CIRJE-F-467

t-Tests in a Structural Equation with Many Instruments

Yukitoshi Matsushita
CIRJE, Faculty of Economics, University of Tokyo

February 2007

t-Tests in a Structural Equation with Many Instruments

Yukitoshi Matsushita *

December 2006

Abstract

This paper studies the properties of t-ratios associated with the limited information maximum likelihood (LIML) estimators in a structural form estimation when the number of instrumental variables is large. Asymptotic expansions are made of the distributions of a large K t-ratio statistic under large-K_n asymptotics. A modified t-ratio statistic is proposed from the asymptotic expansion. The power of the large K t-ratio test dominates the AR test, the K-test by Kleibergen (2002), and the conditional LR test by Moreira (2003); and the difference can be substantial when the instruments are weak.

Key Words

Many instruments, Asymptotic expansions, t-ratio, Limited Information Maximum Likelihood(LIML), Linear Simultaneous Equations System

*I am grateful to Yoichi Arai and Naoto Kunitomo for their support and suggestions, and to T.W. Anderson for giving useful comments to the previous version.

1. Introduction

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), who used 178 instruments in one of their specifications. However, in such cases, it has been found that approximate distributions of estimators and statistics based on the conventional large sample asymptotic theory can be inaccurate. See, for instance, Anderson, Sawa, and Kunitomo (1982); Bound, Jaeger, and Baker (1995); and Anderson, Kunitomo, and Matsushita (2005). In order to overcome this problem, several new test statistics have recently been proposed. 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 the weak instruments. Several papers extend these tests to a more general framework including heteroscedasticity. See, for instance, Kleibergen (2005) and Andrews, Moreira and Stock (2006).

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 (2005). 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.

This paper focuses on the second approach. The main purpose of this paper is to explore the finite sample properties of t-ratio statistics under the large-K_n asymptotic theory. Since the t-test is one of the most commonly used procedures to test hypotheses on a coefficient in a structural equation, there have been several literatures investigating the finite sample properties of the t-ratio. See Richardson and Rohr (1971), Morimune (1989), Hansen et al (2006), for instance. Morimune (1989) derived asymptotic expansions of the distributions of (standard) t-ratio statistics associated with the k-class estimators under the standard large sample asymptotic theory in the case of normal disturbances. This paper extends his work into the case with many instruments. We derive an asymptotic expansion of the null distribution of (large K) t-ratio statistic based on the LIML estimator under the large-K_n asymptotics: both in the case of normal disturbances and non-normal disturbances. An asymptotic expansion of the distribution of the LIML estimator is also derived, which is new in the many endogenous variables case. We find that the absolute values of the second terms of the asymptotic expansion of the (standardized) LIML estimator and large K t-ratio are the same but have different signs, and that this second order term may have a substantial impact on the size distortion of the t-ratio test. Using the asymptotic expansion of the large K t-ratio, a modified t-ratio statistic which does not include terms of order $O(n^{-1/2})$ in the expansion is proposed.

In Section 2, the model and t-tests with many instruments are explained, and a large K t-test is defined. In Section 3, large-K_n asymptotic expansions of the null distributions of the t-ratio statistic are provided both in the cases of normal and non-normal disturbances. Some Monte Carlo experiments are provided in Section 4, and conclusions are provided in Section 5. All derivations of theorems are provided in Appendices.

2. The Model and t-Tests with Many Instruments

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

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

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 normally 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

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

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

$$\begin{pmatrix} \boldsymbol{Y}_2' \\ \boldsymbol{Z}_1' \end{pmatrix} \boldsymbol{P}_Z(\boldsymbol{y}_1 \ \boldsymbol{Y}_2 \ \boldsymbol{Z}_1) \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{LI} \\ -\hat{\gamma}_{LI} \end{pmatrix} = \mathbf{0}. \tag{2.8}$$

Under the conventional (fixed K) asymptotics, both LIML and TSLS estimators are consistent and have the same asymptotic distributions. Let $\boldsymbol{i}$ be a $(G_1 + K_1) \times 1$ column vector of zeros, apart from its ith element which is unity. The standard t-ratio for testing

$$H_0 : \boldsymbol{i}' \begin{pmatrix} \boldsymbol{\beta} \\ \gamma \end{pmatrix} = 0, \tag{2.9}$$

is

$$t_k = \frac{1}{s_k \sqrt{\hat{\boldsymbol{Q}}_{ii}^{-1}(k)}} \boldsymbol{i}' \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}}_k \\ \hat{\gamma}_k \end{pmatrix}, \quad k = LIML, \ TSLS, \tag{2.10}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ can be the LIML and TSLS estimators, s_k^2 is an esimator of σ^2 that is given as

$$s_k^2 = \frac{1}{n - K_1 - G_1} (\boldsymbol{y}_1 - \boldsymbol{Y}_2 \hat{\boldsymbol{\beta}}_k - \boldsymbol{Z}_1 \hat{\gamma}_k)' (\boldsymbol{y}_1 - \boldsymbol{Y}_2 \hat{\boldsymbol{\beta}}_k - \boldsymbol{Z}_1 \hat{\gamma}_k), \tag{2.11}$$

and $\hat{\boldsymbol{Q}}_{ii}^{-1}(k)$ is the ith diagonal element in the matrix, where

$$\hat{\boldsymbol{Q}}^{-1} = n \begin{pmatrix} \boldsymbol{Y}_2' \boldsymbol{Z} (\boldsymbol{Z}' \boldsymbol{Z})^{-1} \boldsymbol{Z}' \boldsymbol{Y}_2 - (k-1) \boldsymbol{Y}_2' \bar{\boldsymbol{P}}_Z \boldsymbol{Y}_2 & \boldsymbol{Y}_2' \boldsymbol{Z}_1 \\ \boldsymbol{Z}_1' \boldsymbol{Y}_2 & \boldsymbol{Z}_1' \boldsymbol{Z}_1 \end{pmatrix}^{-1}. \tag{2.12}$$

Here, $k = 1$ for the TSLS estimator, and $k = 1 + \hat{\lambda}$ for the LIML estimator, and $(s^2 \hat{\boldsymbol{Q}}_{ii}^{-1})$ is a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}' \ \hat{\gamma}')\boldsymbol{i}$ under the null hypothesis of the test.

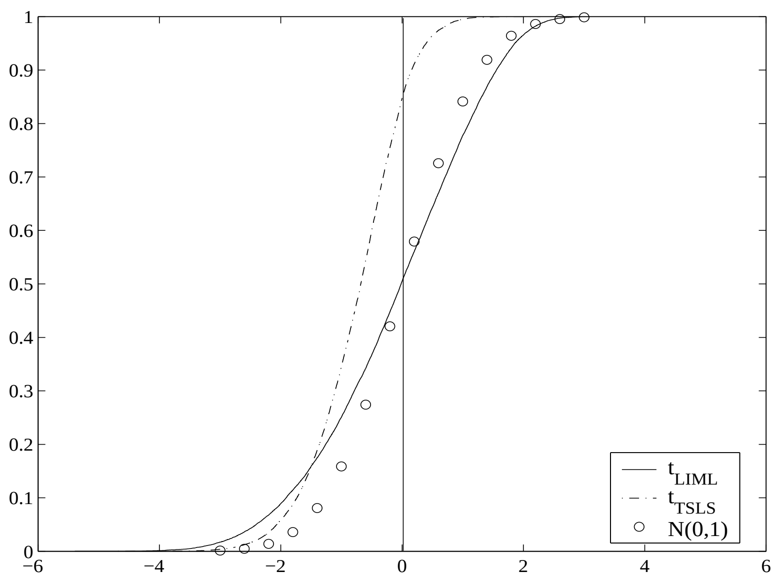

Figure 1: The null distributions of t_{LIML} and t_{TSLS} when the number of excluded instruments is 30 ($n - K = 100, K_2 = 30, \alpha = 0.5, \delta^2 = 30$)

However, when the number of instruments is large classical normal asymptotic approximations may provide poor approximations to the finite-sample distributions of IV statistics. When the number of the instruments is large, the TSLS estimator can be extremely biased. Anderson et al (2005) show that for $K_2 = 10$ and $K_2 = 30$, the median of the TSLS (and GMM) estimators can be lower than -1.0 ASD(asymptotic standard deviation)'s. On the other hand, the LIML estimator has larger variances than the asymptotic variance based on the standard large sample theory. Figure 1 includes the empirical null distributions of the (standard) t-ratio associated with the LIML and TSLS estimators when the number of the excluded instruments is 30.

Bekker (1994) pointed out that the large-K_n asymptotic thoery may be suited better to applications, where the number of the (excluded) instruments (K_2) is allowed to increase with the number of observations (n). In this paper we consider the same situations, that is, the number of the (excluded) instruments (K_2) is allowed to increase proportionally with the number of observations (n):

$$n \rightarrow \infty,$$

$$
\begin{aligned}
K/n &= c_1 + O(n^{-1}), \qquad (0 \le c_1 < 1) \\
K/q &= c_2 + O(n^{-1}), \qquad (0 \le c_2 < \infty)
\end{aligned} \tag{2.13}
$$

where we defined $q = n - K$. Under the large-K_n asymptotics, the asymptotic distributions of the LIML and TSLS estimators are rather different. The LIML estimator is consistent and asymptotic normal while the TSLS estimators even lose consistency. The LIML estimator attains the asymptotic lower bound when the number of instruments is large. See Kunitomo(1982) and Anderson and Kunitomo (2006). For this reason, in this paper, we focus on t-tests based on the LIML estimator. We define $\boldsymbol{D}_2$ as follows:

$$
\boldsymbol{D}_2 = \begin{pmatrix} \boldsymbol{\Pi}_{12} & \boldsymbol{I}_{K_1} \\ \boldsymbol{\Pi}_{22} & \boldsymbol{0} \end{pmatrix} \tag{2.14}
$$

and assume that

$$
\frac{1}{n}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D}_2 \xrightarrow{p} \boldsymbol{Q}, \tag{2.15}
$$

where $\boldsymbol{Q}$ is a nonsingular constant matrix. Under the sequence (2.13), Anderson and Kunitomo (2006) have proved under certain conditions [1]

$$
\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\gamma}}_{LI} - \boldsymbol{\gamma} \end{pmatrix} \xrightarrow{d} N(\boldsymbol{0}\ \boldsymbol{\Psi}),
$$

where

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

which is identical to the Bekker (1994) variance in the case of the normal disturbances. Here we have used the notations that

$$
\begin{aligned}
\boldsymbol{\Xi}_3 &= plim_{n\to\infty}\boldsymbol{D}_2'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i[(1+c_2)a_{ii}^{(n)} - c_2]E[u_i^2\boldsymbol{w}_{2i}'], \\
\boldsymbol{\Gamma}_4 &= E(u_i^2\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}') - \sigma^2E[\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}'], \\
\eta &= (1+c_2)^2 plim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}a_{ii}^{(n)2} - c_2^2, \\
\boldsymbol{q}_2 &= \frac{1}{\sigma^2}(\boldsymbol{\omega}_{21}' - \boldsymbol{\beta}'\boldsymbol{\Omega}_{22}\ \boldsymbol{0}')', \quad \boldsymbol{w}_{2i} = (\mathbf{v}_{2i}'\ \mathbf{0}')' - u_i\mathbf{q}_2,
\end{aligned}
$$

