CIRJE-F-466

Approximate Distributions of the Likelihood Ratio Statistic in a Structural Equation with Many Instruments

Yukitoshi Matsushita
CIRJE, Faculty of Economics, University of Tokyo

February 2007

Approximate Distributions of the Likelihood Ratio Statistic in a Structural Equation with Many Instruments *

Yukitoshi Matsushita †

October 2006

Abstract

This paper studies the properties of 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) statistics under the null hypothesis $H_0 : \beta = \beta_0$: the (large sample) asymptotic expansion and the large-K_n asymptotic theory. The size comparison of two modified LR tests based on these two asymptotics is made with Moreira's conditional likelihood ratio (CLR) test and the large K t-test.

*I would like to thank Naoto Kunitomo for his support and suggestions.

†Graduate School of economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN; Running Head: LR-test with Many Instruments; Key Words: Many instruments, Asymptotic expansion, large-K_n asymptotics

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 IV statistics can be poor. See Nelson and Startz (1990a,b), Bound, Jaeger, and Baker (1995), 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 literatures is Angrist and Krueger (1991), where they used 178 instruments in one of their specifications. Bound, Jaeger, and Baker (1995) shows that the properties of the TSLS estimator can be poor in the face of many weak instruments even when the sample size is huge.

In order to overcome these problems, several new statistical procedures have recently proposed. For the 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-Rubin(1949). Kleibergen (2002) and Moreira (2001) proposed a score-type statistic, while Moreira (2003) proposed a conditional likelihood ratio (CLR) test, both of which are shown to be robust to weak instruments, too. Among these testing procedures, the CLR test has been found to dominate the other tests in terms of power. Andrews, Moreira, and Stock (2006) show that the CLR test is quite close to being uniformly most powerful invariant among a class of two-sided test.

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, 1982) and

Morimune (1983) were the earlier developers of the large-K_n asymptotics, and they derived asymptotic expansions of the distributions of the k-class estimators including the two stage least squares (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 even when the number of instruments is not large. Hansen, Hausman and Newey (2006) consider the same model and show that Bekker (1994) standard error corrects the size problem. Matsushita (2006) have derived an asymptotic expansion of the distributions of 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 the 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 distribution of the LR statistics under the null hypothesis: the (large sample) asymptotic expansion (in the case of normal disturbances) and the 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 distribution of the LR statistic under the null hypothesis is given in Section 3, and an approximate 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 $K(K=K_1+K_2)$ 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 $\boldsymbol{u}=\boldsymbol{v}_1-\boldsymbol{V}_2\boldsymbol{\beta}$, $\boldsymbol{\gamma}=\boldsymbol{\pi}_{11}-\boldsymbol{\Pi}_{12}\boldsymbol{\beta}$, 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 $\omega_{11}-2\boldsymbol{\beta}'\omega_{21}+\boldsymbol{\beta}'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}$.

We define, for any full column matrix $\boldsymbol{F}$,

$$\boldsymbol{P}_F=\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}', \qquad \bar{\boldsymbol{P}}_F=\boldsymbol{I}-\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'. \tag{2.5}$$

The LIML estimator of $(\boldsymbol{\beta}'\ \gamma')'$ is $(\hat{\boldsymbol{\beta}}'_{LI}\ \hat{\gamma}'_{LI})'$ satisfying

$$\left\{\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\\ \boldsymbol{Z}_1'\end{pmatrix}\boldsymbol{P}_Z(\boldsymbol{y}_1\ \boldsymbol{Y}_2\ \boldsymbol{Z}_1)-\hat{\lambda}\begin{pmatrix}\boldsymbol{y}_1'\\ \boldsymbol{Y}_2'\\ \boldsymbol{Z}_1'\end{pmatrix}\bar{\boldsymbol{P}}_Z(\boldsymbol{y}_1\ \boldsymbol{Y}_2\ \boldsymbol{Z}_1)\right\}\begin{pmatrix}1\\ -\hat{\boldsymbol{\beta}}_{LI}\\ -\hat{\gamma}_{LI}\end{pmatrix}=\boldsymbol{0}, \tag{2.6}$$

