Likelihood Ratio Tests in Reduced Rank Regression with Applications to Econometric Structural Equations *

T. W. Anderson † and Naoto Kunitomo ‡

April 1, 2009

Abstract

Likelihood ratio criteria are developed for hypotheses concerning multivariate regression matrices when reduced rank is assumed or equivalently hypotheses concerning linear restrictions on regression matrices. In an econometric simultaneous equation model a single restriction may be called a "structural equation." A model of reduced rank in statistics is often termed a "linear functional relationship." In this paper the likelihood ratio test that the regression matrix satisfies some specified restriction is developed under the assumption of normality. The test for the corresponding hypothesis that the regression matrix of given rank is a specified matrix is also developed. The asymptotic distribution of the test criterion is found under several alternative assumptions on the sequence of models. The "cointegration model" is included in this study. The test for one structural equation is an advancement on the test statistic proposed by Anderson and Rubin (1949 and 1950).

Key Words

Structural Equation, Likelihood Ratio Criterion, Anderson-Rubin test, Weak Instruments, Many Instruments, Reduced Rank Regression, Cointegration.

*This is TWA09-4-1, which is a revised version of Discussion Paper CIRJE-F-499 under the title "On Likelihood Ratio Tests of Structural Coefficients : Anderson-Rubin (1949) Revisited" (Graduate School of Economics, University of Tokyo, May 2007). The authors thank Cheng Hsiao for useful suggestions in the preparation of this paper.

†Department of Statistics and Department of Economics, Stanford University

‡Graduate School of Economics, University of Tokyo

1. Introduction

There are many problems in multivariate statistical analysis that involve tests concerning regressions of reduced rank and of restrictions on regressions. Tests of the general linear hypothesis are well-developed as tests of the rank of a regression matrix. An example of the problem considered in this paper occurs in the area of simultaneous equations in econometrics. Consider a demand and supply model. The dependent (endogenous) variables are the price of a good and the quantity exchanged in a market; the independent (exogenous) variables may consist of other variables of consumers and (manufacturing) producers. The demand function (a structural equation) depends on the income of consumer, but not on the manufacturing variables such as inventory levels. We shall often use the language of econometrics because it furnishes an important and familiar application.

A "reduced rank regression" model can be described in terms of the space spanned by the columns (or by some columns) of expected values of the vector of dependent variables given the independent variables. The same model can be put in terms of the linear restrictions on the regression coefficients. We shall develop the tests in terms of a single linear restriction and then obtain the general theorems. Limiting distributions of the test statistics are developed for the number of observations increasing.

We first develop a likelihood ratio test for a hypothesis about the coefficients of one structural equation in a set of simultaneous equations. The null hypothesis is that the vector of coefficients is a specified vector; the alternative hypothesis is that the structural equation is "identified," that is, that some vector of coefficients satisfies the rank or dimensionality condition. The limiting distribution of -2 times the logarithm of the likelihood ratio criterion as the number of observations increases is often chi-square with degrees of freedom equal to one less than the number of coefficients specified in the null hypothesis.

The asymptotic distribution theory is valid under various assumptions about the model including so-called "weak instruments". *Weak Instruments* means a regression model in which the regression coefficients decrease in size as the number of observations increases. These alternative models permit the choice of asymptotic distributions as an approximation to the exact distribution.

The problem of testing a null hypothesis on the coefficients of the structural equation has been studied by many econometricians since Anderson and Rubin (1949). See Andrews, Moreira and Stock (2006) for a recent review of these studies. The present problem is broadly related to several testing problems in reduced rank regression models, the errors-in-variables models and cointegration models, usually treated separately. We shall explore these relations in a unified way. Statistical problem concerning "reduced rank regression" moldes have been studied in the statistical literature by Anderson (1951, 2003), Reinsel and Valu (1998), and others.

In Section 2 we define the statistical model. Then we give a new derivation of the likelihood ratio test in Section 3 and give some results on its asymptotic distribution under a set of general conditions including some cases of weak instruments and many instruments in Section 4. The extensions of our approach to the reduced rank regression models and the cointegration models are discussed in Sections 5 and 6. Concluding remarks are given in Section 7. The proofs of theorems are in Section 8.

2. The statistical models

The observed data consist of a $T \times G$ matrix of endogenous or dependent variables $\mathbf{Y}$ and a $T \times K$ matrix of exogenous or independent variables $\mathbf{Z}$. A linear model (the reduced form) is

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V} \ , \tag{2.1}$$

where $\mathbf{\Pi}$ is a $K \times G$ matrix of parameters and $\mathbf{V}$ is a $T \times G$ matrix of unobservable disturbances. The rows of $\mathbf{V}$ are assumed independent; each row has a normal distribution $N(\mathbf{0}, \mathbf{\Omega})$. The coefficient matrix $\mathbf{\Pi}$ is estimated by the sample regression matrix

$$\mathbf{P} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \ . \tag{2.2}$$

The covariance matrix $\mathbf{\Omega}$ is estimated by $(1/T)\mathbf{H}$, where

$$\mathbf{H} = (\mathbf{Y} - \mathbf{Z}\mathbf{P})'(\mathbf{Y} - \mathbf{Z}\mathbf{P}) = \mathbf{Y}'\mathbf{Y} - \mathbf{P}'\mathbf{A}\mathbf{P} \ , \tag{2.3}$$

and $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$. The matrices $\mathbf{P}$ and $\mathbf{H}$ constitute a sufficient set of statistics for the model.

A structural or behavioral equation may involve a subset of the endogenous variables, say $\mathbf{Y}_1$, $T \times G_1$, a subset of exogenous variables, say $\mathbf{Z}_1$, $T \times K_1$, and a subset of disturbances, say $\mathbf{V}_1$, $T \times G_1$. The equation of interest is written as

$$\mathbf{Y}_1\boldsymbol{\beta} = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{u} \ , \tag{2.4}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of G_1 and K_1 parameters, respectively, $\mathbf{u} = \mathbf{V}_1\boldsymbol{\beta}$ and $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$; a component of $\mathbf{u}$ has the normal distribution $N(0, \sigma^2)$, where $\sigma^2 = \boldsymbol{\beta}'\mathbf{\Omega}_{11}\boldsymbol{\beta}$ and $\mathbf{\Omega}_{11}$ is the $G_1 \times G_1$ upper-left corner of $\mathbf{\Omega}$ such that

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix} \ .$$

Let $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{V}$ and $\mathbf{\Pi}$ be partitioned accordingly so that the reduced form (2.1) is

$$(\mathbf{Y}_1, \mathbf{Y}_2) = (\mathbf{Z}_1, \mathbf{Z}_2) \begin{bmatrix} \mathbf{\Pi}_{11} & \mathbf{\Pi}_{12} \\ \mathbf{\Pi}_{21} & \mathbf{\Pi}_{22} \end{bmatrix} + (\mathbf{V}_1, \mathbf{V}_2) \ , \tag{2.5}$$

where $\mathbf{Z}_2$ is a $T \times K_2$ matrix. The relation between the structural equation (2.4) and the reduced form (2.5) is

$$\begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11}\boldsymbol{\beta} \\ \boldsymbol{\Pi}_{21}\boldsymbol{\beta} \end{bmatrix} . \tag{2.6}$$

The second part of (2.6),

$$\boldsymbol{\Pi}_{21}\boldsymbol{\beta} = \mathbf{0} , \tag{2.7}$$

defines $\boldsymbol{\beta}$ except for a multiplicative constant if and only if the rank of $\boldsymbol{\Pi}_{21}$ is $G_1 - 1$. In that case the structural equation is said to be *identified*. Since $\boldsymbol{\Pi}_{21}$ is $K_2 \times G_1$, a necessary condition for identification is $K_2 \geq G_1 - 1$.