[1] Anderson and Kunitomo (2006) provided the results only on $\hat{\boldsymbol{\beta}}$ using different notations.

and $a_{ii}^{(n)} = \boldsymbol{z}_i'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_i$. $\boldsymbol{\Psi}$ can be estimated consistently by

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

where

$$\begin{aligned}\hat{\boldsymbol{\Omega}} &= \frac{1}{q}\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}, \quad \hat{\sigma}^2 = \frac{1}{q}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}, \quad \hat{\boldsymbol{q}}_2 = \frac{1}{\hat{\sigma}^2}\frac{1}{q}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}, \\ \hat{\boldsymbol{\Xi}}_3 &= \frac{1}{n}\sum_{i=1}^{n}[\hat{\boldsymbol{D}}_2'\boldsymbol{z}_i(1+\hat{\lambda})a_{ii}^{(n)} - \hat{\lambda}]\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i{}^2\hat{\boldsymbol{w}}_{2i}, \quad \hat{\boldsymbol{D}}_2 = \begin{pmatrix} \hat{\boldsymbol{\Pi}}_{12} & \boldsymbol{I}_{K_1} \\ \hat{\boldsymbol{\Pi}}_{22} & \boldsymbol{0} \end{pmatrix}, \\ \hat{\boldsymbol{\Gamma}}_4 &= \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\hat{\boldsymbol{w}}_{2i}\hat{\boldsymbol{w}}_{2i}' - \hat{\sigma}^2\frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{w}}_{2i}\hat{\boldsymbol{w}}_{2i}', \quad \hat{\eta} = (1+\hat{\lambda})^2\frac{1}{n}\sum_{i=1}^{n}a_{ii}^{(n)2} - \hat{\lambda}^2, \\ \hat{u}_i &= y_{1i} - \boldsymbol{y}_{2i}'\hat{\boldsymbol{\beta}} - \boldsymbol{z}_{1i}'\hat{\boldsymbol{\gamma}}, \text{ and } \hat{\boldsymbol{w}}_{2i} = ((\boldsymbol{y}_{2i} - (\hat{\boldsymbol{\Pi}}_{12}'\hat{\boldsymbol{\Pi}}_{22}')\boldsymbol{z}_i)', \ 0)' - \hat{u}_i\hat{\boldsymbol{q}}_2.\end{aligned}$$

Here $\hat{\boldsymbol{\Pi}}$ is the OLS estimator of $\boldsymbol{\Pi}$ in (2.2), and we have used the notation that $\hat{\boldsymbol{b}} = (1, -\boldsymbol{\beta}_{LI}')'$. The large K t-ratio for testing H_0 is given by

$$t_{largeK} = \frac{1}{\sqrt{\hat{\boldsymbol{\Psi}}_{ii}}}\boldsymbol{i}'\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LI} \\ \hat{\boldsymbol{\gamma}}_{LI} \end{pmatrix}, \tag{2.16}$$

where $\hat{\boldsymbol{\Psi}}_{ii}$ is the i-th diagonal element in the matrix $\hat{\boldsymbol{\Psi}}$. The estimate of the asymptotic variance depends on the estimates of the third and fourth order moments of the distributions of the disturbances, which make it complicated. However, we will see that these terms seem to have little effects even when the distributions of the disturbances are deviated from the normal distribution in Section 4. See Anderson and Kunitomo (2006) for further discussions.

3. Asymptotic Expansions of the Distributions of the Large K t-Ratio Under H_0

3.1 The Case of Normal Disturbances

In order to explore the finite sample properties of the large K t-test, asymptotic expansions of the null distributions of the large K t-ratio are derived under the large-K_n asymptotics in this section.

We consider the case of normal disturbances first. When the rows of $\boldsymbol{V}$ are normally distributed, a consistent estimator of the asymptotic variance of the LIML estimator is given by

$$\begin{aligned}\hat{\boldsymbol{\Psi}} &= \hat{\sigma}^2\hat{\mathbf{Q}}^{-1} \\ &+\frac{K}{n}(1+\hat{\lambda})\hat{\mathbf{Q}}^{-1}\begin{pmatrix} \frac{1}{q}\mathbf{Y}_2'\bar{\mathbf{P}}_Z\mathbf{Y}_2\hat{\sigma}^2 - \frac{1}{q^2}\mathbf{Y}_2'\bar{\mathbf{P}}_Z\mathbf{Y}\hat{\mathbf{b}}\hat{\mathbf{b}}'\mathbf{Y}'\bar{\mathbf{P}}_Z\mathbf{Y}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix}\hat{\mathbf{Q}}^{-1}.\end{aligned} \tag{3.1}$$

A new assumption is necessary for the expansion.

Assumption 1 *There exists a constant positive definite matrix*

$$\boldsymbol{Q} = plim_{n\to\infty}\hat{\boldsymbol{Q}} \quad s.t. \;\; \hat{\boldsymbol{Q}} = \boldsymbol{Q} + O_p(n^{-1}). \tag{3.2}$$

The following theorem is obtained. The derivation is provided in *Appendices* A and B.

Theorem 1 *When the rows of $\boldsymbol{V}$ are normally distributed, the asymptotic expansion of the distribution of the large K t-ratio (2.16) under the sequence (2.13) is given by*

$$P\{t_{largeK} \le \xi\} = \Phi(\xi) - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\Psi_{ii}}}(\boldsymbol{i}'\boldsymbol{\Psi}\boldsymbol{q}_2)\xi^2\phi(\xi) + O(n^{-1}), \tag{3.3}$$

where $\boldsymbol{\xi}$ is a (G_1+K_1) vector and $\Phi(\xi)$ and $\phi(\xi)$ are the cdf and the density function of the standard normal distribution, respectively.

As in the case with the LIML estimator (see *Appendix* A), when $c_1 = 0$, this asymptotic expansion is identical to the result under the standard large sample theory up to $O(n^{-1/2})$. See Morimune (1989). When $G_1 = 1$, we have a simple expression of the expansion of the large K t-ratio for testing $H_0 : \beta = 0$ as follows.

Corollary 1 *When $G_1 = 1$, the asymptotic expansion of the distribution of the large K t-ratio (2.16) for testing $H_0 : \beta = 0$ under the sequence (2.13) is given by*

$$P(t_{largeK} \leq \xi) = \Phi(\xi) + \frac{\alpha}{\mu}\sqrt{\eta}\xi^2\phi(\xi) + o(\mu^{-1}). \tag{3.4}$$

Here, we use the notations $\eta = 1 + \frac{1}{\tau^2}(1 + \frac{\nu^2}{\tau^2})$, $\nu^2 = \lim_{n\to\infty} \frac{\mu^2}{q}$, $\tau^2 = \lim_{n\to\infty} \frac{\mu^2}{L}$, $\mu^2 = \frac{\sigma^2}{|\Omega|}\Pi_{22}' A_{22.1}\Pi_{22}$, $A_{22.1} = \boldsymbol{Z}_2'\boldsymbol{Z}_2 - \boldsymbol{Z}_2'\boldsymbol{Z}_1(\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1'\boldsymbol{Z}_2$, $\alpha = (\omega_{22}\beta - \omega_{21})/\sqrt{|\Omega|}$, and $L = K_2 - 1$.

From Corollary 1 and A.1 (in *Appendix A*), we find that the absolute values of the second terms of the asymptotic expansion of the LIML estimator and large K t-ratio are the same but have different signs. This implies that the distributions of the LIML estimator and large K t-ratio are skewed in opposite directions. We will later see that this second order term may have a substantial impact on the size distortion of the t-test.

We find from the asymptotic expansion (3.3) that there exists a simple adjustment of the t-ratio statistic which does not include terms of order $O(n^{-1/2})$ in the expansion. We propose an adjusted t-ratio as

$$t_{adj} = t_{largeK} - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\hat{\boldsymbol{\Psi}}_{ii}}}(\boldsymbol{i}'\hat{\boldsymbol{\Psi}}\hat{\boldsymbol{q}}_2)t_{largeK}^2, \tag{3.5}$$

where $\hat{\boldsymbol{\Psi}}$ is defined by (3.1), and $\hat{\boldsymbol{q}}_2$ is an estimator of $\boldsymbol{q}_2$ where $\boldsymbol{\omega}_{21}$, $\boldsymbol{\Omega}_{22}$, and $\boldsymbol{\beta}$ are consistently estimated by $\frac{1}{q}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{y}_1$, $\frac{1}{q}\boldsymbol{Y}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}_2$, and $\hat{\boldsymbol{\beta}}_{LI}$, respectively.

3.2 The Case of Non-normal Disturbances

In order to investigate the effects of the normality assumption for disturbances, the asymptotic expansion of the distribution of the large K t-ratio under H_0 is derived in the case of non-normal disturbances. One convenient class of underlying disturbances is the elliptically contoured distribution, which contains many important distributions including the multivariate normal distribution, the multivariate t distribution and the uniform distribution on the sphere in R^p.

When the rows of $\boldsymbol{V}$ are followed by the class of elliptically contoured distribution $EC(\boldsymbol{\Omega})$ [2], the asymptotic variance of the LIML estimator $\boldsymbol{\Psi}^{\dagger}$ is given by

$$\boldsymbol{\Psi}^{\dagger} = \sigma^2 \boldsymbol{Q}^{-1} + \{c_1(1+c_2)+\eta\kappa\}\boldsymbol{Q}^{-1}\left[\begin{pmatrix} \boldsymbol{\Omega}_{22}\sigma^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1}, \quad (3.6)$$

where $\kappa = (E(u_i^4)/\sigma^4 - 3)/3$. (Anderson and Kunitomo (2006))

Hence, the large K t-ratio for testing H_0 is given by

$$t^{\dagger}_{largeK} = \frac{1}{\sqrt{\hat{\boldsymbol{\Psi}}^{\dagger}_{ii}}}\boldsymbol{i}'\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{LI} \\ \hat{\boldsymbol{\gamma}}_{LI} \end{pmatrix}, \quad (3.7)$$

where $\hat{\boldsymbol{\Psi}}^{\dagger}$ is a consistent estimator of $\boldsymbol{\Psi}^{\dagger}$ using $\hat{\eta} = (1+\hat{\lambda})^2\frac{1}{n}\sum_{i=1}^{n} a_{ii}^{(n)2} - \hat{\lambda}^2$ and $\hat{\kappa} = (\frac{\frac{1}{n}\sum_{i=1}^{n}(y_{1i}-y_{2i}'\hat{\beta}-z_{1i}'\hat{\gamma})^4}{\hat{\sigma}^4} - 3)/3$ for estimating η and κ in (3.6), respectively.