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

$$\left| \begin{pmatrix} \boldsymbol{y}_1' \\ \boldsymbol{Y}_2' \\ \boldsymbol{Z}_1' \end{pmatrix} \boldsymbol{P}_Z(\boldsymbol{y}_1 \ \boldsymbol{Y}_2 \ \boldsymbol{Z}_1) - \lambda \begin{pmatrix} \boldsymbol{y}_1' \\ \boldsymbol{Y}_2' \\ \boldsymbol{Z}_1' \end{pmatrix} \bar{\boldsymbol{P}}_Z(\boldsymbol{y}_1 \ \boldsymbol{Y}_2 \ \boldsymbol{Z}_1) \right| = 0. \tag{2.7}$$

When the instruments are weakly correlated to the included endogenous variables, approximations based on the standard asymptotic theory are not satisfactory. In order to overcome this problem, several new statistical procedures robust to weak instruments have recently proposed. The Anderson-Rubin (AR) test is a fundamental building block for developing reliable inference procedures with weak instruments. Kleibergen (2002) and Moreira (2001) proposed a score-type statistic, while Moreira (2003) proposed a conditional likelihood ratio (CLR) test.

- **Anderson-Rubin (AR) Test**

  Anderson and Rubin (AR) statsitic is given by

$$AR = \frac{(1, -\boldsymbol{\beta}_0{}')\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}(1, -\boldsymbol{\beta}_0{}')'}{(1, -\boldsymbol{\beta}_0{}')\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}(1, -\boldsymbol{\beta}_0{}')'/(n-K)}. \tag{2.8}$$

  Because, under the null hypothesis, we have

$$AR = \frac{\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}}{\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}/(n-K)}, \tag{2.9}$$

  the null distribution of the AR statistic does not depend on instrument quality. Thus the AR test is one of the testing procedures which are robust to weak instruments. Under either the standard large sample asymptotics or weak-instrument asymptoics, the limiting distribution of AR statistic under the null hypothesis is $\chi^2(K_2)$

- **Score-type Test**

  Define the statistics

$$\boldsymbol{S} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0(\boldsymbol{b}_0{}'\boldsymbol{\Omega}\boldsymbol{b}_0)^{-1/2} \tag{2.10}$$

  and

$$\boldsymbol{T} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix}\left[(\boldsymbol{\beta}_0, \boldsymbol{I}_{G_1})\boldsymbol{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix}\right]^{-1/2}, \tag{2.11}$$

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 $\boldsymbol{b}_0 = (1, -\boldsymbol{\beta}_0{}')'$. Kliebergen (2002) proposed the statistic

$$K = \hat{\boldsymbol{S}}'\hat{\boldsymbol{T}}(\hat{\boldsymbol{T}}'\hat{\boldsymbol{T}})^{-1}\hat{\boldsymbol{T}}'\hat{\boldsymbol{S}}. \tag{2.12}$$

Kleibergen showed that under either the standard large sample asymptotics or weak-instrument asymptoics, the limiting distribution of 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

$$LR = \frac{\boldsymbol{b}_0'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0}{\boldsymbol{b}_0'\boldsymbol{\Omega}\boldsymbol{b}_0} - \min_{\boldsymbol{b}} \frac{\boldsymbol{b}'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{\Omega}\boldsymbol{b}}. \tag{2.13}$$

  Moreira (2003) showed that the LR statistic is a function of $\boldsymbol{S}$ and $\boldsymbol{T}$ defined in (2.10) and (2.11), 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 instrumens). Moreira (2003) and Andrews, Moreira, and Stock (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 the general conditions (stochastic instruments and nonormal disturbances.)