Consider the null hypothesis

$$H_0 : \ \boldsymbol{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0} , \tag{2.8}$$

where $\boldsymbol{\beta}_0$ is a (non-zero) specified vector. First we shall find the likelihood ratio criterion for H_0 when the alternative hypothesis, say H_2, consists of arbitrary $\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$.

It will be convenient to transform the model so that the two sets of exogenous variables are orthogonal. Let

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{Z}_1 & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 \end{bmatrix} , \tag{2.9}$$

$$\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12} , \qquad \mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} , \tag{2.10}$$

$$\mathbf{P}_{21} = \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y}_1 , \tag{2.11}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} . \tag{2.12}$$

Define $(\boldsymbol{\Pi}_{11}^*, \boldsymbol{\Pi}_{12}^*) = \left(\mathbf{I}_{K_1}, \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)\boldsymbol{\Pi}$. Then

$$\mathbf{Z}\boldsymbol{\Pi} = (\mathbf{Z}_1, \mathbf{Z}_{2.1}) \begin{bmatrix} \boldsymbol{\Pi}_{11}^* & \boldsymbol{\Pi}_{12}^* \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} = (\mathbf{Z}_1, \mathbf{Z}_{2.1})\,\boldsymbol{\Pi}^* .$$

The matrix $\mathbf{Z}_{2.1}$ has the properties $\mathbf{Z}_1'\mathbf{Z}_{2.1} = \mathbf{O}$ and $\mathbf{Z}_{2.1}'\mathbf{Z}_{2.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{A}_{22.1}$. Define also

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_{2.1}' \end{bmatrix} [\mathbf{Z}_1, \mathbf{Z}_{2.1}] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22.1} \end{bmatrix} .$$

In terms of $(\mathbf{Z}_1, \mathbf{Z}_{2.1})$, the sample regression matrix is

$$\mathbf{P}^* = (\mathbf{A}^*)^{-1} \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_{2.1}' \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{A}_{11}^{-1}\mathbf{Z}_1'\mathbf{Y} \\ \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1^* \\ \mathbf{P}_2^* \end{bmatrix}$$

and

$$\mathbf{H} = \mathbf{Y}'\mathbf{Y} - \mathbf{P}^{*'}\mathbf{A}^*\mathbf{P}^* .$$

3. The likelihood ratio test of a specified vector of structural coefficients given that the equation is identified

The null hypothesis $H_0 : \ \mathbf{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$ is relevant if the equation is "identified", that is, if the rank of $\mathbf{\Pi}_{21}$ is $G_1 - 1$. We develop the likelihood ratio criterion for testing H_0 against the alternative hypothesis $H_1 :$ rank $(\mathbf{\Pi}_{21}) = G_1 - 1$ by finding the likelihood function maximized under H_0 and dividing by the likelihood function maximized under H_1. The likelihood function is

$$\begin{aligned}
(3.1) \qquad & L(\mathbf{\Pi}, \mathbf{\Omega}) \\
= \ & (2\pi)^{-\frac{1}{2}TG}|\mathbf{\Omega}|^{-\frac{1}{2}T} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{Y} - \mathbf{Z\Pi})'(\mathbf{Y} - \mathbf{Z\Pi})\mathbf{\Omega}^{-1}\} \\
= \ & (2\pi)^{-\frac{1}{2}TG}|\mathbf{\Omega}|^{-\frac{1}{2}T} \exp\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{P} - \mathbf{\Pi})'\mathbf{A}(\mathbf{P} - \mathbf{\Pi}) + \mathbf{H}\right]\mathbf{\Omega}^{-1}\} \\
= \ & (2\pi)^{-\frac{1}{2}TG}|\mathbf{\Omega}|^{-\frac{1}{2}T} \exp\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{P}^* - \mathbf{\Pi}^*)'\mathbf{A}^*(\mathbf{P}^* - \mathbf{\Pi}^*) + \mathbf{H}\right]\mathbf{\Omega}^{-1}\} \\
= \ & (2\pi)^{-\frac{1}{2}TG}|\mathbf{\Omega}|^{-\frac{1}{2}T} \exp\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{P}_1^* - \mathbf{\Pi}_1^*)'\mathbf{A}_{11}(\mathbf{P}_1^* - \mathbf{\Pi}_1^*)\right. \\
& \left. +(\mathbf{P}_2^* - \mathbf{\Pi}_2)'\mathbf{A}_{22.1}(\mathbf{P}_2^* - \mathbf{\Pi}_2) + \mathbf{H}\right]\mathbf{\Omega}^{-1}\} ,
\end{aligned}$$

where $\mathbf{\Pi}_1^* = (\mathbf{\Pi}_{11}^*, \mathbf{\Pi}_{12}^*)$ and $\mathbf{\Pi}_2 = (\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22})$. The maximum of $L(\mathbf{\Pi}, \mathbf{\Omega})$ with respect to $\mathbf{\Pi}_1^*$ occurs at $\mathbf{\Pi}_1^* = \mathbf{P}_1^*$ and is

$$L(\mathbf{\Pi}_2, \mathbf{\Omega}) = (2\pi)^{-\frac{1}{2}TG} |\mathbf{\Omega}|^{-\frac{1}{2}T} \exp\{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{P}_2^* - \mathbf{\Pi}_2)'\mathbf{A}_{22.1}(\mathbf{P}_2^* - \mathbf{\Pi}_2) + \mathbf{H}\right]\mathbf{\Omega}^{-1}\} . \tag{3.2}$$

The maximum of $L(\mathbf{\Pi}_2, \mathbf{\Omega})$ with respect to $\mathbf{\Omega}$ is

$$L(\mathbf{\Pi}_2) = (2\pi)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} \left|(\mathbf{P}_2^* - \mathbf{\Pi}_2)'\mathbf{A}_{22.1}(\mathbf{P}_2^* - \mathbf{\Pi}_2) + \mathbf{H}\right|^{-\frac{1}{2}T} e^{-\frac{1}{2}TG} . \tag{3.3}$$

By *Lemma 3* of Section 8, the maximum of $L(\mathbf{\Pi}_2)$ with respect to $\mathbf{\Pi}_{22}$ is

$$\begin{aligned} & L(\mathbf{\Pi}_{21}) \\ = \ & (2\pi)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} \left|(\mathbf{P}_{21} - \mathbf{\Pi}_{21})'\mathbf{A}_{22.1}(\mathbf{P}_{21} - \mathbf{\Pi}_{21}) + \mathbf{H}_{11}\right|^{-\frac{1}{2}T} |\mathbf{H}_{22.1}|^{-\frac{1}{2}T} e^{-\frac{1}{2}TG} , \end{aligned} \tag{3.4}$$

where $\mathbf{P}_{21} = \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}'\mathbf{Y}_1$, $\mathbf{H}_{22.1} = \mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^{-1}\mathbf{H}_{12}$, $\mathbf{H}_{11}$ is a $G_1 \times G_1$ submatrix, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} .$$

Then the maximum of $L(\mathbf{\Pi}_{21})$ with respect to $\mathbf{\Pi}_{21}$ is

$$\mathrm{L}_{H_2} = (2\pi)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} e^{-\frac{1}{2}TG} . \tag{3.5}$$

This is the likelihood maximized with respect to $\mathbf{\Pi}$ and $\mathbf{\Omega}$ with no rank restriction on the coefficients.

