as

$$(n - K)\lambda_0 = \lambda_0^{(0)} + \frac{1}{\sqrt{n}}\lambda_0^{(1)} + \frac{1}{n}\lambda_0^{(2)} + O_p(n^{-3/2}), \tag{A.2}$$

where

$$\lambda_0^{(0)} = \boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}/\sigma^2,$$

$$\lambda_0^{(1)} = -\frac{1}{\sigma^2}[\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}x],$$

and

$$\lambda_0^{(2)} = \frac{1}{\sigma^2}\left[\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}\left\{-\frac{1}{3}(x^2 - 2) + x^2\right\}\right].$$

Hence the test statistic is stochastically expanded as

$$l = l^{(0)} + \frac{1}{\sqrt{n}}l^{(1)} + \frac{1}{n}l^{(2)} + O_p(n^{-3/2}), \tag{A.3}$$

where

$$l^{(0)} \equiv v = \lambda_0^{(0)} - \hat{\lambda}^{(0)} = \frac{1}{\sigma^2}\boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u},$$

$$\begin{aligned}l^{(1)} &= \lambda_0^{(1)} - \hat{\lambda}^{(1)} \\ &= \frac{1}{\sigma^2}\{2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} - \boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u}x \\ &\qquad - 2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}\boldsymbol{q}_2'\boldsymbol{e}^{(0)}\},\end{aligned}$$

and

$$
\begin{aligned}
l^{(2)} &= \lambda_0^{(2)} - \hat{\lambda}^{(2)} \\
&= \frac{1}{\sigma^2}\Big\{-\left[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}\right]'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\left[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}\right] \\
&\qquad + 2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})[\boldsymbol{e}^{(1)} - (x - 2\boldsymbol{q}_2'\boldsymbol{e}^{(0)})\boldsymbol{e}^{(0)}] \\
&\qquad - \frac{1}{\sigma^2}\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}[2(\boldsymbol{w}_{12} - \boldsymbol{\beta}'\boldsymbol{W}_{22}, \boldsymbol{0})\boldsymbol{e}^{(0)} \\
&\qquad - 4\sigma^2\boldsymbol{q}_2'\boldsymbol{e}^{(0)}x + 2\sigma^2\boldsymbol{q}_2'\boldsymbol{e}^{(1)} - \sigma^2\boldsymbol{e}^{(0)'}(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{e}^{(0)}] \\
&\qquad + \boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u}\left[-\frac{1}{3}(x^2 - 2) + x^2\right]\Big\}.
\end{aligned}
$$

We shall derive an asymptotic expansion of the distribution of l by inverting the characteristic function of l up to order n^{-1}:

$$
\begin{aligned}
\text{(A.4)}\qquad C(t) \;=\; & E(\exp(itv)) + \frac{1}{\sqrt{n}}E(itE(l^{(1)} \mid v)\exp(itv)) \\
& + \frac{1}{n}E(itE(l^{(2)} \mid v)\exp(itv)) \\
& + \frac{1}{2n}E(i^2tE(l^{(1)^2} \mid v)\exp(itv)) + O(-n^{-3/2}).
\end{aligned}
$$

Validity of the method can be given following the same method used by Kunitomo *et al.* (1983). To calculate the conditional expectations given the first order term v, we use the following formula which was developed by Morimune and Tsukuda (1984):

$$
\text{(A.5)}\qquad E(\boldsymbol{e}^{(0)'}\boldsymbol{C}_j\boldsymbol{e}^{(0)} \mid v) = \frac{v}{G_1}\sigma^2\,\mathrm{tr}(\boldsymbol{Q}^{-1}\boldsymbol{C}_j), \qquad j = 1, 2,
$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are defined by (3.2) and (3.3) respectively.