3 Asymptotic Expansion of the distribution of LR statistic under H_0

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

We consider testing a hypothesis that the coefficients of the endogenous variables are zero ($H_0 : \boldsymbol{\beta} = 0$). The likelihood ratio test statistic for this hypothesis can be defined as

$$l = (n - K)[\lambda_0 - \hat{\lambda}], \tag{3.14}$$

where

$$\lambda_0 = \frac{\boldsymbol{b}_0'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0}{\boldsymbol{b}_0'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\boldsymbol{b}_0}, \tag{3.15}$$

$$\hat{\lambda} = \min_{\boldsymbol{b}} \frac{\boldsymbol{b}'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\boldsymbol{b}}, \tag{3.16}$$

where $\boldsymbol{b}' = (1, -\boldsymbol{\beta}')$, $\boldsymbol{b}_0' = (1, -\boldsymbol{\beta}_0')$, $\boldsymbol{P}_F$ is $\boldsymbol{F}(\boldsymbol{F}'\boldsymbol{F})^{-1}\boldsymbol{F}'$, and $\bar{\boldsymbol{P}}_F = \boldsymbol{I} - \boldsymbol{P}_F$ for any full column matrix $\boldsymbol{F}$.

We consider a modification of the likelihood ratio test based on an asymptotic expansion of the distribution of the LR statistic under H_0. The following notations are used throughout this chapter:

$$\boldsymbol{q}_2' = \frac{1}{\sigma^2}(\boldsymbol{\omega}_{12} - \beta'\boldsymbol{\Omega}_{22}, \boldsymbol{0}) : \quad 1 \times p, \tag{3.17}$$

$$\boldsymbol{C}_1 = \boldsymbol{q}_2\boldsymbol{q}_2' : \quad p \times p, \tag{3.18}$$

$$\boldsymbol{C}_2 = \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{C}_1 : \quad p \times p, \tag{3.19}$$

$$\boldsymbol{X} = \boldsymbol{Z}\begin{pmatrix} \boldsymbol{\Pi}_{12} & \boldsymbol{I}_{K_1} \\ \boldsymbol{\Pi}_{22} & \boldsymbol{0} \end{pmatrix} : \quad n \times p, \tag{3.20}$$

and

$$\tilde{\boldsymbol{Q}} = \boldsymbol{X}'\boldsymbol{X} : \quad p \times p. \tag{3.21}$$

We give the large sample asymptotic expansion of the distribution of the LR statistic (3.14) under H_0 in the case of the normal disturbances, which is similar to *Theorem 1* of Morimune and Tsukuda (1984).

Theorem 1 *Assume there exists a constant positive definite matrix* $\boldsymbol{Q} = plim_{n\to\infty} n^{-1}\tilde{\boldsymbol{Q}}$ *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:*

$$\begin{aligned} P(l \le \xi) &= G_{G_1}(\xi) \\ &\quad -\frac{\xi}{n}\{\frac{1}{G_1}\sigma^2 tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_2)L - \frac{1}{2}[G_1 - 2 - \xi]\} g_{G_1}(\xi) + O(n^{-3/2}), \end{aligned} \tag{3.22}$$

where G_{G_1} *and* g_{G_1} *are the* χ^2 *distribution function and* χ^2 *density function with* G_1 *degrees of freedom, respectively.*

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

$$u_\alpha + \frac{u_\alpha}{n}\{\frac{1}{G_1} tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_2)\sigma^2 L - \frac{1}{2}(G_1 - 2 - u_\alpha)\}, \tag{3.23}$$

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

$$\hat{\boldsymbol{Q}}^{-1} = n \begin{pmatrix} \boldsymbol{Y}_2'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}_2 - \hat{\lambda}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2 & \boldsymbol{Y}_2'\boldsymbol{Z}_1 \\ \boldsymbol{Z}_1'\boldsymbol{Y}_2 & \boldsymbol{Z}_1'\boldsymbol{Z}_1 \end{pmatrix}^{-1} \tag{3.24}$$

and

$$\hat{\boldsymbol{C}}_2 = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2/q - \frac{1}{\hat{\sigma}^4}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2/q^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \tag{3.25}$$

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 $\hat{\boldsymbol{b}} = (1, -\hat{\boldsymbol{\beta}}')'$. We propose a modified LR test (LR_{m1}) using the critical value

$$u_\alpha + \frac{u_\alpha}{n}\{\frac{1}{G_1} tr(\hat{\boldsymbol{Q}}^{-1}\hat{\boldsymbol{C}}_2)\hat{\sigma}^2 L - \frac{1}{2}(G_1 - 2 - u_\alpha)\}, \tag{3.26}$$

instead of u_α.