We obtain the next result. The derivation is provided in *Appendix* C.

Theorem 2 *Let the rows of* $\boldsymbol{V}$ *be followed by the class of elliptically contoured distribution* $EC(\boldsymbol{\Omega})$. *In addition to Assumption 1, we assume that* $\frac{1}{n}\sum_{i=1}^{n} a_{ii}^{(n)} = plim\frac{1}{n}\sum_{i=1}^{n} a_{ii}^{(n)} + O_p(n^{-1})$. *The asymptotic expansion of the null distribution of the large K t-ratio (3.7) under the sequence (2.13) is given by*

$$P\{t^{\dagger}_{largeK} \le \xi\} = \Phi(\xi) - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\boldsymbol{\Psi}^{\dagger}}_{ii}}(\boldsymbol{i}'\boldsymbol{\Psi}^{\dagger}\boldsymbol{q}_2)\xi^2\phi(\xi) + O(n^{-1}). \quad (3.8)$$

From *Theorem 2*, when the disturbances are followed by the class of the elliptically contoured distribution, the asymptotic expansion of the distribution of the large K t-ratio under H_0 has the same form as that in the case of normal disturbances with $\boldsymbol{\Psi}$ replaced by $\boldsymbol{\Psi}^{\dagger}$.

4. Monte Carlo Experiments

Empirical distributions by Monte Carlo studies are obtained in order to examine the quality of the preceding asymptotic approximations to the finite sample distributions of the LIML estimator and t-ratio statistics. We considered models with

[2] The precise definition of elliptically contoured (EC) distribution has been given by Section 2.7 of Anderson (2003).

two endogenous variables, i.e., $G_1 = 1$. In this case, the distributions of the LIML estimator and the t-ratios for a coefficient on endogenous variables 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 $\hat{\mathbf{\Omega}}$;

$$\delta^2 = \frac{\mathbf{\Pi}_{22}' \boldsymbol{A}_{22.1} \mathbf{\Pi}_{22}}{\omega_{22}}, \tag{4.1}$$

the noncentrality parameter associated with (2.1); and

$$\alpha = \frac{\omega_{22}\beta - \omega_{21}}{|\mathbf{\Omega}|^{1/2}} = -\frac{\rho}{(1-\rho^2)^{1/2}}, \tag{4.2}$$

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

$$\boldsymbol{y}_1 = \boldsymbol{y}_2 \beta^{(0)} + \boldsymbol{Z}_1 \gamma^{(0)} + \boldsymbol{u}, \tag{4.3}$$

and

$$\boldsymbol{y}_2 = \boldsymbol{Z} \mathbf{\Pi}_2^{(0)} + \boldsymbol{V}_2, \tag{4.4}$$

where $K_1 = 1$, $\boldsymbol{Z} \sim N(\mathbf{0}, I_K \otimes I_n)$, $(\boldsymbol{u}, \boldsymbol{V}) \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes I_n)$, $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, 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 $\mathbf{\Pi}_2^{(0)} = c(1, \cdots, 1)'$. The number of repetitions in each experiment is 20,000. The accuracy of our simulation method has been carefully examined by Kunitomo and Matsushita (2003a).

4.1 Distributions of t-Ratios on β Under H_0

The empirical small sample distributions are calculated for a standardized form of the LIML estimator

$$\hat{e}_\beta = \frac{\sqrt{\mathbf{\Pi}_{22}' A_{22.1} \mathbf{\Pi}_{22}}}{\sigma}(\hat{\beta} - \beta),$$

and the three types of t-ratios– t_{LIML}, the large K t-ratio (t_{largeK}), and the adjusted large K t-ratio (t_{adj}) –under H_0. The null hypothesis H_0 was imposed so that the true coefficient is zero. In Tables 1-3, the 5, 10, 50, 90, and 95 percentiles, and the observed sizes at the 10% and 5% asymptotic critical values are tabulated. In Figure 2, graphs of $N(0,1)$, and empirical null distribuitions of t, t_{largeK}, t_{adj} are given in the case of $n - K = 100, K_2 = 30, \alpha = 1$, and $\delta^2 = 30$.

From the tables, the distribution of t_{LIML} is close to the standard normal distribution when α and K_2 are small. As α increases, a slight asymmetry is observed; and as K_2 increases, the tails become long, which causes a large difference between actual and nominal sizes. For given α, K_2, and n, the size distortion become small as δ^2 increases. For given α, δ^2, and n, the size distortion increases with K_2.

The distribution of the large K t-ratio (t_{largeK}) is closer to the standard normal distribution than t_{LIML} irrespective of whether K_2 is small or large. This implies that the large K asymptotics are more accurate than the standard large sample asymptotics, which agrees with the results reported in Bekker (1994). However, the distribution is still skewed when $|\alpha|$ is large. It is often the case in numerous applications that the sign of the parameters is known from the economic theory and that the one-sided test is used. In such cases, the size distortion of the large K t-test can be rather large.

The distribution of the adjusted large K t-ratio (t_{adj}) is the closest to the standard normal distribution in all the cases. It is close to being symmetric even when $|\alpha|$ and K_2 are large and the difference between the nominal and real sizes is small in all cases including the many weak instruments cases ($\delta^2/K_2 = 1$; see Moreira (2003)).

We note that the small sample distributions of the t-ratio statistics are rather different from that of the (standardized) LIML estimator. The distributions of t_{LIML} and t_{largeK} have long left tails for $\alpha > 0$ (and long right tails for $\alpha < 0$), while the

Table 1: Distributions of t ratios under H_0: $n - K = 30, K_2 = 3, \delta^2 = 30$

		$\alpha = 0.1$				$\alpha = 1$			
	normal	$\hat{e}_\beta$	t	t_{largeK}	t_{adj}	$\hat{e}_\beta$	t	t_{largeK}	t^{adj}
X05	-1.65	-1.64	-1.72	-1.61	-1.74	-1.37	-1.95	-1.92	-1.66
X10	-1.28	-1.26	-1.34	-1.26	-1.31	-1.10	-1.48	-1.45	-1.29
MEDN	0	0.01	0.01	0.01	0.01	-0.02	-0.01	-0.01	-0.01
X90	1.28	1.32	1.28	1.19	1.30	1.50	1.13	1.1	1.26
X95	1.65	1.75	1.60	1.50	1.69	2.09	1.39	1.37	1.63
$P(t < z_{05})$	5.0	4.9	5.8	4.7	5.9	1.9	8.0	7.6	5.2
$P(t < z_{10})$	10.0	9.7	11.1	9.6	10.4	6.3	13.2	12.7	10.2
$P(t > z_{90})$	10.0	10.7	9.9	8.2	10.3	12.9	6.8	6.3	9.6
$P(t > z_{95})$	5.0	6.0	4.4	3.4	5.4	8.5	2.3	2.0	4.8

distribution of the LIML estimator has a long right tail for $\alpha > 0$ (and a long left tail for $\alpha < 0$). The abovementioned observations agree with the asymptotic expansions of the cdfs of $\hat{e}^*_\beta$ and t_{largeK} in *Appendix* A and Section 3, respectively.

In summary, it may be stated that the distribution of t_{adj} is closest to the standard normal distribution in all the cases. The distribution of t_{LIML} is skewed and extremely deviated from the normal distribution, particularly when α and K_2 are large. The distribution of t_{largeK} is closer to the normal distribution than t_{LIML}; however, it is still skewed and the size distortion can be large. However, the difference between t_{adj} and the standard normal distribution is small in all cases except when δ^2 is too small. The actual size of t_{adj} is close to the nominal size.

4.2 Power Comparison

We conduct power comparisons of the large K t-ratio statistic with the Anderson-Rubin (AR) statistic (Anderson and Rubin (1949)), the K statistic (Kleibegen(2002)), and the conditional likelihood ratio (CLR) statistic (Moreira(2003)). [3]

[3] We do not report power results for the likelihood ratio (LR) test and the standard t-ratio test because their size properties appear to be rather poor in the situation considered here.

Table 2: Distributions of t ratios under H_0: $n - K = 100, K_2 = 30, \delta^2 = 30$

		$\alpha = 0.1$				$\alpha = 1$			
	normal	$\hat{e}_\beta$	t	t_{largeK}	t_{adj}	$\hat{e}_\beta$	t	t_{largeK}	t_{adj}
X05	-1.65	-1.98	-2.27	-1.57	-1.86	-1.34	-2.60	-2.00	-1.67
X10	-1.28	-1.40	-1.85	-1.23	-1.39	-1.08	-1.95	-1.51	-1.29
MEDN	0	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01
X90	1.28	1.47	1.71	1.15	1.36	1.68	1.23	0.96	1.21
X95	1.65	2.14	2.04	1.43	1.77	2.46	1.42	1.13	1.50
$P(t < z_{05})$	5.0	7.4	13.1	4.1	6.8	2.1	13.5	8.4	5.3
$P(t < z_{10})$	10.0	11.7	19.5	9.2	11.7	6.0	19.1	13.3	10.1
$P(t > z_{90})$	10.0	12.3	18.3	7.4	11.4	14.8	8.4	2.3	8.5
$P(t > z_{95})$	5.0	8.3	11.1	2.7	6.4	10.3	1.6	0.2	3.2

Table 3: Distributions of t ratios under H_0: $n - K = 100, K_2 = 50, \delta^2 = 50$

		$\alpha = 0.1$				$\alpha = 1$			
	normal	$\hat{e}_\beta$	t	t_{largeK}	t_{adj}	$\hat{e}_\beta$	t	t_{largeK}	t_{adj}
X05	-1.65	-1.87	-2.50	-1.65	-1.86	-1.39	-2.62	-1.97	-1.68
X10	-1.28	-1.35	-2.02	-1.29	-1.41	-1.13	1.98	-1.48	-1.28
MEDN	0	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	-0.01
X90	1.28	1.43	1.85	1.19	1.35	1.60	1.38	1.03	1.25
X95	1.65	2.03	2.28	1.50	1.78	2.22	1.62	1.24	1.57
$P(t < z_{05})$	5.0	6.7	15.3	5.1	7.1	2.3	13.8	8.0	5.5
$P(t < z_{10})$	10.0	11.1	21.4	10.2	12.0	6.9	19.3	13.1	10.1
$P(t > z_{90})$	10.0	12.0	20.0	8.2	11.1	14.2	12.1	4.2	9.4
$P(t > z_{95})$	5.0	7.7	13.2	3.4	6.3	9.5	4.7	0.7	4.3