Then we have the conditional expectations given the first order term v as follows:

$$
\text{(A.6)}\qquad E(l^{(1)} \mid v) = 0,
$$

$$
\text{(A.7)}\qquad E(l^{(2)} \mid v) = 2v + \mathrm{tr}(\boldsymbol{Q}^{-1}\boldsymbol{C}_2\sigma^2)L,
$$

$$
\text{(A.8)}\qquad E(l^{(1)^2} \mid v) = 4\,\mathrm{tr}(\boldsymbol{Q}^{-1}\boldsymbol{C}_2\sigma^2)L + 2v^2.
$$

The probablity $P(l \le \xi)$ is approximated to the order n^{-1} by the Fourier inverse transformation of the characteristic function (A.4). The inverse transformation of the first term is $G_{G_1}(\xi)$ which is the χ^2 cdf function with G_1 degrees of freedom. We also use the next Fourier Inversion formula which was developed by Kunitomo *et al.* (1983):

$$
\begin{aligned}
\text{(A.9)}\qquad & \int_{x=0}^{\xi}\frac{1}{2\pi}\int_t(-it)^p\exp(-itx)E[\exp(itv)v^j]dtdx \\
& = \frac{2^j\Gamma\left(\dfrac{G_1}{2} + j\right)}{\Gamma\left(\dfrac{G_1}{2}\right)} \cdot g_{G_1+2j}^{(p-1)}(\xi),
\end{aligned}
$$

where $i = \sqrt{-1}$, j is any integer $(G_1 + 2j > 0)$, and $g^{(p-1)}_{G_1+2j}(\xi)$ is the $(p-1)$-th order derivative of g_{G_1+2j}, which is the χ^2 density function with $G_1 + 2j$ degrees of freedom. Theorem 1 follows after simplifications.

Appendix B

Derivation of Theorem 2

The variance ratio (2.12) is exactly rewritten as

$$\text{(B.1)} \qquad \hat{\lambda} = \frac{\left\{\boldsymbol{u} - \frac{1}{\sqrt{n}}[\boldsymbol{Z}\boldsymbol{D}_2 + (\boldsymbol{V}_2, \boldsymbol{0})]\hat{\boldsymbol{e}}\right\}' \boldsymbol{P}_Z \left\{\boldsymbol{u} - \frac{1}{\sqrt{n}}[\boldsymbol{Z}\boldsymbol{D}_2 + (\boldsymbol{V}_2, \boldsymbol{0})]\hat{\boldsymbol{e}}\right\}}{\left\{\boldsymbol{u} - \frac{1}{\sqrt{n}}(\boldsymbol{V}_2, \boldsymbol{0})\hat{\boldsymbol{e}}\right\}' \bar{\boldsymbol{P}}_Z \left\{\boldsymbol{u} - \frac{1}{\sqrt{n}}(\boldsymbol{V}_2, \boldsymbol{0})\hat{\boldsymbol{e}}\right\}}$$

where

$$\hat{\boldsymbol{e}} = \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\gamma}}_{LI} - \boldsymbol{\gamma} \end{pmatrix}$$

and

$$\boldsymbol{D} = (\boldsymbol{D}_1 \;\; \boldsymbol{D}_2) = \left(\begin{pmatrix} \boldsymbol{\pi}_{11} \\ \boldsymbol{\pi}_{21} \end{pmatrix} \;\; \begin{pmatrix} \boldsymbol{\Pi}_{12} & \boldsymbol{I}_{K_1} \\ \boldsymbol{\Pi}_{22} & \boldsymbol{0} \end{pmatrix} \right).$$

The large-K_n asymptotics of $\hat{\boldsymbol{e}}$ is expanded in terms of $n^{-1/2}$ as

$$\text{(B.2)} \qquad \hat{\boldsymbol{e}} = \boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{e}^{(1)} + O_p(n^{-1}).$$