Now consider maximizing the likelihood function under the condition $\mathbf{H}_1 :\ \mathrm{rank}(\mathbf{\Pi}_{21}) = G_1 - 1$, that is, $\mathbf{\Pi}_{21}\boldsymbol{\beta} = \mathbf{0}$ for some $\boldsymbol{\beta}$. The matrix $\mathbf{\Pi}_{21}$ of rank $G_1 - 1$ can be parameterized as

$$\mathbf{\Pi}_{21} = \boldsymbol{\mu}\mathbf{\Gamma}' , \tag{3.6}$$

where $\boldsymbol{\mu}$ is $K_2 \times (G_1 - 1)$ of rank $G_1 - 1$ and $\mathbf{\Gamma}$ is $G_1 \times (G_1 - 1)$ of rank $G_1 - 1$ such that

$$\mathbf{\Gamma}'\boldsymbol{\beta} = \mathbf{0} . \tag{3.7}$$

Lemma 1 : The minimum of

$$
\left|(\mathbf{P}_{21}-\boldsymbol{\mu}\boldsymbol{\Gamma}^{'})^{'}\mathbf{A}_{22.1}(\mathbf{P}_{21}-\boldsymbol{\mu}\boldsymbol{\Gamma}^{'})+\mathbf{H}_{11}\right| \tag{3.8}
$$

with respect to $\boldsymbol{\mu}$ is

$$
|\mathbf{H}_{11}|\left[1+\frac{\boldsymbol{\beta}^{'}\mathbf{G}_{11}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{H}_{11}\boldsymbol{\beta}}\right] , \tag{3.9}
$$

where

$$
\mathbf{G}_{11}=\mathbf{P}_{21}^{'}\mathbf{A}_{22.1}\mathbf{P}_{21} . \tag{3.10}
$$

Proof : The determinant (3.8) is

$$
|\mathbf{H}_{11}|\left|(\boldsymbol{\mu}\boldsymbol{\Gamma}^{'}-\mathbf{P}_{21})\mathbf{H}_{11}^{-1}(\boldsymbol{\Gamma}\boldsymbol{\mu}^{'}-\mathbf{P}_{21}^{'})\mathbf{A}_{22.1}+\mathbf{I}_{K_2}\right| , \tag{3.11}
$$

which is minimized at

$$
\hat{\boldsymbol{\mu}}=\mathbf{P}_{21}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1} . \tag{3.12}
$$

The determinant (3.8) is then

$$
\begin{aligned}
&\Big|[\mathbf{P}_{21}-\mathbf{P}_{21}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}]^{'}\mathbf{A}_{22.1} \\
&\qquad\times[\mathbf{P}_{21}-\mathbf{P}_{21}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}]+\mathbf{H}_{11}\Big| \\
=\ &\left|[\mathbf{I}_{G_1}-\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}]\mathbf{G}_{11}[\mathbf{I}_{G_1}-\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}]+\mathbf{H}_{11}\right| \\
=\ &|\mathbf{H}_{11}|\Big|\left[\mathbf{I}_{G_1}-\mathbf{H}_{11}^{-1/2}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1/2}\right](\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2}) \\
&\qquad\times\left[\mathbf{I}_{G_1}-\mathbf{H}_{11}^{-1/2}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{'}\mathbf{H}_{11}^{-1/2}\right]+\mathbf{I}_{G_1}\Big| \\
=\ &|\mathbf{H}_{11}|\left|\left[\mathbf{I}_{G_1}-\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}\right]\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2}\left[\mathbf{I}_{G_1}-\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}\right]+\mathbf{I}_{G_1}\right| ,
\end{aligned} \tag{3.13}
$$

where $\mathbf{Q}=\mathbf{H}_{11}^{-1/2}\boldsymbol{\Gamma}$. The matrix $\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}$ is idempotent of rank G_1-1 and $\mathbf{I}_{G_1}-\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}$ is idempotent of rank $G_1-(G_1-1)=1$. Then $\mathbf{I}_{G_1}-\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}=$

$\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$ and $\mathbf{Q}'\mathbf{x} = \mathbf{0}$ for $\mathbf{x} = \mathbf{H}_{11}^{1/2}\boldsymbol{\beta}$. Then (3.13) is

$$
\begin{aligned}
(3.14) \qquad & |\mathbf{H}_{11}| \left| \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' + \mathbf{I}_{G_1} \right| \\
= \; & |\mathbf{H}_{11}| \left[1 + \frac{\boldsymbol{\beta}'\mathbf{G}_{11}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{H}_{11}\boldsymbol{\beta}} \right]
\end{aligned}
$$

by Corollary A.3.1 of Anderson (2003). **Q.E.D.**

The likelihood maximized over $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is

$$
(3.15) \qquad \mathrm{L}_{H_0} = (2\pi e)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} \left[1 + \frac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0} \right]^{-\frac{1}{2}T} .
$$

The likelihood maximized over $\boldsymbol{\beta}$ is

$$
\begin{aligned}
(3.16) \qquad \mathrm{L}_{H_1} &= \max_{\boldsymbol{\beta}} (2\pi e)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} \left[1 + \frac{\boldsymbol{\beta}'\mathbf{G}_{11}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{H}_{11}\boldsymbol{\beta}} \right]^{-\frac{1}{2}T} \\
&= (2\pi e)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} \left[1 + \min_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta}'\mathbf{G}_{11}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{H}_{11}\boldsymbol{\beta}} \right]^{-\frac{1}{2}T} \\
&= (2\pi e)^{-\frac{1}{2}TG} T^{\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} \left[1 + \frac{\hat{\boldsymbol{\beta}}'\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}} \right]^{-\frac{1}{2}T} ,
\end{aligned}
$$

where $\hat{\boldsymbol{\beta}}$ satisfies

$$
(3.17) \qquad \mathbf{G}_{11}\hat{\boldsymbol{\beta}} = \nu_1 \mathbf{H}_{11}\hat{\boldsymbol{\beta}}
$$

and ν_1 as the smallest root of

$$
(3.18) \qquad |\mathbf{G}_{11} - \nu \mathbf{H}_{11}| = 0 \; ;
$$

that is, $\hat{\boldsymbol{\beta}}$ is the Limited Information Maximum Likelihood Estimator of $\boldsymbol{\beta}$.

Theorem 1: The likelihood ratio criterion for testing the null hypothesis $H_0 : \mathbf{\Pi}_{21}$ has rank $G_1 - 1$ and $\mathbf{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$ vs. $H_1 : \mathbf{\Pi}_{21}$ has rank $G_1 - 1$ is

$$
(3.19) \qquad \frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \left[\frac{1 + \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}}}{1 + \dfrac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}} \right]^{\frac{1}{2}T} = \left[\frac{1 + \min_{\mathbf{b}} \dfrac{\mathbf{b}'\mathbf{G}_{11}\mathbf{b}}{\mathbf{b}'\mathbf{H}_{11}\mathbf{b}}}{1 + \dfrac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}} \right]^{\frac{1}{2}T}
$$

$$= \left[\frac{1+\nu_1}{1+\dfrac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}} \right]^{\frac{1}{2}T} .$$

The null hypothesis H_0 that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ is rejected if (3.19) is less than a suitable constant; that is, if

$$\frac{1+\dfrac{\hat{\boldsymbol{\beta}}'\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}}}{1+\dfrac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}} < c(K_2, T-K) . \tag{3.20}$$

We call the left-hand side of (3.20) the *Rank-Adjusted Anderson-Rubin* (RAAR) criterion.

Comments :

1. The RAAR criterion does not depend on a normalization of the vector of coefficients. The ratio $\boldsymbol{\beta}_0'\mathbf{P}_{21}'\mathbf{A}_{22.1}\mathbf{P}_{21}\boldsymbol{\beta}_0/\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0$ is unchanged by replacing $\boldsymbol{\beta}_0$ by $\boldsymbol{\beta}_0$ times an arbitrary constant. Similarly, $\hat{\boldsymbol{\beta}}'\mathbf{P}_{21}'\mathbf{A}_{22.1}\mathbf{P}_{21}\hat{\boldsymbol{\beta}}/\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}$ is unchanged by replacing the LIML estimator multiplied by a constant. The normalization of $\boldsymbol{\beta}_0$ does not have to be the same as of $\hat{\boldsymbol{\beta}}$.