4 Large-K_n Asymptotic Approximation of the distribution of the LR statistic under H_0

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

$$\begin{aligned} n &\to \infty, \\ K/n &= c_1 + O(n^{-1}), \qquad (0 \le c_1 < 1) \\ K/q &= c_2 + O(n^{-1}), \qquad (0 \le c_2 < \infty) \end{aligned} \tag{4.27}$$

where we defined $q = n - K$.

Under the sequences (4.27), 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 $\boldsymbol{Q} = plim_{n\to\infty} n^{-1}\tilde{\boldsymbol{Q}}$ such that $\boldsymbol{Q} = n^{-1}\tilde{\boldsymbol{Q}} + O_p(n^{-1})$. Then, under H_0, under the sequences (4.27),*

$$l \xrightarrow{d} \frac{1}{\sigma^2}\boldsymbol{UQU}, \tag{4.28}$$

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

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

The limit distribution may 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 $\boldsymbol{Q\Psi}/\sigma^2$.

Here we have used the notations that $\boldsymbol{\Xi}_3 = plim_{n\to\infty}\boldsymbol{D}_2'\frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i[(1+c_2)a_{ii}^{(n)} - c_2]E[u_i^2\boldsymbol{w}_{2i}']$, $\eta = (1+c_2)^2 plim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n a_{ii}^{(n)2} - c_2^2$, $a_{ii}^{(n)} = \boldsymbol{z}_i'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_i$, $\boldsymbol{\Gamma}_4 = E(u_i^2\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}') - \sigma^2E[\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}']$, and $\boldsymbol{w}_{2i} = (\mathbf{v}_{2i}'\ \mathbf{0}')' - u_i\mathbf{q}_2$.

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 (3.24) and

$$\begin{aligned}\hat{\boldsymbol{\Psi}} &= \hat{\sigma}^2 \hat{\boldsymbol{Q}}^{-1} \\ &+ \frac{K}{n}(1+\hat{\lambda})\hat{\boldsymbol{Q}}^{-1} \begin{pmatrix} \frac{1}{q}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2\hat{\sigma}^2 - \frac{1}{q^2}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \hat{\boldsymbol{Q}}^{-1},\end{aligned} \tag{4.29}$$

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

In the case of non-normality, $\boldsymbol{\Psi}$ 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 are deviated 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 using the critical value based on the approximation by large-K_n asymptotics, LR_{largeK}.

5 Size Comparison with the CLR statistic

5.1 The Case of Normal Disturbances

We conduct the 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 $\hat{\boldsymbol{\Omega}}$;

$$\delta^2 = \frac{\boldsymbol{\Pi}_{22}'\boldsymbol{A}_{22.1}\boldsymbol{\Pi}_{22}}{\omega_{22}}, \tag{5.30}$$

the noncentrality parameter associated with (2.1); and

$$\alpha = \frac{\omega_{22}\beta - \omega_{21}}{|\boldsymbol{\Omega}|^{1/2}} = -\frac{\rho}{(1-\rho^2)^{1/2}}, \tag{5.31}$$

where ρ is a correlation 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 use the DGP

$$\boldsymbol{y}_1 = \boldsymbol{y}_2\beta^{(0)} + \boldsymbol{Z}_1\gamma^{(0)} + \boldsymbol{u}, \tag{5.32}$$

and

$$\boldsymbol{y}_2 = \boldsymbol{Z}\boldsymbol{\Pi}_2^{(0)} + \boldsymbol{V}_2, \tag{5.33}$$

where $K_1 = 1$, $\boldsymbol{Z} \sim N(\boldsymbol{0}, I_K \otimes I_n)$, $(\boldsymbol{u}, \boldsymbol{V}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes I_n)$, $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, 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 $\boldsymbol{\Pi}_2^{(0)} = c(1,\cdots,1)'$.