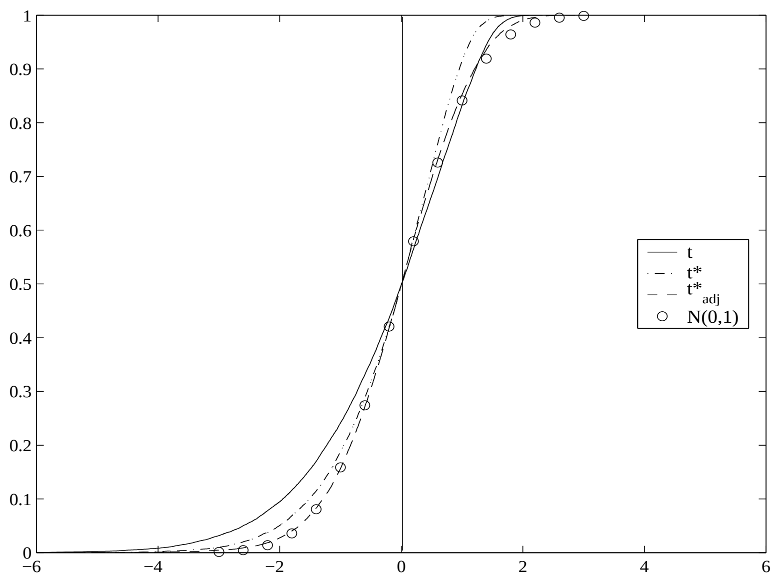

Figure 2: The null distributions of t, t_{largeK}, and t_{adj}: $n - K = 100, K_2 = 30, \alpha = 1, \delta^2 = 30$

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

  Anderson and Rubin (AR) statsitic is given by

$$AR = \frac{(1, -\boldsymbol{\beta}_0{}')\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}(1, -\boldsymbol{\beta}_0{}')'}{(1, -\boldsymbol{\beta}_0{}')\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}(1, -\boldsymbol{\beta}_0{}')'/(n-K)}. \tag{4.5}$$

  Because, under the null hypothesis, we have

$$AR = \frac{\boldsymbol{u}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{u}}{\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}/(n-K)}, \tag{4.6}$$

  the null distribution of the AR statistic does not depend on δ^2. 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, $AR \xrightarrow{d} \chi^2(K_2)$ under the null hypothesis.

- **Score-type Test**

  Define the statistics

$$\boldsymbol{S} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0(\boldsymbol{b}_0{}'\boldsymbol{\Omega}\boldsymbol{b}_0)^{-1/2} \tag{4.7}$$

and

$$\boldsymbol{T} = (\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix}\left[(\boldsymbol{\beta}_0, \boldsymbol{I}_{G_1})\boldsymbol{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_0{}' \\ \boldsymbol{I}_{G_1} \end{pmatrix}\right]^{-1/2}, \qquad (4.8)$$

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

$$K = \hat{\boldsymbol{S}}'\hat{\boldsymbol{T}}(\hat{\boldsymbol{T}}'\hat{\boldsymbol{T}})^{-1}\hat{\boldsymbol{T}}'\hat{\boldsymbol{S}}. \qquad (4.9)$$

Kleibergen showed that under either the standard large sample asymptotics or weak-instrument asymptoics, $K \xrightarrow{d} \chi^2(G_1)$ under the null hypothesis, 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

$$LR = \frac{\boldsymbol{b}_0'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}_0}{\boldsymbol{b}_0'\boldsymbol{\Omega}\boldsymbol{b}_0} - \min_{\boldsymbol{b}} \frac{\boldsymbol{b}'\boldsymbol{Y}'(\boldsymbol{P}_Z - \boldsymbol{P}_{Z_1})\boldsymbol{Y}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{\Omega}\boldsymbol{b}}. \qquad (4.10)$$

  Moreira (2003) showed that the LR statistic is a function of $\boldsymbol{S}$ and $\boldsymbol{T}$ defined in (4.7) and (4.8), 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) suggested computing the null distribution by Monte Carlo simulation or numerical integration. In practice, $\boldsymbol{\Omega}$ 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.)

We generate 5,000 datasets from DGP (4.3) and (4.4) for various values of β and report size-corrected power curves at the 5% significance level. 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. Figures 3-6 display the power curves in the case in which $K_2 = 3$. Figures 7-10 display the power curves in the case of many (weak) instruments–$K_2 = 30$ and $\delta^2 = 30$.

Our results are similar to the results of Kleibergen (2002), Moreira(2003), and Guggenberger and Smith (2005): (i) The power of the AR test decreases substantially when the number of instruments increases, (ii) The CLR test is usually more powerful than the AR and the K tests, and (iii) The most important finding of our experiments is power curve of the large K t-ratio uniformly dominates the power curves of the other tests. Occasionally, their differences are rather large.

4.3 Effects of Normality

Since the distributions of estimators and t-ratio 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 $n - K = 100, K_2 = 30, \alpha = 1$, and $\delta^2 = 30$ and reported the 5, 10, 50, 90, and 95 percentiles of the null distributions and the observed sizes at the 10% and 5% asymptotic critical values of t, t_{largeK}, t_{adj}, and $t^{\dagger}_{largeK}$ in Table 4. From these experiments, the size properties of the three t-ratio statistics, t, t_{largeK}, t_{adj}, which are derived under the assumption of normal disturbances, are approximately valid even if the distributions of disturbances are deviated from normal. The power curves change slightly when the distributions of the disturbances have long tails. Figures 11-12 contain the power curves when the distributions of disturbances are $t(3)$ distributions. However, the large K t-test continues to uniformly dominate the others in these cases.

5. Conclusions

When the number of instruments is large, the null distribution of the standard t-ratio (t_{LIML}) is skewed and extremely deviated from the normal distribution. The

Table 4: Distributions of t ratios under H_0 (The Cases of Non-normal Disturbances): $n - K = 100, K_2 = 30, \delta^2 = 30$

		$n - K = 100, K_2 = 30, \delta^2 = 30,\ \alpha = 1$							
		$u_i = (\chi^2(3) - 3)/\sqrt{6}$				$u_i = t(3)$			
	normal	t	t_{largeK}	t_{adj}	$t^{\dagger}_{largeK}$	t	t_{largeK}	t_{adj}	$t^{\dagger}_{largeK}$
X05	-1.65	-2.39	-1.85	-1.74	-1.84	-2.60	-2.02	-1.69	-2.01
X10	-1.28	-1.82	-1.41	-1.32	-1.40	-1.94	-1.50	-1.29	-1.49
MEDN	0	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
X90	1.28	1.38	1.06	1.25	1.06	1.22	0.95	1.18	0.94
X95	1.65	1.68	1.31	1.62	1.31	1.43	1.13	1.47	1.13
$P(t < z_{05})$	5.0	12.1	7.0	5.9	7.0	13.7	8.4	5.5	8.2
$P(t < z_{10})$	10.0	18.0	11.9	10.6	11.8	19.1	13.4	10.2	13.3
$P(t > z_{90})$	10.0	12.1	5.5	9.5	5.4	8.4	2.6	8.1	2.6
$P(t > z_{95})$	5.0	5.4	1.4	4.8	1.4	2.2	0.3	3.2	0.3

null distribution of a large K t-ratio (t_{largeK}) is closer to the normal distribution, but it continues to be skewed and the size distortion can be large, particularly for the one-sided test.

In order to explore the finite sample properties of the large K t-ratio, we derived an asymptotic expansion of the null distribution both in the cases of the normal and non-normal disturbances. We proposed an adjusted large K t-ratio (t_{adj}) from the asymptotic exapansion. The actual size of t_{adj} is shown to be close to the nominal size.

We also have found that the power of the large K t-ratio test dominates the AR test, the K-test, and the conditional LR test. It may be stated that the large K t-test should be used when the number of instruments is large (except when the instruments are too weak). When we know the sign of the parameter from the economic theory, the use of the modified large K t-ratio statistic is recommended as a more accurate test procedure.

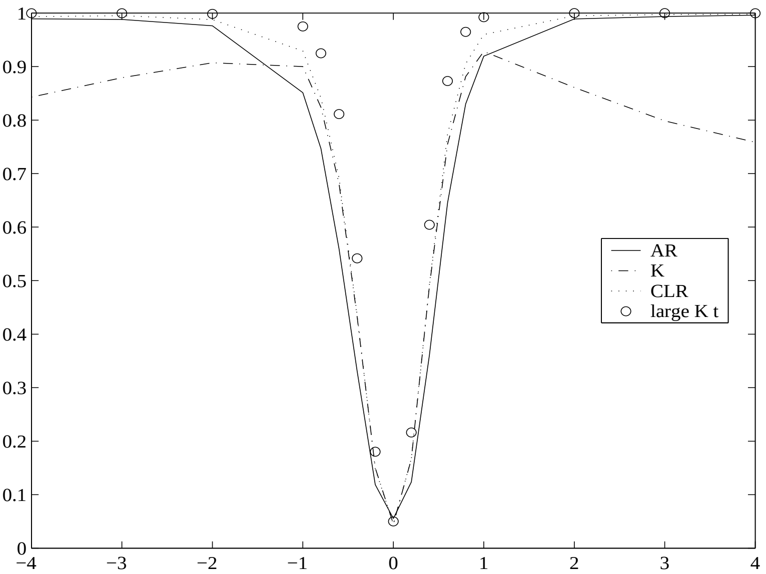

Figure 3: Power of tests: $n - K = 30, K_2 = 3, \alpha = 0.1, \delta^2 = 30$

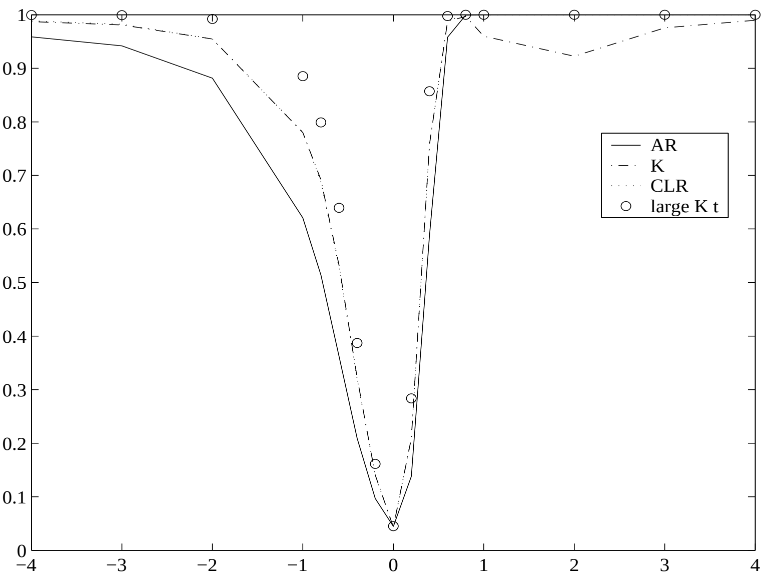

Figure 4: Power of tests: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 30$