2. The RAAR criterion compares the hypothesized $\boldsymbol{\beta}_0$ with the LIML estimator $\hat{\boldsymbol{\beta}}$.

3. The RAAR criterion is a function of the sufficient statistics $\mathbf{P}$ and $\mathbf{H}$.

4. The RAAR criterion is invariant with respect to linear transformations $\mathbf{Y}_1 \to \mathbf{Y}_1\mathbf{C}$, $\boldsymbol{\beta}_0 \to \mathbf{C}^{-1}\boldsymbol{\beta}_0$ and $\mathbf{Z}_2 \to \mathbf{Z}_2\mathbf{D}$ for $\mathbf{C}$ and $\mathbf{D}$ nonsingular.
The only invariants of $\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0/\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\beta}}'\mathbf{G}_{11}\hat{\boldsymbol{\beta}}/\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}$ are $\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0/\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0$ and the roots of (3.18).

5. The logarithm of the criterion (3.19) is

$$(3.21) \qquad -2 \ \log \frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = -\mathrm{T} \left[\log \left(1 + \frac{\hat{\boldsymbol{\beta}}' \mathbf{G}_{11} \hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}' \mathbf{H}_{11} \hat{\boldsymbol{\beta}}} \right) - \log \left(1 + \frac{\boldsymbol{\beta}_0' \mathbf{G}_{11} \boldsymbol{\beta}_0}{\boldsymbol{\beta}_0' \mathbf{H}_{11} \boldsymbol{\beta}_0} \right) \right] ,$$

which is approximately

$$(3.22) \qquad \frac{\boldsymbol{\beta}_0' \mathbf{G}_{11} \boldsymbol{\beta}_0}{\boldsymbol{\beta}_0' \frac{1}{T} \mathbf{H}_{11} \boldsymbol{\beta}_0} - \min_{\mathbf{b}} \frac{\mathbf{b}' \mathbf{G}_{11} \mathbf{b}}{\mathbf{b}' \frac{1}{T} \mathbf{H}_{11} \mathbf{b}} .$$

6. The RAAR is a likelihood ratio criterion. In many statistical inference problems concerning normal distributions a likelihood ratio test has optimum properties.

Moreira (2003) arrived at a statistic similar to (3.22) by a somewhat different route. He calls the crtiterion a *conditional likelihood statistic*.

4. Limiting Distributions

4.1 The Standard Case

The likelihood ratio criterion for testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ vs. $H_1 : \mathrm{rank}(\boldsymbol{\Pi}_{21}) = G_1 - 1$ has been derived on the basis of the rows of $\mathbf{V}$ being independently distributed according to $N(\mathbf{0}, \boldsymbol{\Omega})$. For the limiting distribution of -2 times the logarithm of (3.19) we assume that the rows of $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_T)'$ satisfy

$$(\mathbf{I}) \qquad \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_t \mathbf{z}_t' \xrightarrow{p} \mathbf{M} \ \ (\text{as } T \to \infty) ,$$

$$(\mathbf{II}) \qquad \frac{1}{T} \max_{1 \le t \le T} \|\mathbf{z}_t\|^2 \xrightarrow{p} 0$$

as $T \to \infty$. The limiting distribution of -2 times (3.19) holds under relaxed conditions on $\mathbf{Z}$ and $\mathbf{V}$. The conditions allow for components of $\mathbf{z}_t$ being components of $\mathbf{y}_{t-1}, \cdots, \mathbf{y}_1$ and components of $\mathbf{v}_t$ depending on $\mathbf{z}_t, \cdots, \mathbf{z}_0$. Let the $\sigma-$field $\mathcal{F}_t$ be generated by $\mathbf{z}_0, \mathbf{v}_0, \cdots, \mathbf{z}_t, \mathbf{v}_t,\ t = 1, \cdots, T$, and $\mathcal{F}_0$ is the initial $\sigma-$field generated by $\mathbf{y}_1$. (See Anderson and Kunitomo (1992), for instance.) Partition the $(G_1 + G_2)$-vector $\mathbf{v}_t^{'} = (\mathbf{v}_{1t}^{'}, \mathbf{v}_{2t}^{'})$ $(t = 1, \cdots, T)$. We assume that $\mathcal{E}(\mathbf{v}_t | \mathcal{F}_t) = \mathbf{0}$ *a.s.*, $\mathcal{E}(\mathbf{v}_t \mathbf{v}_t^{'} | \mathcal{F}_{t-1}) = \mathbf{\Omega}_t$ *a.s.*, and $\mathbf{\Omega}_t$ is a function of $\mathbf{z}_1, \mathbf{v}_1, \cdots, \mathbf{z}_{t-1}, \mathbf{v}_{t-1}, \mathbf{z}_t$. Since $u_t = \mathbf{v}_{1t}^{'} \boldsymbol{\beta}$, we have $\mathcal{E}(u_t | \mathcal{F}_t) = \mathbf{0}$ *a.s.* and $\mathcal{E}(u_t^2 | \mathcal{F}_t) = \sigma_t^2 = \boldsymbol{\beta}^{'} \mathbf{\Omega}_{11}^{(t)} \boldsymbol{\beta}$ *a.s.*, where $\mathbf{\Omega}_t$ is a $(G_1 + G_2) \times (G_1 + G_2)$ matrix

$$\mathbf{\Omega}_t = \begin{bmatrix} \mathbf{\Omega}_{11}^{(t)} & \mathbf{\Omega}_{12}^{(t)} \\ \mathbf{\Omega}_{21}^{(t)} & \mathbf{\Omega}_{22}^{(t)} \end{bmatrix} .$$

Suppose

$$\textbf{(III)} \qquad \frac{1}{T} \sum_{t=1}^{T} \mathbf{\Omega}_{11}^{(t)} \otimes \mathbf{z}_t \mathbf{z}_t^{'} \xrightarrow{p} \mathbf{\Omega}_{11} \otimes \mathbf{M} \ \ (\text{as } T \to \infty) ,$$

$$\textbf{(IV)} \qquad \frac{1}{T} \sum_{t=1}^{T} \mathbf{\Omega}_{11}^{(t)} \xrightarrow{p} \mathbf{\Omega}_{11} \ \ (\text{as } T \to \infty) ,$$

$$\textbf{(V)} \qquad \sup_{t \geq 1} \mathcal{E}[\mathbf{v}_{1t}^{'} \mathbf{v}_{1t} I(\mathbf{v}_{1t}^{'} \mathbf{v}_{1t} > c) | \mathcal{F}_t] \xrightarrow{p} 0 \ \ (\text{as } c \to \infty) ,$$

where $I(\cdot)$ is the indicator function, and $\mathbf{M}$ and $\mathbf{\Omega}_{11}$ are nonsingular (constant) matrices. Conditions (IV) and (V) imply

$$\text{(4.1)} \qquad \frac{1}{T} \sum_{t=1}^{T} \mathbf{v}_{1t} \mathbf{v}_{1t}^{'} \xrightarrow{p} \mathbf{\Omega}_{11} \ \ (\text{as } T \to \infty)$$

and $\sigma^2 = \boldsymbol{\beta}^{'} \mathbf{\Omega}_{11} \boldsymbol{\beta}$ (> 0).

Comments :

1. We allow some heteroscedasticity of disturbances and only require second-order moments. Thus the conditions on disturbances are minimal.

2. The conditions (I) and (II) on instruments include the situations that the lagged endogenous variables are subsets of instruments when they follow a stationary autoregressive process, for instance.

Although the RAAR statistic is invariant with respect to normalization, we shall find it convenient to normalize $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$ so $\boldsymbol{\beta}_0 = (1, -\boldsymbol{\beta}_2^{0'})'$ and $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$.