Tables 1-4 contain the empirical sizes of the statistics at the 10, 5, and 1% for various values of δ^2, K_2, and α. The number of repetitions is 10,000 in each experiments. 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.

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 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 robust to weak instruments and has good power properties, this finding seems to have some importance. When

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

the number of the instruments is small (less than five), The LR_{m1} test and CLR test improve the size properties. However, as the number of the instruments increases the LR_{m1} test as well as the CLR become 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% ans 1% asymptotic critical values of LR, LR_{m1}, CLR, t_{largeK} and LR_{largeK} in Tables 5-6. We calculated the critical values of the LR_{m1}, t_{largeK} and LR_{largeK} using the asymptotic variance assuming normal disturbances. From these experiments, the size properties of all these statistics, which are derived under the assumption of normal disturbances, are approximately valid even if the distributions of disturbances are deviated from normal.

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

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

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 be also 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, Morimune, and Tsukuda (1983) and Morimune and Tsukuda (1984). The variance ratio $\hat{\lambda}$ defined by (3.16) is stochastically expanded as

$$(n-K)\hat{\lambda} = \hat{\lambda}^{(0)} + \frac{1}{\sqrt{n}}\hat{\lambda}^{(1)} + \frac{1}{n}\hat{\lambda}^{(2)} + O_p(n^{-3/2}), \tag{A.34}$$

where

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

where

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

and

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

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 (3.15) is expanded as

$$(n-K)\lambda_0 = \lambda_0^{(0)} + \frac{1}{\sqrt{n}}\lambda_0^{(1)} + \frac{1}{n}\lambda_0^{(2)} + O_p(n^{-3/2}), \tag{A.35}$$

where

$$\begin{aligned}
\lambda_0^{(0)} &= \boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}/\sigma^2, \\
\lambda_0^{(1)} &= -\frac{1}{\sigma^2}[\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}x], \\
&\text{and} \\
\lambda_0^{(2)} &= \frac{1}{\sigma^2}[\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}\{-\frac{1}{3}(x^2-2)+x^2\}].
\end{aligned}$$

Hence the test statistic is stochastically expanded as

$$l = l^{(0)} + \frac{1}{\sqrt{n}}l^{(1)} + \frac{1}{n}l^{(2)} + O_p(n^{-3/2}), \tag{A.36}$$

where

$$l^{(0)} \equiv v = \lambda_0^{(0)} - \hat{\lambda}^{(0)} = \frac{1}{\sigma^2}\boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u}, \tag{A.37}$$

$$\begin{aligned}
l^{(1)} &= \lambda_0^{(1)} - \hat{\lambda}^{(1)} \qquad \text{(A.38)} \\
&= \frac{1}{\sigma^2}\{2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} - \boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u}x - 2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}\boldsymbol{q}'\boldsymbol{e}^{(0)}\},
\end{aligned}$$