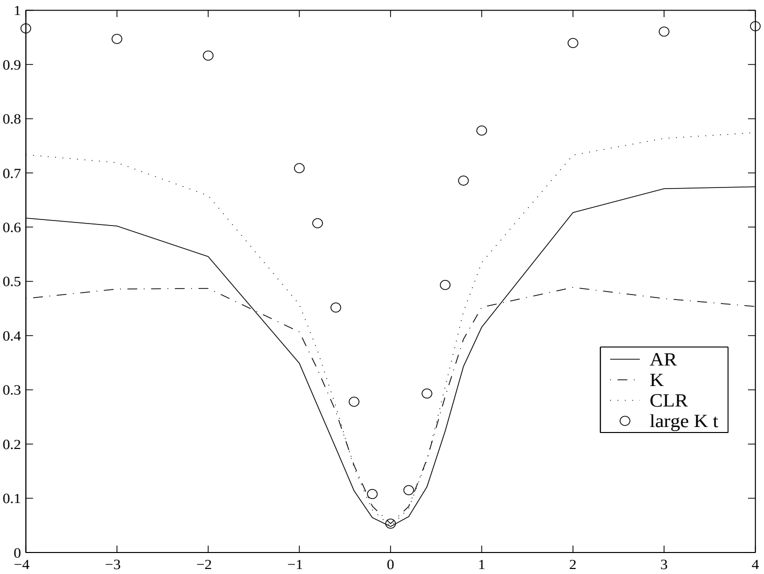

Figure 5: Power of tests: $n - K = 30, K_2 = 3, \alpha = 0.1, \delta^2 = 10$

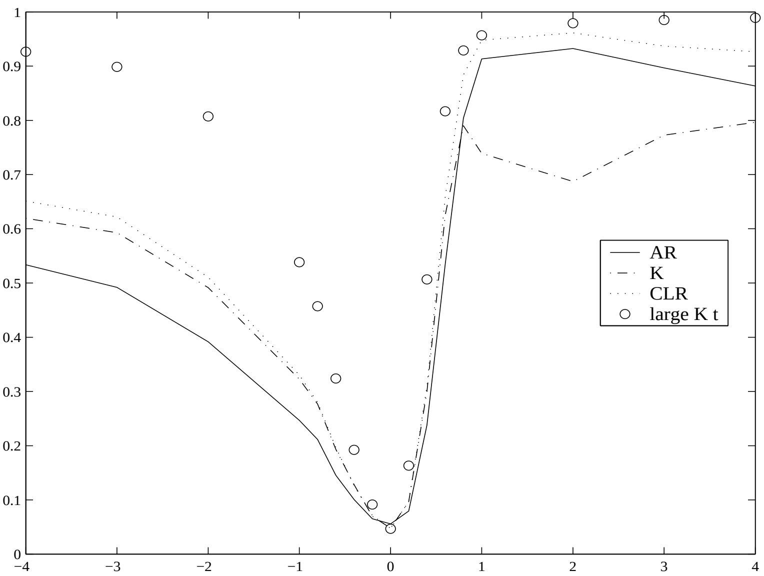

Figure 6: Power of tests: $n - K = 30, K_2 = 3, \alpha = 1, \delta^2 = 10$

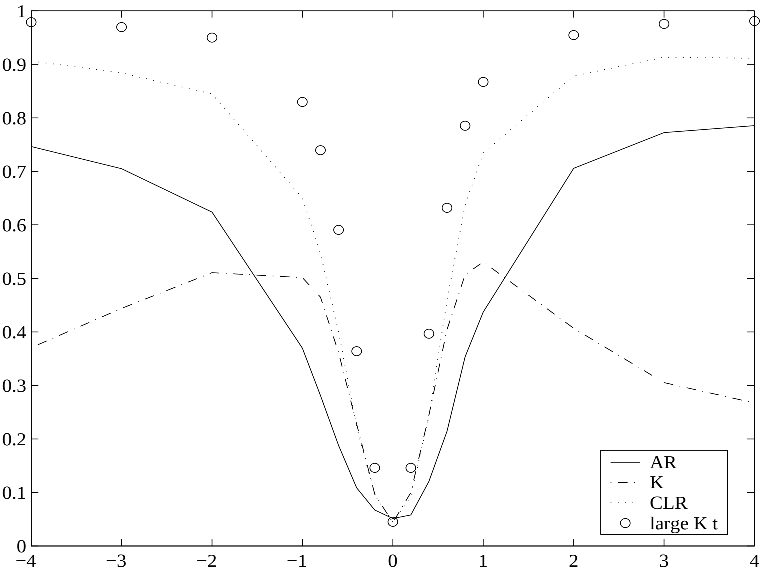

Figure 7: Power of tests: $n - K = 100, K_2 = 30, \alpha = 0.1, \delta^2 = 30$

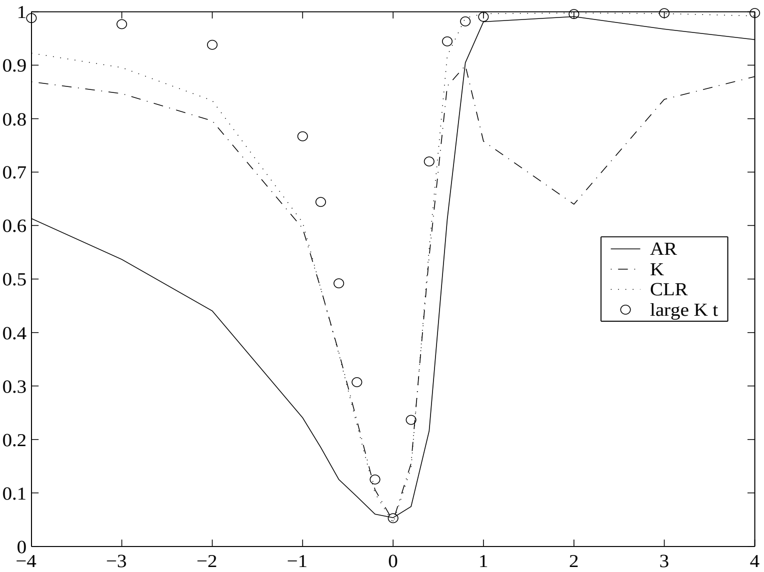

Figure 8: Power of tests: $n - K = 100, K_2 = 30, \alpha = 1, \delta^2 = 30$

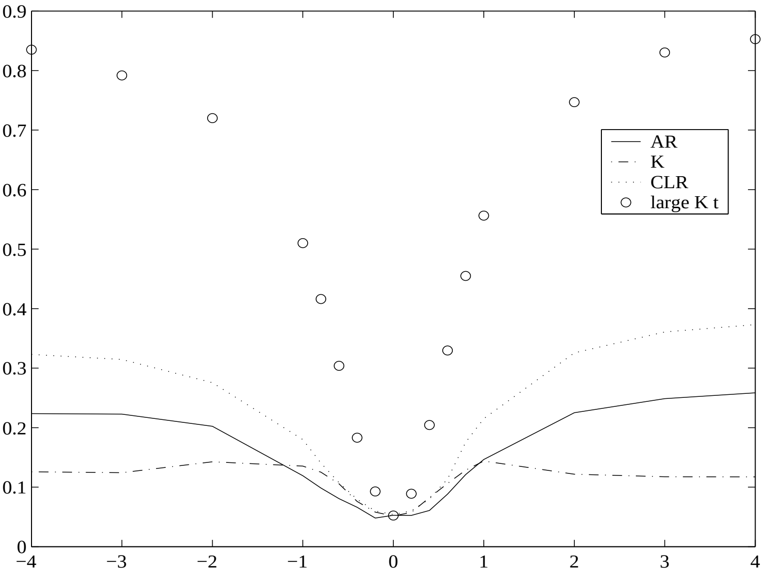

Figure 9: Power of tests: $n - K = 100, K_2 = 30, \alpha = 0.1, \delta^2 = 10$

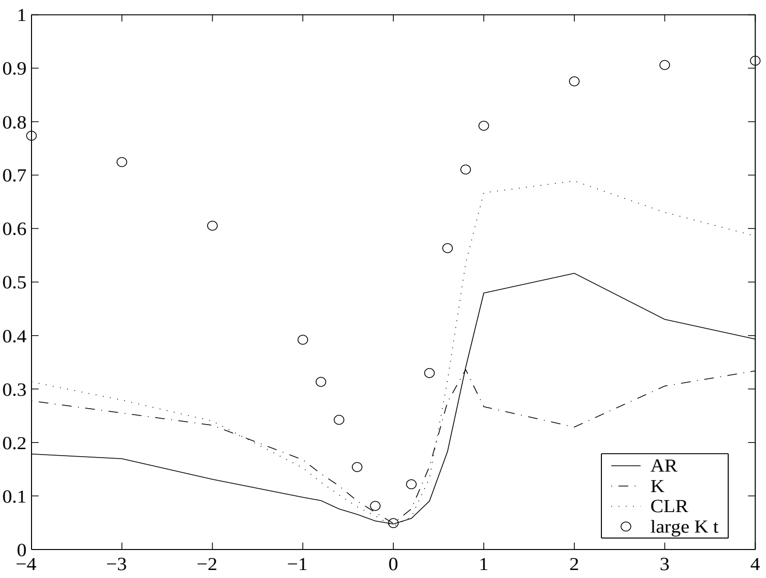

Figure 10: Power of tests: $n - K = 100, K_2 = 30, \alpha = 0.1, \delta^2 = 10$

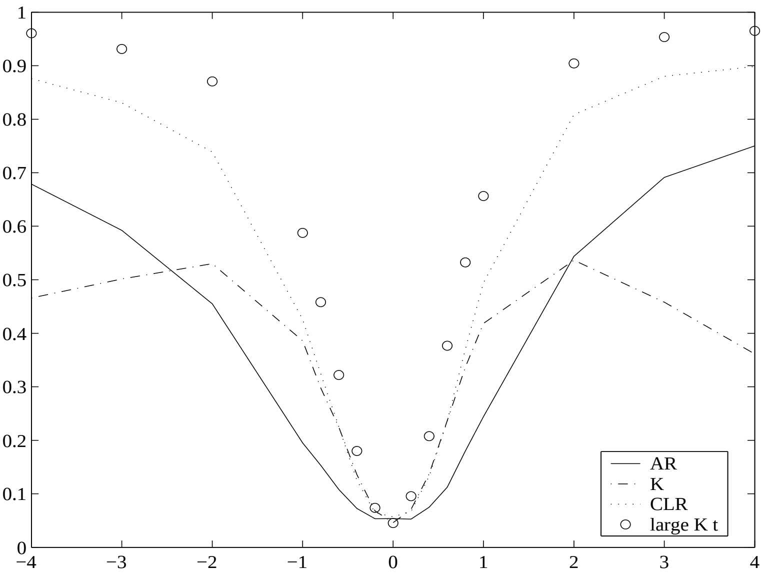

Figure 11: Power of tests: $n - K = 100, K_2 = 30, \alpha = 0.1, \delta^2 = 30, u_i = t(3)$