In order to investigate the limiting null distribution and the local power of the LRC, we consider a sequence of local alternatives $\boldsymbol{\Pi}^{(T)}$ such that

$$(4.2) \qquad \begin{bmatrix} \boldsymbol{\Pi}_{11}^{(T)} & \boldsymbol{\Pi}_{12}^{(T)} \\ \boldsymbol{\Pi}_{21}^{(T)} & \boldsymbol{\Pi}_{22}^{(T)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} + \frac{1}{\sqrt{T}} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} ,$$

where $\boldsymbol{\xi}_i$ is a $K_i \times 1$ vector $(i = 1, 2)$, each element of the $(K_1+K_2)\times(G_1+G_2)$ matrix $\boldsymbol{\Pi}$ is a function of T (say $\boldsymbol{\Pi}_T$) partitioned as $\boldsymbol{\Pi}_T = (\boldsymbol{\Pi}_{ij}^{(T)})$ and rank$(\boldsymbol{\Pi}_{21}) = G_1 - 1$. Hence $\lim_{T\to\infty} \boldsymbol{\Pi}_{21}^{(T)} = \boldsymbol{\Pi}_{21}$ and $\boldsymbol{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$ as the limit $(T \to \infty)$ in (2.8). (See (2.6) and (2.7) in Section 2.) Then Theorem 2 is an extension of Theorem 4 of Anderson and Kunitomo (1994). The proof is given in Section 8.

Theorem 2 : Assume Conditions (I)-(V). Under the local alternative sequences (4.2), as $T \to \infty$ the limiting distribution of

$$(4.3) \quad LR_1 = -2\log \frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \mathrm{T}\left[\log\left(1 + \frac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}\right) - \log\left(1 + \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{G}_{11}\mathbf{b}}{\mathbf{b}'\mathbf{H}_{11}\mathbf{b}}\right)\right]$$

is noncentral χ^2 with $G_1 - 1$ degrees of freedom and the noncentrality parameter $\kappa_1 = \theta_1\sigma^{-2}$, where $\sigma^2 = \boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0$, $\mathbf{M}_{22.1} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$,

$$(4.4) \qquad \theta_1 = \boldsymbol{\xi}_2'\mathbf{M}_{22.1}\boldsymbol{\Pi}_2^*(\boldsymbol{\Pi}_2^{*'}\mathbf{M}_{22.1}\boldsymbol{\Pi}_2^*)^{-1}\boldsymbol{\Pi}_2^{*'}\mathbf{M}_{22.1}\boldsymbol{\xi}_2 ,$$

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} , \ \boldsymbol{\Pi}_2^* = \boldsymbol{\Pi}_{21} \begin{bmatrix} \mathbf{0}' \\ \mathbf{I}_{G_1-1} \end{bmatrix} ,$$

in which we assume that $\boldsymbol{\Pi}_2^*$ has rank $G_1 - 1$.

Let $\boldsymbol{\Pi}_{21} = (\boldsymbol{\pi}_1, \boldsymbol{\Pi}_2^*)$. Then (2.7) is

$$(4.5) \qquad \mathbf{0} = \boldsymbol{\Pi}_{21}\boldsymbol{\beta} = (\boldsymbol{\pi}_1, \boldsymbol{\Pi}_2^*)\begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{\pi}_1 - \boldsymbol{\Pi}_2^*\boldsymbol{\beta}_2 .$$

Thus when $\boldsymbol{\beta}$ is normalized as $(1, -\boldsymbol{\beta}_2)$ and $\mathbf{\Pi}$ has rank $G_1 - 1$, $\mathbf{\Pi}_2^*$ must have rank $G_1 - 1$ to solve (4.5) for $\boldsymbol{\beta}_2$. The noncentrality parameter θ_1 is invariant with respect to multiplying $\mathbf{\Pi}_2^*$ on the right by an arbitrary nonsingular $(G_1-1)\times(G_1-1)$ matrix.

Under H_0 ($\boldsymbol{\xi} = \mathbf{0}$), the limiting distribution of LR_1 is χ^2 with $G_1 - 1$ degrees of freedom under the general conditions on the disturbances. Then by using the χ^2 distribution in Theorem 2 when T is large in (2.18), we can take the rejection region as

$$\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} > [1+\nu_1]\, e^{\frac{1}{T}\chi^2_{G_1-1}(\epsilon)} - 1 \tag{4.6}$$

by using $\chi^2(\epsilon)$ with $G_1 - 1$ degrees of freedom. It is also possible to investigate the power function under the local alternative hypotheses of (4.2).

4.2 Weak Instruments

Next, we consider a case of so-called *weak* instruments. Let $\mathbf{\Pi}_T = \mathbf{C}/T^{\delta}$ for a constant matrix $\mathbf{C}$ and $\delta > 0$; as T grows the regression matrix $\mathbf{\Pi}_T$ becomes smaller. The $(K_1 + K_2) \times (G_1 + G_2)$ matrices $\mathbf{\Pi}_T = (\mathbf{\Pi}_{ij}^{(T)})$ and $\mathbf{C} = (\mathbf{C}_{ij})$ are partitioned as $\mathbf{\Pi}_T$. Then Condition (I) implies

$$(\mathbf{I}^{'}) \qquad \frac{1}{T^{1-2\delta}} \sum_{t=1}^{T} \mathbf{\Pi}_T^{'}\mathbf{z}_t\mathbf{z}_t^{'}\mathbf{\Pi}_T \xrightarrow{p} \mathbf{C}^{'}\mathbf{M}\mathbf{C} \quad (\text{as } T \to \infty)\,.$$

We rewrite (2.1), (2.12) and (3.10) as

$$(\mathbf{Y}_1^{(T)}, \mathbf{Y}_2^{(T)}) = \left(\mathbf{Z}_1^{(T)}, \mathbf{Z}_2^{(T)}\right) \begin{bmatrix} \mathbf{\Pi}_{11}^{(T)} & \mathbf{\Pi}_{12}^{(T)} \\ \mathbf{\Pi}_{21}^{(T)} & \mathbf{\Pi}_{22}^{(T)} \end{bmatrix} + \left(\mathbf{V}_1^{(T)}, \mathbf{V}_2^{(T)}\right)\,, \tag{4.7}$$

$$\mathbf{G}_{11}^{(T)} = \mathbf{P}_{21}^{(T)'}\mathbf{A}_{22.1}\mathbf{P}_{21}^{(T)}\,,$$

and

$$\mathbf{H}_{11}^{(T)} = \mathbf{Y}_1^{(T)'}\mathbf{Y}_1^{(T)} - \mathbf{P}_{21}^{(T)'}\mathbf{A}_{22.1}\mathbf{P}_{21}^{(T)}\,.$$

Define $\nu_1^{(T)}$ as the smallest root of $\left|\mathbf{G}_{11}^{(T)} - \lambda^{(T)}\mathbf{H}_{11}^{(T)}\right| = 0$; that is,

$$\nu_1^{(T)} = \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{G}_{11}^{(T)}\mathbf{b}}{\mathbf{b}'\mathbf{H}_{11}^{(T)}\mathbf{b}} = \frac{\hat{\boldsymbol{\beta}}^{(T)'}\mathbf{G}_{11}^{(T)}\hat{\boldsymbol{\beta}}^{(T)}}{\hat{\boldsymbol{\beta}}^{(T)'}\mathbf{H}_{11}^{(T)}\hat{\boldsymbol{\beta}}^{(T)}} , \tag{4.8}$$

and the LIML estimator $\hat{\boldsymbol{\beta}}^{(T)} = (1, -\hat{\boldsymbol{\beta}}_2^{(T)'})'$ satisfying $\left(\mathbf{G}_{11}^{(T)} - \nu_1^{(T)}\mathbf{H}_{11}^{(T)}\right)\hat{\boldsymbol{\beta}}^{(T)} = \mathbf{0}$.

The weak instruments case is different from the standard situation for (2.1) and (2.4). The limiting distribution of LR_1 depends on the weakness of instruments, which can be measured by the parameter δ. Theorem 3 states the limiting distribution of the LR statistic when $0 < \delta < 1/2$. The proof is similar to Theorem 1 and is omitted.