and

$$\begin{aligned}
l^{(2)} &= \lambda_0^{(2)} - \hat{\lambda}^{(2)} \qquad \text{(A.39)} \\
&= \frac{1}{\sigma^2}\{-[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}]'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})[(\boldsymbol{V}_2, \boldsymbol{0})\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{X}\boldsymbol{e}^{(1)}] \\
&\quad +2\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)(\boldsymbol{V}_2, \boldsymbol{0})[\boldsymbol{e}^{(1)} - (x - 2\boldsymbol{q}'\boldsymbol{e}^{(0)})\boldsymbol{e}^{(0)}] \\
&\quad -\frac{1}{\sigma^2}\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_X)\boldsymbol{u}[2(\boldsymbol{w}_{12} - \boldsymbol{\beta}'\boldsymbol{W}_{22}, \boldsymbol{0})\boldsymbol{e}^{(0)} \\
&\qquad -4\sigma^2\boldsymbol{q}'\boldsymbol{e}^{(0)}x + 2\sigma^2\boldsymbol{q}'\boldsymbol{e}^{(1)} - \sigma^2\boldsymbol{e}^{(0)'}(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{e}^{(0)}] \\
&\quad +\boldsymbol{u}'(\boldsymbol{P}_X - \boldsymbol{P}_{Z_1})\boldsymbol{u}[-\frac{1}{3}(x^2-2)+x^2]\}.
\end{aligned}$$

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

$$\begin{aligned}
C(t) &= E(exp(itv)) + \frac{1}{\sqrt{n}}E(itE(l^{(1)}|v)exp(itv)) \qquad \text{(A.40)} \\
&+ \frac{1}{n}E(itE(l^{(2)}|v)exp(itv)) + \frac{1}{2n}E(i^2tE({l^{(1)}}^2|v)exp(itv)) + O(-n^{-3/2}).
\end{aligned}$$

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):

$$E(\boldsymbol{e}^{(0)\prime}\boldsymbol{C}_j\boldsymbol{e}^{(0)}|v) = \frac{v}{G_1}\sigma^2 tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_j), \quad j = 1, 2, \tag{A.41}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are defined by (3.18) and (3.19) respectively.

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

$$\begin{aligned} E(l^{(1)}|v) &= 0, & \text{(A.42)}\\ E(l^{(2)}|v) &= 2v + tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_2\sigma^2)L, & \text{(A.43)}\\ E({l^{(1)}}^2|v) &= 4tr(\boldsymbol{Q}^{-1}\boldsymbol{C}_2\sigma^2)L + 2v^2. & \text{(A.44)} \end{aligned}$$

The probablity $P(l \leq \xi)$ is approximated to the order n^{-1} by the Fourier inverse transformation of the characteristic function (A.40). 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):

$$\begin{aligned} &\int_{x=0}^{\xi} \frac{1}{2\pi} \int_t (-it)^p exp(-itx) E[exp(itv)v^j] dtdx \\ &\qquad = \frac{2^j \Gamma(\frac{G_1}{2} + j)}{\Gamma(\frac{G_1}{2})} \cdot g_{G_1+2j}^{(p-1)}(\xi), \end{aligned} \tag{A.45}$$

where $i = \sqrt{-1}$, j is any integer ($G_1 + 2j > 0$), and $g_{G_1+2j}^{(p-1)}(\xi)$ 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. (Q.E.D.)

B Derivation of Theorem 2

The variance ratio (3.16) is exactly rewritten as

$$\hat{\lambda} = \frac{\{\boldsymbol{u} - \frac{1}{\sqrt{n}}[\boldsymbol{Z}\boldsymbol{D}_2 + (\boldsymbol{V}_2, \boldsymbol{0})]\hat{\boldsymbol{e}}\}'\boldsymbol{P}_Z\{\boldsymbol{u} - \frac{1}{\sqrt{n}}[\boldsymbol{Z}\boldsymbol{D}_2 + (\boldsymbol{V}_2, \boldsymbol{0})]\hat{\boldsymbol{e}}\}}{\{\boldsymbol{u} - \frac{1}{\sqrt{n}}(\boldsymbol{V}_2, \boldsymbol{0})\hat{\boldsymbol{e}}\}'\bar{\boldsymbol{P}}_Z\{\boldsymbol{u} - \frac{1}{\sqrt{n}}(\boldsymbol{V}_2, \boldsymbol{0})\hat{\boldsymbol{e}}\}} \tag{B.46}$$

where

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

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

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

$$\hat{\boldsymbol{e}} = \boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{e}^{(1)} + O_p(n^{-1}). \tag{B.47}$$