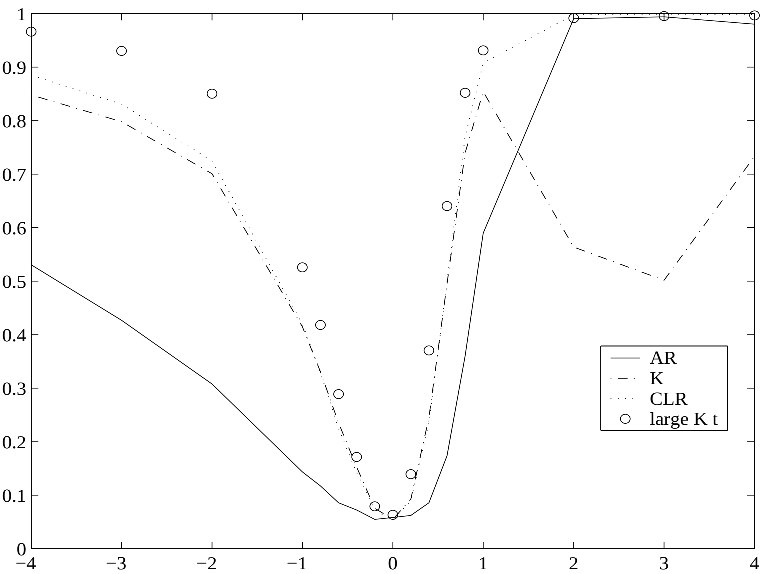

Figure 12: Power of tests: $n - K = 100, K_2 = 30, \alpha = 1, \delta^2 = 30, u_i = t(3)$

APPENDIX

A. Asymptotic expansion of the distribution of the LIML estimator

In this section we give the asymptotic expansion of the distribution of the LIML estimator

$$\hat{\boldsymbol{e}}_{LI} = \begin{pmatrix} \hat{\boldsymbol{e}}_{\beta} \\ \hat{\boldsymbol{e}}_{\gamma} \end{pmatrix} = \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{LI} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\gamma}}_{LI} - \boldsymbol{\gamma} \end{pmatrix} \tag{A.1}$$

under the large-K_n asymptotics.

Theorem A.1 *When the rows of $\boldsymbol{V}$ are normally distributed, the asymptotic expansion of the joint distribuiton of $\hat{\boldsymbol{e}}_{LI}$ under the sequence (2.13) is given by*

$$P\{\hat{\boldsymbol{e}}_{LI} \leq \boldsymbol{\xi}\} = \Phi_{\Psi}(\boldsymbol{\xi}) + \frac{1}{\sqrt{n}}(\boldsymbol{q}_2'\boldsymbol{\xi})\boldsymbol{\xi}\phi_{\Psi}(\xi) + O(n^{-1}), \tag{A.2}$$

where $\boldsymbol{\xi}$ is a $(G_1 + K_1)$ vector and $\Phi_{\Psi}(\boldsymbol{\xi})$ and $\phi_{\Psi}(\boldsymbol{\xi})$ are the cdf and the density function of the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Psi}$, respectively.

We note that when $c_1 = 0$, this asymptotic expansion is identical to the result under the standard large sample theory up to $O(n^{-1/2})$. See Fujikoshi et al (1982). When $G_1 = 1$, integrating (A.2) with respect to the last K_1 elements of $\boldsymbol{\xi}$, the asymptotic expansion of the marginal joint distribution of a standardized statistic

$$\hat{e}_{\beta}^{*} = \frac{\sqrt{\Pi_{22}' A_{22.1} \Pi_{22}}}{\sigma} \frac{1}{\sqrt{\eta}}(\hat{\beta}_{LI} - \beta) \tag{A.3}$$

is derived. Here, we use the notations $\eta = 1 + \frac{1}{\tau^2}(1 + \frac{\nu^2}{\tau^2})$, $\nu^2 = \lim_{n\to\infty} \frac{\mu^2}{q}$, $\tau^2 = \lim_{n\to\infty} \frac{\mu^2}{L}$, $\mu^2 = \frac{\sigma^2}{|\Omega|}\Pi_{22}' A_{22.1}\Pi_{22}$, $A_{22.1} = \boldsymbol{Z}_2'\boldsymbol{Z}_2 - \boldsymbol{Z}_2'\boldsymbol{Z}_1(\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1'\boldsymbol{Z}_2$, and $L = K_2 - 1$.

Corollary A.1 *When* $G_1 = 1$*, the asymptotic expansion of the distribution of* $\hat{e}^*_\beta$ *under the sequence (2.13) is given by*

$$P\{\hat{e}^*_\beta \le \xi\} = \Phi(\xi) - \frac{\alpha}{\mu}\sqrt{\eta}\xi^2\phi(\xi) + o(\mu^{-1}), \tag{A.4}$$

where $\alpha = (\omega_{22}\beta - \omega_{21})/\sqrt{|\Omega|}$.

This expansion of the distribution of $\hat{\mathbf{e}}^*_\beta$ is identical to the result of Morimune (1983) up to $O(n^{-1/2})$. The derivation of the asymptotic expansion is provided next.

Derivation of Theorem A.1:

In order to derive the asymptotic expansion of the distribution of the LIML estimator, we derive the stochastic expansion of the LIML estimator $\hat{\mathbf{e}}_{LI}$ first. The LIML estimator is defined by (2.6), which is rewritten as follows:

$$\begin{aligned} &\begin{pmatrix} \boldsymbol{Y}' \\ \boldsymbol{Z}_1' \end{pmatrix} (\boldsymbol{P}_Z - \hat{\lambda}\bar{\boldsymbol{P}}_Z)(\boldsymbol{Y},\ \boldsymbol{Z}_1)\hat{\mathbf{e}}_{LI} \qquad \text{(A.5)} \\ &= \sqrt{n}\begin{pmatrix} \boldsymbol{Y}' \\ \boldsymbol{Z}_1' \end{pmatrix} (\boldsymbol{P}_Z - \hat{\lambda}\bar{\boldsymbol{P}}_Z)(\boldsymbol{Y},\ \boldsymbol{Z}_1)\begin{pmatrix} 1 \\ -\boldsymbol{\beta} \\ -\boldsymbol{\gamma} \end{pmatrix}. \end{aligned}$$

Defining

$$\boldsymbol{D} = (\boldsymbol{D}_1\ \ \boldsymbol{D}_2) = \left(\begin{pmatrix} \boldsymbol{\pi}_{11} \\ \boldsymbol{\pi}_{21} \end{pmatrix} \ \begin{pmatrix} \boldsymbol{\Pi}_{12} & \boldsymbol{I}_{K_1} \\ \boldsymbol{\Pi}_{22} & \mathbf{0} \end{pmatrix}\right), \tag{A.6}$$

we can write

$$\begin{aligned} &\begin{pmatrix} \boldsymbol{Y}' \\ \boldsymbol{Z}_1' \end{pmatrix} (\boldsymbol{P}_Z - \hat{\lambda}\bar{\boldsymbol{P}}_Z)(\boldsymbol{Y},\ \boldsymbol{Z}_1) \qquad \text{(A.7)} \\ &= \{\boldsymbol{Z}\boldsymbol{D} + (\boldsymbol{V}\ \mathbf{0})\}'(\boldsymbol{P}_Z - \hat{\lambda}\bar{\boldsymbol{P}}_Z)\{\boldsymbol{Z}\boldsymbol{D} + (\boldsymbol{V}\ \mathbf{0})\} \\ &= \boldsymbol{D}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{D} + \boldsymbol{D}'\boldsymbol{Z}'(\boldsymbol{V}\ \mathbf{0}) + \begin{pmatrix} \boldsymbol{V}' \\ \mathbf{0}' \end{pmatrix}\boldsymbol{Z}\boldsymbol{D} + \begin{pmatrix} \boldsymbol{V}' \\ \mathbf{0}' \end{pmatrix}(\boldsymbol{P}_Z - \hat{\lambda}\bar{\boldsymbol{P}}_Z)(\boldsymbol{V}\ \mathbf{0}). \end{aligned}$$

We define $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ such that

$$\frac{1}{K}\begin{pmatrix} \boldsymbol{V}' \\ \mathbf{0}' \end{pmatrix}\boldsymbol{P}_Z(\boldsymbol{V}\ \mathbf{0}) = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \frac{1}{\sqrt{K}}\boldsymbol{E}_1, \tag{A.8}$$

and

$$\frac{1}{q}\begin{pmatrix} \boldsymbol{V}' \\ \boldsymbol{0}' \end{pmatrix} \bar{\boldsymbol{P}}_Z(\boldsymbol{V}\ \boldsymbol{0}) \;=\; \begin{pmatrix} \boldsymbol{\Omega} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} + \frac{1}{\sqrt{q}}\boldsymbol{E}_2. \tag{A.9}$$

We use the following notations:

$$\boldsymbol{W}_2 = (\boldsymbol{V}_2\ \boldsymbol{0}) - \boldsymbol{u}\boldsymbol{q}_2', \tag{A.10}$$

and $\boldsymbol{q}_2 = \frac{1}{\sigma^2}(\boldsymbol{\omega}_{21}' - \boldsymbol{\beta}'\boldsymbol{\Omega}_{22}\ \boldsymbol{0}')'$. By substituting (A.7), (A.8), and (A.9) into (A.5) and putting

$$\hat{\boldsymbol{e}}_{LI} \;=\; \boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{e}^{(1)} + o_p(n^{-1/2}), \tag{A.11}$$

$$\hat{\lambda} \;=\; \lambda^{(0)} + \frac{1}{\sqrt{n}}\lambda^{(1)} + o_p(n^{-1/2}), \tag{A.12}$$

we can determine each $\boldsymbol{e}^{(i)}$ and $\lambda^{(i)}$ successively, which is given as follows:

$$\lambda^{(0)} = c_2, \tag{A.13}$$

$$\lambda^{(1)} = \frac{c_2}{\sigma^2}\{\frac{1}{\sqrt{c_1}}(\frac{1}{\sqrt{K}}\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u}) - \sqrt{\frac{c_2}{c_1}}\frac{1}{\sqrt{q}}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u}\}, \tag{A.14}$$

$$\boldsymbol{e}^{(0)} = \boldsymbol{Q}^{-1}[\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{u} + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u} - \frac{\sqrt{c_1c_2}}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}], \tag{A.15}$$

$$\begin{aligned}\boldsymbol{e}^{(1)} = -\boldsymbol{Q}^{-1}[\{&\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'(\boldsymbol{V}_2\ \boldsymbol{0}) + \frac{\sqrt{c_1}}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z(\boldsymbol{V}_2\ \boldsymbol{0}) \\ &-\sqrt{c_1c_2}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2\ \boldsymbol{0})\}\boldsymbol{e}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{W}_2'\boldsymbol{Z}\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\ &-\frac{c_1}{c_2}\lambda^{(1)}\left[\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^2\right]\boldsymbol{e}^{(0)} + \sqrt{\frac{c_1}{c_2}}\lambda^{(1)}\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}].\end{aligned} \tag{A.16}$$

Each $\boldsymbol{\lambda}^{(i)}$ is obtained by premultiplying $(1, -\boldsymbol{\beta}', -\boldsymbol{\gamma}')$ to (A.5). Each $\boldsymbol{e}^{(i)}$ is obtained by using the last $G_1 + K_1$ rows of (A.5).