The terms of $\boldsymbol{e}^{(0)}$ and $\boldsymbol{e}^{(1)}$ are given in Matsushita (2006) as

$$\text{(B.3)} \qquad \boldsymbol{e}^{(0)} = \boldsymbol{Q}^{-1}\left[\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{u} + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u} - \frac{\sqrt{c_1 c_2}}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\right],$$

$$\begin{aligned} \text{(B.4)} \qquad \boldsymbol{e}^{(1)} = -\boldsymbol{Q}^{-1}\Bigg[&\left\{\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'(\boldsymbol{V}_2 \;\; \boldsymbol{0}) + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z(\boldsymbol{V}_2 \;\; \boldsymbol{0})\right. \\ &\left. - \sqrt{c_1 c_2}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2 \;\; \boldsymbol{0})\right\}\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{W}_2'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\ &- \frac{n}{q}\lambda^{(1)}\left[\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^2\right]\boldsymbol{e}^{(0)} \\ &+ \sqrt{\frac{n}{q}}\lambda^{(1)}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\Bigg]. \end{aligned}$$

We first make the large-K_n stochastic expansion of the variance ratio (2.12). Substituting (B.2) into (B.1), the numerator of the variance ratio devided by K becomes

$$\text{(B.5)} \qquad \sigma^2 + \frac{1}{\sqrt{n}}\left\{\sqrt{\frac{n}{K}}\sqrt{K}\left(\frac{1}{K}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u} - \sigma^2\right) - 2(\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(0)}\right\}$$

$$+\frac{1}{n}\Bigg\{-2\sqrt{\frac{n}{K}}\sqrt{K}\left[\frac{1}{K}\boldsymbol{b}_0'\boldsymbol{V}'\boldsymbol{P}_Z(\boldsymbol{V}_2,\boldsymbol{0})-(\boldsymbol{b}_0'\boldsymbol{\Omega},\boldsymbol{0})\boldsymbol{J}_2\right]\boldsymbol{e}^{(0)}$$
$$-2\frac{n}{K}\frac{1}{\sqrt{n}}\boldsymbol{u}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)}+\frac{n}{K}\boldsymbol{e}^{(0)\prime}\frac{1}{n}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)}$$
$$+\boldsymbol{e}^{(0)\prime}\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{e}^{(0)}-2(\boldsymbol{b}_0'\boldsymbol{\Omega},\boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(1)}\Bigg\}$$

to terms of $O_p(n^{-1})$. The denominator devided by $q(=n-K)$ becomes

$$\text{(B.6)}\qquad \sigma^2+\frac{1}{\sqrt{n}}\left\{\sqrt{\frac{n}{q}}\sqrt{q}\left(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}-\sigma^2\right)-2(\boldsymbol{b}_0'\boldsymbol{\Omega},\boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(0)}\right\}$$
$$+\frac{1}{n}\Bigg\{-2\sqrt{\frac{n}{q}}\sqrt{q}\left[\frac{1}{q}\boldsymbol{b}_0'\boldsymbol{V}'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2,\boldsymbol{0})-(\boldsymbol{b}_0'\boldsymbol{\Omega},\boldsymbol{0})\boldsymbol{J}_2\right]\boldsymbol{e}^{(0)}$$
$$-2\frac{n}{K}\frac{1}{\sqrt{n}}\boldsymbol{u}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)}+\frac{n}{K}\boldsymbol{e}^{(0)\prime}\frac{1}{n}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)}$$
$$+\boldsymbol{e}^{(0)\prime}\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{e}^{(0)}-2(\boldsymbol{b}_0'\boldsymbol{\Omega},\boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(1)}\Bigg\}$$

to terms of $O_p(n^{-1})$.

Multiplying Taylor's expansion of the inverse of (B.6) to (B.5) it follows the large-K_n stochastic expansion of the variance ratio (2.12):

$$\text{(B.7)}\qquad \hat{\lambda}=\hat{\lambda}^{(0)}+\frac{1}{\sqrt{n}}\hat{\lambda}^{(1)}+\frac{1}{n}\hat{\lambda}^{(2)}+O_p(n^{-3/2}),$$

where

$$\hat{\lambda}^{(0)}=c_2,$$
$$\hat{\lambda}^{(1)}=\frac{c_2}{\sigma^2}\left\{\sqrt{\frac{n}{K}}\left(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}\right)-\sqrt{\frac{n}{q}}\left(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\right)\right\},$$
$$\hat{\lambda}^{(2)}=\frac{c_2}{\sigma^2}\Bigg\{-\frac{n}{K}\boldsymbol{e}^{(0)\prime}\boldsymbol{Q}\boldsymbol{e}^{(0)}$$
$$-\sqrt{\frac{n}{q}}\frac{1}{\sigma^2}\left[\sqrt{\frac{n}{K}}\left(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}\right)-\sqrt{\frac{n}{q}}\left(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\right)\right]$$
$$\times\left[\sqrt{q}\left(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}-\sigma^2\right)\right]\Bigg\}.$$