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

$$\boldsymbol{e}^{(0)} = \boldsymbol{Q}^{-1}[\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{u} + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u} - \frac{\sqrt{c_1c_2}}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}], \tag{B.48}$$

$$\begin{aligned} \boldsymbol{e}^{(1)} = & -\boldsymbol{Q}^{-1}[\{\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'(\boldsymbol{V}_2 \ \boldsymbol{0}) + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z(\boldsymbol{V}_2 \ \boldsymbol{0}) \\ & -\sqrt{c_1c_2}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2 \ \boldsymbol{0})\}\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{W}_2'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\ & -\frac{c_1}{c_2}\lambda^{(1)}\left[\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^2\right]\boldsymbol{e}^{(0)} + \sqrt{\frac{c_1}{c_2}}\lambda^{(1)}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}]. \end{aligned} \tag{B.49}$$

We first make the large-K_n stochastic expansion of the variance ratio (3.16). Substituting (B.47) into (B.46), the numerator of the variance ratio devided by K becomes

$$\begin{aligned} \sigma^2 & + \frac{1}{\sqrt{n}}\{\sqrt{\frac{1}{c_1}}\sqrt{K}(\frac{1}{K}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u} - \sigma^2) - 2(\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(0)}\} \\ & + \frac{1}{n}\{-2\sqrt{\frac{1}{c_1}}\sqrt{K}[\frac{1}{K}\boldsymbol{b}_0'\boldsymbol{V}'\boldsymbol{P}_Z(\boldsymbol{V}_2, \boldsymbol{0}) - (\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2]\boldsymbol{e}^{(0)} \\ & \quad -2\frac{1}{c_1}\frac{1}{\sqrt{n}}\boldsymbol{u}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} + \frac{1}{c_1}\boldsymbol{e}^{(0)\prime}\frac{1}{n}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\ & \quad +\boldsymbol{e}^{(0)\prime}\begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{e}^{(0)} - 2(\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(1)}\} \end{aligned} \tag{B.50}$$

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

$$\begin{aligned} \sigma^2 & + \frac{1}{\sqrt{n}}\{\sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2) - 2(\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(0)}\} \\ & + \frac{1}{n}\{-2\sqrt{\frac{c_2}{c_1}}\sqrt{q}[\frac{1}{q}\boldsymbol{b}_0'\boldsymbol{V}'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2, \boldsymbol{0}) - (\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2]\boldsymbol{e}^{(0)} \\ & \quad -2\frac{1}{c_1}\frac{1}{\sqrt{n}}\boldsymbol{u}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} + \frac{1}{c_1}\boldsymbol{e}^{(0)\prime}\frac{1}{n}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\ & \quad +\boldsymbol{e}^{(0)\prime}\begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{e}^{(0)} - 2(\boldsymbol{b}_0'\boldsymbol{\Omega}, \boldsymbol{0})\boldsymbol{J}_2\boldsymbol{e}^{(1)}\} \end{aligned} \tag{B.51}$$

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

Multiplying Taylor's expansion of the inverse of (B.51) to (B.50) it follows the large-K_n stochastic expansion of the variance ratio (3.16):

$$\hat{\lambda} = \hat{\lambda}^{(0)} + \frac{1}{\sqrt{n}}\hat{\lambda}^{(1)} + \frac{1}{n}\hat{\lambda}^{(2)} + O_p(n^{-3/2}), \tag{B.52}$$

where

$$\begin{aligned}
\hat{\lambda}^{(0)} &= c_2, \\
\hat{\lambda}^{(1)} &= \frac{c_2}{\sigma^2}\{\frac{1}{\sqrt{c_1}}(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}) - \sqrt{\frac{c_2}{c_1}}(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u})\}, \\
\hat{\lambda}^{(2)} &= \frac{c_2}{\sigma^2}\{-\frac{1}{c_1}\boldsymbol{e}^{(0)\prime}\boldsymbol{Q}\boldsymbol{e}^{(0)} \\
&\quad -\sqrt{\frac{c_2}{c_1}}\frac{1}{\sigma^2}[\frac{1}{\sqrt{c_1}}(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}) - \sqrt{\frac{c_2}{c_1}}(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u})][\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2)]\}
\end{aligned}$$