It should be noted that $\boldsymbol{W}_2$ and $\boldsymbol{u}$ are uncorrelated when $\{\boldsymbol{v}_i\}$ are independently distributed. In light of this fact, we note that the Cornish-Fisher expansion of $\frac{1}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u}$ and $\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}$ can be written as

$$\frac{1}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u} \;=\; \boldsymbol{X}_0 + O_p(n^{-1}) \tag{A.17}$$

$$\frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u} \;=\; \boldsymbol{Y}_0 + O_p(n^{-1}), \tag{A.18}$$

where $\boldsymbol{X}_0$ and $\boldsymbol{Y}_0$ are distributed independently as both $N(\boldsymbol{0}\ \sigma^2\boldsymbol{C}_2)$, where $\boldsymbol{C}_2 = E[\boldsymbol{w}_{2i}\boldsymbol{w}'_{2i}]$. Hence, we can rewrite $\hat{\boldsymbol{e}}_{LI}$ as

$$\hat{\boldsymbol{e}}_{LI} = \tilde{\boldsymbol{e}}^{(0)} + \frac{1}{\sqrt{n}}\boldsymbol{e}^{(1)} + o_p(n^{-1/2}), \tag{A.19}$$

where

$$\tilde{\boldsymbol{e}}^{(0)} \equiv \boldsymbol{x} = \boldsymbol{Q}^{-1}[\frac{1}{\sqrt{n}}\boldsymbol{D}'_2\boldsymbol{Z}'\boldsymbol{u} + \sqrt{c_1}\boldsymbol{X}_0 + \sqrt{c_1c_2}\boldsymbol{Y}_0]. \tag{A.20}$$

We derive an asymptotic expansion of the distribution of $\hat{\boldsymbol{e}}_{LI}$ by inverting the characteristic function of $\hat{\boldsymbol{e}}_{LI}$ up to order $n^{-1/2}$:

$$C(t) = E[exp(i\boldsymbol{t}'\boldsymbol{x})] + \frac{1}{\sqrt{n}}E[i\boldsymbol{t}'E(\boldsymbol{e}^{(1)}|\boldsymbol{x})exp(i\boldsymbol{t}'\boldsymbol{x})] + O(n^{-1}), \tag{A.21}$$

where $\boldsymbol{t} = (t_i)$ is a $(G_1+K_1)\times 1$ vector of real variables and $i^2 = -1$. The conditional expectation of $\boldsymbol{e}^{(1)}$, given the first order term $\boldsymbol{x}$, is calculated as

$$E(\boldsymbol{e}^{(1)}|\boldsymbol{x}) = -\boldsymbol{x}(\boldsymbol{x}'\boldsymbol{q}_2) + O_p(n^{-1/2}). \tag{A.22}$$

The probability $P(\hat{\boldsymbol{e}} \leq \boldsymbol{\xi})$ is approximated to the order $n^{-1/2}$ by the Fourier inverse transformation of the characteristic function (A.21). The inverse transformation of the first term is $\Phi_\Psi(\xi)$. We also use the next Fourier Inversion formula that was developed by Fujikoshi et al (1982): for any polynomials $h(\cdot)$ and $g(\cdot)$,

$$\mathcal{F}^{-1}[h(-i\boldsymbol{t})E(g(\boldsymbol{x})exp(i\boldsymbol{t}'\boldsymbol{x}))]_{\boldsymbol{x}=\boldsymbol{\xi}} = h\left(\frac{\partial}{\partial\boldsymbol{\xi}}\right)g(\boldsymbol{\xi})\phi_\Psi(\boldsymbol{\xi}), \tag{A.23}$$

where $\partial/\partial\boldsymbol{\xi}' = (\partial/\partial\xi_1, \cdots, \partial/\partial\xi_{G_1+K_1})$.

Then, we have

$$P(\hat{\boldsymbol{e}}_{LI} \leq \boldsymbol{\xi}) = \Phi_\Psi(\boldsymbol{\xi}) + \frac{1}{\sqrt{n}}(\boldsymbol{q}'_2\boldsymbol{\xi})\boldsymbol{\xi}\phi_\Psi(\xi) + O(n^{-1}), \tag{A.24}$$

where $\boldsymbol{\xi}$ is a $(G_1+K_1)\times 1$ vector and, $\Phi_\Psi(\boldsymbol{\xi})$ and $\phi_\Psi(\xi)$ are the multivariate normal cdf and density function with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Psi}$, respectively.

B. Asymptotic expansion of the null distribution of the large K t-ratio statistic

In order to derive theasymptotic expansion of the null distribution of the large K t-ratio statistic, we need to expand stochastically each term of $\hat{\boldsymbol{\Psi}}$ as follows:

$$\hat{\boldsymbol{Q}} = \boldsymbol{Q} + \frac{1}{\sqrt{n}}[\frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'(\boldsymbol{V}_2\ \boldsymbol{0}) + \frac{1}{\sqrt{n}}\begin{pmatrix} \boldsymbol{V}_2' \\ \boldsymbol{0}' \end{pmatrix}\boldsymbol{Z}\boldsymbol{D}_2 \tag{B.1}$$

$$+\sqrt{c_1}\boldsymbol{J}_2'\boldsymbol{E}_1\boldsymbol{J}_2 + \sqrt{c_1c_2}\boldsymbol{J}_2'\boldsymbol{E}_2\boldsymbol{J}_2 - \lambda^{(1)}\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}] + O_p(n^{-1}).$$

Here, we have used the $(1+G_1+K_1)\times(G_1+K_1)$ choice matrix $\boldsymbol{J}_2 = (\boldsymbol{0}\ \boldsymbol{I}_{G_1+K_1})'$. Hence,

$$\hat{\boldsymbol{Q}}^{-1} = \boldsymbol{Q}^{-1} - \frac{1}{\sqrt{n}}\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1} + O_p(n^{-1}), \tag{B.2}$$

where

$$\boldsymbol{B} = \frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'(\boldsymbol{V}_2\ \boldsymbol{0}) + \frac{1}{\sqrt{n}}\begin{pmatrix} \boldsymbol{V}_2' \\ \boldsymbol{0}' \end{pmatrix}\boldsymbol{Z}\boldsymbol{D}_2$$

$$+\sqrt{c_1}\boldsymbol{J}_2'\boldsymbol{E}_1\boldsymbol{J}_2 - \sqrt{c_1c_2}\boldsymbol{J}_2'\boldsymbol{E}_2\boldsymbol{J}_2 - \lambda^{(1)}\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

$$\frac{1}{q}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}} \tag{B.3}$$

$$= \{\boldsymbol{b}_0 - \frac{1}{\sqrt{n}}\boldsymbol{J}_2\boldsymbol{e}_\beta^{(0)} - \frac{1}{n}\boldsymbol{J}_2\boldsymbol{e}_\beta^{(1)} + \cdots\}'$$

$$\times\{\boldsymbol{\Omega} + \frac{1}{\sqrt{n}}[\sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{V}'\bar{\boldsymbol{P}}_Z\boldsymbol{V} - \boldsymbol{\Omega})]\}\{\boldsymbol{b}_0 - \frac{1}{\sqrt{n}}\boldsymbol{J}_2\boldsymbol{e}_\beta^{(0)} - \frac{1}{n}\boldsymbol{J}_2\boldsymbol{e}_\beta^{(1)} + \cdots\}$$

$$= \sigma^2 + \frac{1}{\sqrt{n}}[-2\boldsymbol{e}_\beta^{(0)'}\boldsymbol{J}_2'\boldsymbol{\Omega}\boldsymbol{b}_0 + \sqrt{\frac{c_2}{c_1}}\sqrt{q}\boldsymbol{b}_0'(\frac{1}{q}\boldsymbol{V}'\bar{\boldsymbol{P}}_Z\boldsymbol{V} - \boldsymbol{\Omega})\boldsymbol{b}_0] + O_p(n^{-1}),$$

where we have used the notation $\boldsymbol{b}_0 = (1, -\boldsymbol{\beta}')'$. Similarly,

$$\frac{1}{q^2}\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y}\hat{\boldsymbol{b}}\hat{\boldsymbol{b}}'\boldsymbol{Y}'\bar{\boldsymbol{P}}_Z\boldsymbol{Y} \tag{B.4}$$

$$= \boldsymbol{\Omega}\boldsymbol{b}_0\boldsymbol{b}_0'\boldsymbol{\Omega}$$

$$+\frac{1}{\sqrt{n}}[-\boldsymbol{\Omega}\boldsymbol{b}_0\boldsymbol{e}_\beta^{(0)'}\boldsymbol{J}_2'\boldsymbol{\Omega} + \sqrt{\frac{c_2}{c_1}}\boldsymbol{\Omega}\boldsymbol{b}_0\boldsymbol{b}_0'\sqrt{q}(\frac{1}{q}\boldsymbol{V}'\bar{\boldsymbol{P}}_Z\boldsymbol{V} - \boldsymbol{\Omega})$$

$$-\boldsymbol{\Omega}\boldsymbol{J}_2\boldsymbol{e}_\beta^{(0)}\boldsymbol{b}_0'\boldsymbol{\Omega} + \sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{V}'\bar{\boldsymbol{P}}_Z\boldsymbol{V} - \boldsymbol{\Omega})\boldsymbol{b}_0\boldsymbol{b}_0'\boldsymbol{\Omega}] + O(n^{-1})$$