Similarly λ_0 defined by (2.11) is expanded as

$$\text{(B.8)}\qquad \lambda_0=\lambda_0^{(0)}+\frac{1}{\sqrt{n}}\lambda_0^{(1)}+\frac{1}{n}\lambda_0^{(2)}+O_p(n^{-3/2}),$$

where

$$\lambda_0^{(0)}=c_2,$$

$$\lambda_0^{(1)} = \frac{c_2}{\sigma^2}\left\{\sqrt{\frac{n}{K}}\left(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}\right) - \sqrt{\frac{n}{q}}\left(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\right)\right\},$$

$$\lambda_0^{(2)} = -\frac{c_2}{\sigma^4}\sqrt{\frac{n}{q}}\left[\sqrt{\frac{n}{K}}\left(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}\right) - \sqrt{\frac{n}{q}}\left(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\right)\right] \times \left[\sqrt{q}\left(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2\right)\right].$$

Hence we have the relation that

$$l = \frac{n-K}{n}(\lambda_0^{(2)} - \hat{\lambda}^{(2)}) = \frac{1}{\sigma^2}\boldsymbol{e}^{(0)\prime}\boldsymbol{Q}\boldsymbol{e}^{(0)} + o_p(1). \tag{B.9}$$

Anderson *et al.* (2006) show that

$$\boldsymbol{e}^{(0)} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}). \tag{B.10}$$

Then we have the desired result.

Acknowledgements

The author wishes to thank Naoto Kunitomo and the referees for helpful comments.

References

Anderson, T. W. (2005). Origins of the limited information maximum likelihood and two-stage least squares estimators, *Journal of Econometrics*, **127**, 1–16.

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

Anderson, T. W., Kunitomo, N. and Matsushita, Y. (2006). A New Light from Old Wisdom: Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models, Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2005/2005cf321.pdf).

Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2006). Optimal two-sided invariant similar tests for instrumental variables regression, *Ecoonometrica*, **74**, 715–752.

Angrist, J. D. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings, *Quarterly Journal of Economics*, **106**, 979–1014.

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

Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, *Journal of the American Statistical Association*, **90**, 443–450.

Hansen, C., Hausman, J. and Newey, W. K. (2006). Estimation with many instrumental variables, Unpublished manuscript.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression, *Econometrica*, **70**(5), 1781–1803.

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., Morimune, K. and Tsukuda, Y. (1983). Asymptotic expansions of the distributions of the test statistics for overidentifying restrictions in a system of simultaneous equations, *International Economic Review*, **24**, 199–215.

Matsushita, Y. (2006). t-Tests in A Structural Equation with Many Instruments, Discussion Paper CIRJE-F-467, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007/2007cf467.pdf).

Moreira, M. (2001). Tests with Correct Size when Instruments can be Arbitrarily Weak, Center for Laor Economics Working Paper Series, 37, UC Berkeley.

Moreira, M. (2003). A conditional likelihood ratio test for structural models, *Econometrica*, **71**(4), 1027–1048.

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

Morimune, K. and Tsukuda, Y. (1984). Testing a subset of coefficients in a structural equation, *Econometrica*, **52**, 427–448.

Nelson, C. R. and Startz, R. (1990a). Some further results on the exact small sample properties of the instrumental variables estimator, *Econometrica*, **58**, 967–976.

Nelson, C. R. and Startz, R. (1990b). The distributions of the instrumental variables estimator and its t ratio when the instrument is a poor one, *Journal of Business*, **63**, S125–S140.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments, *Econometrica*, **65**, 557–586.