Similarly λ_0 defined by (3.15) is expanded as

$$\lambda_0 = \lambda_0^{(0)} + \frac{1}{\sqrt{n}}\lambda_0^{(1)} + \frac{1}{n}\lambda_0^{(2)} + O_p(n^{-3/2}), \tag{B.53}$$

where

$$\begin{aligned}
\lambda_0^{(0)} &= c_2, \\
\lambda_0^{(1)} &= \frac{c_2}{\sigma^2}\{\frac{1}{\sqrt{c_1}}(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}) - \sqrt{\frac{c_2}{c_1}}(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\}, \\
\lambda_0^{(2)} &= -\frac{c_2}{\sigma^4}\sqrt{\frac{c_2}{c_1}}[\frac{1}{\sqrt{c_1}}(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}) - \sqrt{\frac{c_2}{c_1}}(\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u})][\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2)].
\end{aligned}$$

Hence we have the relation that

$$l = \frac{n-K}{n}(\lambda_0^{(2)} - \hat{\lambda}^{(2)}) = \frac{1}{\sigma^2}\boldsymbol{e}^{(0)\prime}\boldsymbol{Q}\boldsymbol{e}^{(0)} + o_p(1). \tag{B.54}$$

Anderson, Kunitomo and Matsushita (2006) show that

$$\boldsymbol{e}^{(0)} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Psi}). \tag{B.55}$$

Then we have the desired result.

References

[1] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, John-Wiley, 3rd Edition.

[2] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[3] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[4] Anderson, T.W., N. Kunitomo, and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol. 50, 1009-1027.

[5] Anderson, T.W., N. Kunitomo, and Y. Matsushita (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)

[6] Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Ecoonometrica*, Vol.74, 715-752.

[7] Angrist, J. D. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, Vol. 106, 979-1014.

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

[9] Bound, J., D.A. Jaeger, and R.M. Baker (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*, Vol. 90, 443-450.

[10] Fujikoshi, Y., K. Morimune, N. Kunitomo, and M. Taniguchi (1982), "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System," *Journal of Econometrics*, Vol. 18, 2, 191-205.

[11] Guggenberger, P. and R.J. Smith (2005), "Generalized Empirical Likelihood Tests under Partial, Weak and Strong Identification," *Econometric Theory*, Vol. 21, forthcoming.

[12] Hansen, C., J. Hausman, and W.K.Newey (2005), "Estimation with Many Instrumental Variables," Unpublished Manuscript.

[13] Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, Vol.70-5, 1781-1803.

[14] Kunitomo, N. (1980), "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, Vol. 75, 693-700.

[15] Kunitomo, N. and Y. Matsushita (2003a), "Finite Sample Distributions of the Empirical Likelihood Estimator and GMM Estimator," Discussion Paper CIRJE-F-200, Graduate School of Economics, University of Tokyo. (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2003/2003cf200.pdf)

[16] Kunitomo,N. and Matsushita, Y. (2005), "Asymptotic Expansions of the Distribuitons of Semi-Parametric Estimators in A Linear Simultaneous Equations System," Preprint.

[17] Matsushita, Y. (2006), "t-Tests in A Structural Equation with Many Instruments," Preprint.

[18] Moreira, M. (2003), "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, Vol. 71-4, 1027-1048.

[19] Morimune, K. (1983), "Approximate Distribiutions of k-class Estimators when the Degree of Overidentification is Large Compared With Sample Size," *Econometrica*, Vol.51-3, 821-841.

[20] Morimune, K and Y. Tsukuda (1984), "Testing a Subset of Coefficients in a Structural Equation," *Econometrica*, Vol.52, 427-448.

[21] Staiger, D. and J. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, Vol. 65, 557-586.