Then, we have

$$\hat{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \frac{1}{\sqrt{n}}\boldsymbol{\Psi}^{(1)} + O_p(n^{-1}), \tag{B.5}$$

where

$$\begin{aligned}
\boldsymbol{\Psi}^{(1)} &= \boldsymbol{Q}^{-1}[-2\boldsymbol{e}^{(0)\prime}\boldsymbol{q}_2\sigma^2 + \sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{P}_Z\boldsymbol{u} - \sigma^2)] \\
&\quad -\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1}\sigma^2 \\
&\quad +c_1(1+c_2)\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1} \\
&\quad -c_1(1+c_2)\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1}\left[\begin{pmatrix} \sigma^2\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1} \\
&\quad -c_1(1+c_2)\mathbf{Q}^{-1}\left[\begin{pmatrix} \sigma^2\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1}, \\
\boldsymbol{A} &= -2\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{e}_\beta^{\prime(0)}\boldsymbol{q}_2\sigma^2 + \sqrt{\frac{c_2}{c_1}}\sqrt{q}\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}(\frac{1}{q}\boldsymbol{u}'\bar{P}_Z\boldsymbol{u} - \sigma^2) \\
&\quad +\sqrt{\frac{c_2}{c_1}}\boldsymbol{J}_2'\boldsymbol{E}_2\sigma^2 + \boldsymbol{q}_2\sigma^2\boldsymbol{e}_\beta^{\prime(0)}\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \\
&\quad -\sqrt{\frac{c_2}{c_1}}\boldsymbol{q}_2\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z(\boldsymbol{V}_2,\boldsymbol{0}) - \boldsymbol{q}_2'\sigma^2)\sigma^2 \\
&\quad +\begin{pmatrix} \boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{e}_\beta^{(0)}\boldsymbol{q}_2'\sigma^2 - \sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\begin{pmatrix} \boldsymbol{V}_2' \\ \boldsymbol{0}' \end{pmatrix}\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \boldsymbol{q}_2\sigma^2)\boldsymbol{q}_2'\sigma^2.
\end{aligned}$$

Then, the inequality $\{t \leq \xi\}$ is approximated as

$$x_i + \frac{1}{\sqrt{n}}t^{(1)} \leq \xi + O_p(n^{-1}), \tag{B.6}$$

where

$$x_i = \frac{\boldsymbol{i}'\tilde{\boldsymbol{e}}^{(0)}}{\sqrt{\Psi_{ii}}} \tag{B.7}$$

$$t^{(1)} = \frac{\boldsymbol{i}'\boldsymbol{e}^{(1)}}{\sqrt{\Psi_{ii}}} - \frac{1}{2}\frac{\Psi^{(1)}{}_{ii}}{\Psi_{ii}}x_i. \tag{B.8}$$

The first order term x_i is distributed as the standard normal random variable, and the expectation of $t^{(1)}$ conditional upon x_i is calculated as

$$E(t^{(1)}|x_i) = \frac{1}{\sqrt{\Psi_{ii}}}(\boldsymbol{i}'\boldsymbol{\Psi}\boldsymbol{q}_2)x_i^2 + O_p(n^{-1/2}). \tag{B.9}$$

The asymptotic expansion of the distribution is derived by using the same formula as (A.23).

$$P(t^* \le \xi) = \Phi(\xi) - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\Psi_{ii}}}(\boldsymbol{i}'\boldsymbol{\Psi}\boldsymbol{q}_2)\xi^2\phi(\xi) + O(n^{-1}), \tag{B.10}$$

where $\Phi(\xi)$ and $\phi(\xi)$ are the standard normal cdf and density function, respectively.

The validity of the expansion is given by the same method as that in Appendix C in Fujikoshi et al (1982). The random variables that appear in our analyses are $\boldsymbol{x}_1 = \frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{u}$, $\boldsymbol{x}_2 = \frac{1}{\sqrt{n}}\boldsymbol{D}_2'\boldsymbol{Z}'\boldsymbol{W}_2$, $\boldsymbol{w}_1 = \frac{1}{\sqrt{K}}(\boldsymbol{u}'\boldsymbol{P}_Z\boldsymbol{u} - \sigma^2)$, $\boldsymbol{w}_2 = \frac{1}{\sqrt{q}}(\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2)$, $\boldsymbol{w}_3 = \frac{1}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{u}$, $\boldsymbol{w}_4 = \frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{u}$, $\boldsymbol{w}_5 = \frac{1}{\sqrt{K}}\boldsymbol{W}_2'\boldsymbol{P}_Z\boldsymbol{W}_2$, and $\boldsymbol{w}_6 = \frac{1}{\sqrt{q}}\boldsymbol{W}_2'\bar{\boldsymbol{P}}_Z\boldsymbol{W}_2$. We use the space J_n where each element of $\boldsymbol{x}_i, where i = 1, 2$, is in the interval $(-2c\sqrt{\log n}, 2c\sqrt{\log n})$ and c is a standard deviation of each random variable; each element of $\boldsymbol{w}_i$, where $i = 1, \cdots, 6$, is in the interval $(-2\log n, 2\log n)$. Then, $P(J_n) = 1 - o(n^2)$, which is proved by Anderson (1974). We see that each element of $\mathbf{e}^{(j)}$ and $t^{(j)}$ is a homogeneous polynomial of degree $j + 1$ in the elements of $\boldsymbol{x}_i$ and $\boldsymbol{w}_i$. The remainder terms of (A.15) and (B.6) are of the order $O(n^{-1})$ uniformly in J_n. Therefore, the analysis subsequent to (C.3) in Fujikoshi et al (1982) is applicable.

C. Derivation of Theorem 3

Anderson and Kunitomo (2006) showed that the limiting distribution of $\boldsymbol{e}^{(0)}$ under the sequence (2.13) is $N(\mathbf{0}, \boldsymbol{\Psi}^\dagger)$ when the disturbances are followed by the elliptically contoured distribution. Moreover, we notice that $E[\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}'\boldsymbol{w}_{2i}u_i] = \mathbf{0}$ and $E[\boldsymbol{w}_{2i}\boldsymbol{w}_{2i}'\boldsymbol{w}_{2i}{u_i}^3] = \mathbf{0}$ in this case. Using these facts, we can calculate, in the same manner as that in the case of normal disturbances, that

$$x_i = \frac{\boldsymbol{i}'\tilde{\boldsymbol{e}}^{(0)}}{\sqrt{\Psi_{ii}^\dagger}} \tag{C.1}$$

$$t^{(1)} = \frac{\boldsymbol{i}'\boldsymbol{e}^{(1)}}{\sqrt{\Psi_{ii}^\dagger}} - \frac{1}{2}\frac{{\Psi^{\dagger(1)}}_{ii}}{\Psi_{ii}^\dagger}x_i, \tag{C.2}$$

where

$$
\begin{aligned}
\boldsymbol{\Psi}^{\dagger(1)} &= \boldsymbol{Q}^{-1}[-2\boldsymbol{e}^{(0)\prime}\boldsymbol{q}_2\sigma^2 + \sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{P}_Z\boldsymbol{u} - \sigma^2)] \\
&\quad -\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1}\sigma^2 \\
&\quad +\{c_1(1+c_2)+\kappa\eta\}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1} \\
&\quad -\{c_1(1+c_2)+\kappa\eta\}\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1}\left[\begin{pmatrix} \sigma^2\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1} \\
&\quad -\{c_1(1+c_2)+\kappa\eta\}\mathbf{Q}^{-1}\left[\begin{pmatrix} \sigma^2\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}^{-1} \\
&\quad +\kappa^*\eta\mathbf{Q}^{-1}\left[\begin{pmatrix} \sigma^2\boldsymbol{\Omega}_{22} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} - \boldsymbol{q}_2\boldsymbol{q}_2'\sigma^4\right]\boldsymbol{Q}^{-1}, \\
\kappa^* &= \frac{1}{3\sigma^2\sqrt{n}}[-\frac{4}{n}\sum_{i=1}^n u_i^3\boldsymbol{w}_{2i}'\boldsymbol{e}^{(0)} - \frac{4}{n}\sum_{i=1}^n u_i^3\boldsymbol{z}_i'\boldsymbol{D}_2\boldsymbol{e}^{(0)} \\
&\quad +\sqrt{n}(\frac{1}{n}\sum_{i=1}^n u_i^4 - E(u_i^4)) - \frac{2E(u_i^4)}{\sigma^2}\sqrt{\frac{c_2}{c_1}}\sqrt{q}(\frac{1}{q}\boldsymbol{u}'\bar{\boldsymbol{P}}_Z\boldsymbol{u} - \sigma^2)],
\end{aligned}
$$

and the first order term x_i is distributed as the standard normal random variable. Using *Lemma 4.3* given in Kunitomo and Matsushita (2005) and the fact that any odd moments of the elliptically contoured distribution is 0, the expectation of $t^{(1)}$ conditional upon x_i is calculated as

$$
E(t^{(1)}|x_i) = \frac{1}{\sqrt{\Psi_{ii}^{\dagger}}}(\boldsymbol{i}'\boldsymbol{\Psi}^{\dagger}\boldsymbol{q}_2)x_i^2 + O_p(n^{-1/2}). \qquad \text{(C.3)}
$$

Hence, we can derive an asymptotic expansion of the density function of the large K t-ratio by inverting the characteristic function of t^* up to $O(n^{-1/2})$, which can be written as

$$
C(t) = E[exp(itx_i)] + \frac{1}{\sqrt{n}}E[itE(t^{(1)}|x_i)exp(itx_i)] + O(n^{-1}). \qquad \text{(C.4)}
$$

Since x_i is asymptotically normal, we can invert this characteristic function following the same discussion as that in Section 4 in Kunitomo and Matsushita (2005). Then, we obtain the result.

References

[1] Anderson, T.W. (1974), "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System," *Journal of the American Statistical Association*, Vol. 69, 565-573.

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

[3] 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.

[4] 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.

[5] 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.

[6] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "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)

[7] Anderson, T.W., and N. Kunitomo (2006), "On Asymptotic Properties of the LIML Estimator with possibly Many Instruments," Preprint

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

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

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

[11] 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.

[12] 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.

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

[14] Hansen, C., J. Hausman, and W.K.Newey (2006), "Estimation with Many Instrumental Variables," Unpublished Manuscript.

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

[16] Kleibergen, F. (2005), "Testing Parameters in GMM without assuming that they are identified", *Econometrica*, Vol.73, 1103-1124.

[17] 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.

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

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

[19] 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)

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

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

[22] 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.

[23] Morimune, K. (1989), "t Test in a Structural Equation," *Econometrica*, Vol.57-6, 1341-1360.

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