Theorem 3 : Assume $\boldsymbol{\Pi}_T = \mathbf{C}/T^{\delta}$ for a (constant) $K \times G$ matrix $\mathbf{C}$ with $0 < \delta < 1/2$ and Conditions (I) – (V). Under a local alternative sequence (4.2) as $T \to \infty$ the limiting distribution of LR_1 is noncentral χ^2 with $G_1 - 1$ degrees of freedom and the noncentrality parameter $\kappa_2 = \theta_2\sigma^{-2}$, where $\sigma^2 = \boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0$,

$$\theta_2 = \boldsymbol{\xi}_2'\mathbf{M}_{22.1}\mathbf{C}_2^{*}\left[\mathbf{C}_2^{*'}\mathbf{M}_{22.1}\mathbf{C}_2^{*}\right]^{-1}\mathbf{C}_2^{*'}\mathbf{M}_{22.1}\boldsymbol{\xi}_2 \tag{4.9}$$

and

$$\mathbf{C}_2^{*} = \mathbf{C}_{21}\begin{bmatrix} \mathbf{0}' \\ \mathbf{I}_{G_1-1} \end{bmatrix} \tag{4.10}$$

has rank $G_1 - 1$.

If

$$\begin{bmatrix} \boldsymbol{\Pi}_{11}^{(T)} & \boldsymbol{\Pi}_{12}^{(T)} \\ \boldsymbol{\Pi}_{21}^{(T)} & \boldsymbol{\Pi}_{22}^{(T)} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} + \frac{1}{\mathrm{T}^{\eta}}\begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \tag{4.11}$$

and $\eta > 1/2$ then the statistic LR_1 has the limiting distribution of central χ^2 with $G_1 - 1$ degrees of freedom.

Note that the noncentrality parameter in Theorem 3 is the same as in Theorem 2 except $\boldsymbol{\Pi}_T = (1/T)^{\delta}\mathbf{C}$. A large value of δ corresponds to a small value $\boldsymbol{\Pi}_T$ (for fixed

C); hence a relatively small value of θ_2. The limiting distribution of LR_1 under the null hypothesis in Theorem 3 is the same as in Theorem 2.

Staiger and Stock (1997) considered statistical inference for the model (2.1) and (2.2) with *weak instruments* defined by $\mathbf{\Pi}_T = \mathbf{C}/T^{1/2}$ for a fixed $\mathbf{C}$. Note that this case of weak instruments is not included in the study here of $\mathbf{\Pi}_T = \mathbf{C}/T^{\delta}$ for $0 < \delta < 1/2$. Staiger and Stock (1997) get nonstandard distributions for various estimators and test statistics.

4.3 Many Weak Instruments

The model of $K_2 \to \infty$ as $T \to \infty$ has been called the case of *many weak* instruments, recently discussed in econometrics. An alternative formulation of *many weak* instruments is to let each element, as well as the size, of $\mathbf{\Pi}$ be a function of T. We denote K_{2T} and K_T for K_2 and K, respectively. In this model we denote a sequence of $K_T \times G$ $(K_T = K_1 + K_{2T}, T \geq 3)$ matrices $\mathbf{\Pi}_T$, each matrix is partitioned into $(K_1 + K_{2T}) \times (G_1 + G_2)$ submatrices

$$\mathbf{\Pi}_T = \begin{bmatrix} \mathbf{\Pi}_{11}^{(T)} & \mathbf{\Pi}_{12}^{(T)} \\ \mathbf{\Pi}_{21}^{(T)} & \mathbf{\Pi}_{22}^{(T)} \end{bmatrix} .$$

Suppose

$$(\mathbf{VI}) \qquad \frac{K_T}{T} \longrightarrow 0 .$$

Instead of Conditions (I)-(III), we suppose the conditions

$$(\mathbf{I}'') \qquad \frac{1}{T}\sum_{t=1}^{T} \mathbf{\Pi}_T' \mathbf{z}_t^{(T)} \mathbf{z}_t^{(T)'} \mathbf{\Pi}_T \stackrel{p}{\longrightarrow} \mathbf{\Phi} \ \ (\text{as } T \to \infty) ,$$

$$(\mathbf{II}'') \qquad \frac{1}{T} \max_{1 \leq t \leq T} \|\mathbf{\Pi}_T' \mathbf{z}_t^{(T)}\|^2 \stackrel{p}{\longrightarrow} 0 \ \ (\text{as } T \to \infty) ,$$

$$(\mathbf{III}'') \qquad \frac{1}{T}\sum_{t=1}^{T} \mathbf{\Omega}_{11}^{(t)} \otimes \mathbf{\Pi}_T' \mathbf{z}_t^{(T)} \mathbf{z}_t^{(T)'} \mathbf{\Pi}_T \stackrel{p}{\longrightarrow} \mathbf{\Omega}_{11} \otimes \mathbf{\Phi} \ \ (\text{as } T \to \infty) ,$$

where $\mathbf{\Omega}_{11}$ is a positive definite constant matrix, $\mathbf{\Phi}$ is a nonnegative definite constant matrix (the upper-left $G_1 \times G_1$ sub-matrix of $\mathbf{\Phi}$ is of rank $G_1 - 1$), and $\mathbf{z}_t^{(T)}$ is a $K_T \times 1$ vector of instruments.

In this model, there can be alternative assumptions about the relative magnitudes of T, K_T and $\mathbf{\Pi}_T$. The condition (VI) is a necessary and sufficient condition for the next result [1]. The many weak instruments case is different from the standard situation for (2.1) and (2.4) with fixed K and K_2. We have omitted the proof of Theorem 4 because it is similar to those of Theorem 2.

Theorem 4 : Let $\mathbf{z}_t^{(T)}, t = 1, \cdots, T$, be a sequence of $K_T \times 1$ vectors of instruments. For a sequence of $K_T \times G$ coefficient matrices $\mathbf{\Pi}_T$, $T = K+1, \cdots$, assume Conditions (I)$''$-(III)$''$, (IV)-(V) and (VI). Under the local alternative sequence

$$\mathbf{\Pi}_T \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} + \frac{1}{\sqrt{T}} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_{2T} \end{bmatrix} , \tag{4.12}$$

as $T \to \infty$ the statistic LR_1 has the limiting distribution of noncentral χ^2 with $G_1 - 1$ degrees of freedom and the noncentrality parameter $\kappa_3 = \theta_3 \sigma^{-2}$, provided that the probability limits of

$$\theta_3 = \left[\text{plim} \frac{1}{T} \boldsymbol{\xi}_{2T}' \mathbf{A}_{22.1} \mathbf{\Pi}_{2T}\right] \left[\text{plim} \frac{1}{T} \mathbf{\Pi}_{2T}' \mathbf{A}_{22.1} \mathbf{\Pi}_{2T}\right]^{-1} \left[\text{plim} \frac{1}{T} \mathbf{\Pi}_{2T}' \mathbf{A}_{22.1} \boldsymbol{\xi}_{2T}\right] , \tag{4.13}$$

exist and θ_3 is positive for a sequence of the $K_{2T} \times 1$ vectors $\boldsymbol{\xi}_{2T}$, the $K_{2T} \times 1$ sub-vectors $\mathbf{z}_{2t}^{(T)}$ of $\mathbf{z}_t^{(T)}$, a sequence of the $K_{2T} \times K_{2T}$ matrices

$$\mathbf{A}_{22.1} = \sum_{t=1}^{T} \mathbf{z}_{2t}^{(T)} \mathbf{z}_{2t}^{(T)\prime} - \sum_{t=1}^{T} \mathbf{z}_{2t}^{(T)} \mathbf{z}_{1t}' \left[\sum_{t=1}^{T} \mathbf{z}_{1t} \mathbf{z}_{1t}'\right]^{-1} \sum_{t=1}^{T} \mathbf{z}_{1t} \mathbf{z}_{2t}^{(T)\prime} ,$$

and a sequence of $K_{2T} \times (G_1 - 1)$ matrices

$$\mathbf{\Pi}_{2T} = \mathbf{\Pi}_{21}^{(T)} \begin{bmatrix} \mathbf{0}' \\ \mathbf{I}_{G_1 - 1} \end{bmatrix} .$$

[1] Recently, Matsushita (2007) has investigated the finite sample distribution of LR_1 without Condition (VI). The related problem on estimation with many instruments has been explored by Anderson, Kunitomo and Matsushita (2005, 2008), for instance.

Thus we find that rejection regions and confidence regions based on the χ^2 distribution with $G_1 - 1$ degrees of freedom are asymptotically valid for some cases of weak instruments including some many weak instruments. The assumptions of Theorem 3 on weak instruments (with $\boldsymbol{\xi}_2 = \mathbf{C}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$) or Theorem 4 (with $\theta_3 = 0$) on many instruments are sufficient for χ^2 with $G_1 - 1$ degrees of freedom as the asymptotic null distribution.

5. Blocks of Structural Equations and Reduced Rank Regression

The likelihood ratio criterion we have developed can be extended to more general multivariate models such as the errors-in-variable model and the linear functional relationship model. See Anderson (1976), Anderson (1984), and Anderson and Kunitomo (2007).

A block of structural equations is

$$\mathbf{Y}_1\mathbf{B} = \mathbf{Z}_1\boldsymbol{\Gamma}_1 + \mathbf{U} \ , \tag{5.1}$$

where $\mathbf{B}$ is $G_1 \times r$ of rank r, $\boldsymbol{\Gamma}_1$ is $K_1 \times r$ and $\mathbf{U}$ is $T \times r$. Corresponding to (2.6),

$$\begin{bmatrix} \boldsymbol{\Gamma}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B} \\ \mathbf{O} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11}\mathbf{B} \\ \boldsymbol{\Pi}_{21}\mathbf{B} \end{bmatrix} . \tag{5.2}$$

The second part of the above equation $\boldsymbol{\Pi}_{21}\mathbf{B} = \mathbf{O}$ has a solution for $\mathbf{B}$ that is unique except for multiplication on the right by a nonsingular $r \times r$ matrix if and only if the rank of $\boldsymbol{\Pi}_{21}$ is $G_1 - r$. We say the block of equations (5.1) is *block-identified.*

Consider the null hypothesis

$$H_0^{'} : \ \boldsymbol{\Pi}_{21}\mathbf{B}_0 = \mathbf{O} \ , \tag{5.3}$$

where $\mathbf{B}_0$ is a specified $G_1 \times r$ matrix of rank r ($1 \leq r < G_1$). When the set of alternatives $H_2^{'}$ includes all $K_2 \times G_1$ matrices $\mathbf{\Pi}_{21}$, the likelihood ratio criterion is

$$\text{(5.4)} \qquad \frac{\mathrm{L}_{H_0^{'}}}{\mathrm{L}_{H_2^{'}}} = \left[\frac{\left|\mathbf{B}_0^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}_0\right|}{\left|\mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0\right|} \right]^{-T/2} .$$

(See Anderson (2003), Section 8.3.)

The likelihood ratio criterion for testing the null hypothesis

$$\text{(5.5)} \qquad H_1^{'}: \ \ \mathrm{rank}(\mathbf{\Pi}_{21}) = G_1 - r$$

against the alternatives $H_2^{'}$ is

$$\text{(5.6)} \qquad \frac{\mathrm{L}_{H_1{'}}}{\mathrm{L}_{H_2{'}}} = \left[\min_{\mathbf{B}} \frac{\left|\mathbf{B}^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}\right|}{|\mathbf{B}^{'}\mathbf{H}_{11}\mathbf{B}|} \right]^{-T/2} = \prod_{i=1}^{r}(1+\nu_i)^{-T/2} ,$$

where $\nu_1, \cdots, \nu_r$ are the r smallest roots of (3.18) and $\nu_1 \leq \nu_1 \leq \cdots \leq \nu_{G_1}$. (See Anderson (1951), Theorem 2.)

The likelihood ratio criterion for testing $H_0^{'}$ against $H_1^{'}$ is

$$\text{(5.7)} \qquad \frac{\mathrm{L}_{H_0{'}}}{\mathrm{L}_{H_1{'}}} = \left[\frac{\min_{\mathbf{B}} \frac{\left|\mathbf{B}^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}\right|}{|\mathbf{B}^{'}\mathbf{H}_{11}\mathbf{B}|}}{\frac{\left|\mathbf{B}_0^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}_0\right|}{\left|\mathbf{B}^{'}\mathbf{H}_{11}\mathbf{B}\right|}} \right]^{T/2} = \left[\frac{\prod_{i=1}^{r}(1+\nu_i)}{\frac{\left|\mathbf{B}_0^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}_0\right|}{\left|\mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0\right|}} \right]^{T/2} .$$

This is the likelihood ratio criterion for testing $H_0^{'}: \mathbf{B} = \mathbf{B}_0$ given that $\mathbf{B}$ is block-identified.

The criterion (5.7) can also be derived by the method of obtaining Theorem 1. Parametrize $\mathbf{\Pi}_{21}$ as

$$\text{(5.8)} \qquad \mathbf{\Pi}_{21} = \boldsymbol{\mu}\mathbf{\Gamma}^{'} ,$$

where $\boldsymbol{\mu}$ is $K_2 \times (G_1 - r)$ of rank $G_1 - r$ and $\boldsymbol{\Gamma}$ is $G_1 \times (G_1 - r)$ of rank $G_1 - r$. Then $\boldsymbol{\Gamma}'\mathbf{B} = \mathbf{O}$ for a $G_1 \times r$ matrix $\mathbf{B}$ if and only if $\boldsymbol{\Pi}_{21}\mathbf{B} = \mathbf{O}$. The proof of Lemma 1 proceeds with $\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ being idempotent of rank $G_1 - r$, and $\mathbf{I}_{G_1} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ being idempotent of rank $G_1 - (G_1 - r) = r$. Then $\mathbf{I}_{G_1} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{Q}'\mathbf{X} = \mathbf{0}$ for $\mathbf{X} = \mathbf{H}_{11}^{1/2}\mathbf{B}$. Hence

$$
\begin{aligned}
(5.9) \qquad & \left|(\mathbf{P}_{21} - \boldsymbol{\mu}\boldsymbol{\Gamma}')'\mathbf{A}_{22.1}(\mathbf{P}_{21} - \boldsymbol{\mu}\boldsymbol{\Gamma}') + \mathbf{H}_{11}\right| \\
&= |\mathbf{H}_{11}||\mathbf{I}_r + \mathbf{B}'\mathbf{G}_{11}\mathbf{B}(\mathbf{B}'\mathbf{H}_{11}\mathbf{B})^{-1}| \ .
\end{aligned}
$$

Then the maximum of (5.9) for $\mathbf{B} = \mathbf{B}_0$ divided by the maximum of (5.9) with respect to $\mathbf{B}$ is (5.7).

Define

$$
\begin{aligned}
(5.10) \qquad LR_2 &= -2\log \frac{\mathrm{L}_{H_0'}}{\mathrm{L}_{H_1'}} \\
&= T\log \frac{\left|\mathbf{B}_0'(\mathbf{G}_{11} + \mathbf{H}_{11})\mathbf{B}_0\right|}{\left|\mathbf{B}_0'\mathbf{H}_{11}\mathbf{B}_0\right|} - T\sum_{i=1}^{r}\log(1+\nu_i) \ .
\end{aligned}
$$

When the null hypothesis H_0 is true, the first term on the right-hand side has a limiting χ^2-distribution with K_2 degrees of freedom; the second term has a limiting χ^2-distribution with $[K_2 - (G_1 - r)]r$ degrees of freedom. These facts suggest that the limiting distribution of LR_1 is the χ^2-distribution with $K_2 r - [K_2 - (G_1 - r)]r = (G_1 - r)r$ degrees of freedom. This assertion was proved in Anderson and Kunitomo (2007) for $r = 1$ and in Section 8 by another method (Theorem 2).

The regression matrix $\boldsymbol{\Pi}_{21}$ is said to be of reduced rank if $\boldsymbol{\Pi}_{21} = \boldsymbol{\mu}\boldsymbol{\Gamma}'$, where the ranks of $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ are lower than K_2 and G_1. In such a case the columns of $\boldsymbol{\Pi}_{21}'$ lie in a $(G_1 - r)-$dimensional subspaces of the G_1-space, which is spanned by the columns of $\boldsymbol{\Gamma}$. This space is orthogonal to the $r-$dimensional subspace spanned by the columns of $\mathbf{B}$ satisfying $\boldsymbol{\Gamma}'\mathbf{B} = \mathbf{O}$. The $\boldsymbol{\Gamma}-$space is equivalently spanned by the columns of $\boldsymbol{\Gamma}\mathbf{A}$, where $\mathbf{A}$ is $(G_1 - r) \times (G_1 - r)$ of rank $G_1 - r$; the $\mathbf{B}-$space is equivalently spanned by the columns of $\mathbf{BC}$, where $\mathbf{C}$ is $r \times r$ matrix of rank r.

Now consider the null hypothesis

$$H_0^{''}: \quad \mathbf{\Gamma} = \mathbf{\Gamma}_0 \ , \tag{5.11}$$

where $\mathbf{\Gamma}_0$ is a specified $G_1 \times (G_1 - r)$ matrix of rank $G_1 - r$. There exists a $G_1 \times r$ matrix $\mathbf{B}_0$ of rank r such that

$$\mathbf{\Gamma}_0^{'}\mathbf{B}_0 = \mathbf{O} \ . \tag{5.12}$$

Then the likelihood maximized with respect to $\mathbf{\Gamma}$ at $\mathbf{\Gamma} = \mathbf{\Gamma}_0$ is exactly the likelihood maximized with respect to $\mathbf{B}$ at $\mathbf{B} = \mathbf{B}_0$. Thus (5.7) is $\mathrm{L}_{H_0^{''}}/\mathrm{L}_{H_1^{'}}$. The likelihood ratio test of $H_0^{''}: \mathbf{\Gamma} = \mathbf{\Gamma}_0$ is also the likelihood ratio test of $H_0^{'}: \mathbf{B} = \mathbf{B}_0$ as long as $\mathbf{\Gamma}_0^{'}\mathbf{B}_0 = \mathbf{O}$. Note that $\mathbf{B}_0$ can be replaced by $\mathbf{B}_0\mathbf{C}$, where $\mathbf{C}$ is $r \times r$ of rank r.

Lemma 2 : If

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{B}_2 \end{bmatrix} , \ \mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_2 \\ \mathbf{I}_{G_1 - r} \end{bmatrix} , \tag{5.13}$$

and

$$\mathbf{\Gamma}^{'}\mathbf{B} = \mathbf{\Gamma}_2^{'} + \mathbf{B}_2 = \mathbf{O} \ , \tag{5.14}$$

then

$$|\mathbf{\Gamma}^{'}\mathbf{H}^{-1}\mathbf{\Gamma}| = \frac{|\mathbf{B}^{'}\mathbf{H}\mathbf{B}|}{|\mathbf{H}|} \tag{5.15}$$

for $\mathbf{H}$ positive definite.

Proof of Lemma 2 : Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \tag{5.16}$$

and use the partitioned formula for $\mathbf{H}^{-1}$ of Theorem A-3-3 in Anderson (2003). Then

$$\mathbf{B}^{'}\mathbf{H}\mathbf{B} = \mathbf{H}_{11.2} + (\mathbf{H}_{21}\mathbf{H}_{22}^{-1} + \mathbf{B}_2^{'})\mathbf{H}_{22}(\mathbf{H}_{22}^{-1}\mathbf{H}_{12} + \mathbf{B}_2) \tag{5.17}$$

$$\mathbf{\Gamma}^{'}\mathbf{H}^{-1}\mathbf{\Gamma} = \mathbf{H}_{22}^{-1} + (\mathbf{H}_{22}^{-1}\mathbf{H}_{21} + \mathbf{B}_2)\mathbf{H}_{11.2}^{-1}(\mathbf{H}_{12}\mathbf{H}_{22}^{-1} + \mathbf{B}_2^{'}) \ , \tag{5.18}$$

where $\mathbf{H}_{11.2} = \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21}$.

For any $r \times (G_1 - r)$ matrix $\mathbf{C}$ we have

$$|\mathbf{C}\mathbf{C}^{'} + \mathbf{I}_r| = |\mathbf{C}^{'}\mathbf{C} + \mathbf{I}_{G_1^r}| \ . \tag{5.19}$$

By using this relation, we find

$$\begin{aligned} |\mathbf{\Gamma}^{'}\mathbf{H}^{-1}\mathbf{\Gamma}| &= |\mathbf{H}_{22}^{-1}| \times |\mathbf{I}_{G_1 - r} + \mathbf{H}_{22}^{1/2}(\mathbf{H}_{22}^{-1}\mathbf{H}_{21} + \mathbf{B}_2)\mathbf{H}_{11.2}^{-1}(\mathbf{H}_{12}\mathbf{H}_{22}^{-1} + \mathbf{B}_2^{'})\mathbf{H}_{22}^{1/2}| \ , \\ &= |\mathbf{H}^{-1}| \times |\mathbf{H}_{11.2} + (\mathbf{H}_{21}\mathbf{H}_{22}^{-1} + \mathbf{B}_2^{'})\mathbf{H}_{22}(\mathbf{H}_{22}^{-1}\mathbf{H}_{12} + \mathbf{B}_2)| \ , \end{aligned}$$

which is the result. **Q.E.D.**

The likelihood ratio criterion for testing the hypothesis that $\mathbf{\Gamma} = \mathbf{\Gamma}_0$, where

$$\mathbf{\Pi}_{21} = \boldsymbol{\mu}\mathbf{\Gamma}^{'} \tag{5.20}$$

is

$$\begin{aligned} & \left[\frac{\min_{\mathbf{\Gamma}} \frac{\left|\mathbf{\Gamma}^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}\right|}{\left|\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right|}}{\frac{\left|\mathbf{\Gamma}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}_0\right|}{\left|\mathbf{\Gamma}_0^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}_0\right|}} \right]^{T/2} \\ = & \left[\frac{\prod_{i=1}^{r}(1 + \nu_i)}{\frac{\left|\mathbf{\Gamma}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}_0\right|}{\left|\mathbf{\Gamma}_0^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}_0\right|} \times \frac{|\mathbf{G}_{11} + \mathbf{H}_{11}|}{|\mathbf{H}_{11}|}} \right]^{T/2} . \end{aligned} \tag{5.21}$$

6. Cointegration

The "cointegration" problem in econometrics can be formulated in terms of the reduced rank regression in Section 5. It is a multivariate time series model with stationary components and (a nonstationary) random walk components. The main interest is to make statistical inferences on linear relationships. (See Johansen (1995) and Anderson (2000), for instance.)

Let a $G_1 \times 1$ autoregressive model $\{\mathbf{x}_t\}$ be defined by

$$
\begin{aligned}
\Delta \mathbf{x}_t &= \left[\mathbf{\Pi}'(1), \cdots, \mathbf{\Pi}'(p)\right] \begin{bmatrix} \Delta \mathbf{x}_{t-1} \\ \vdots \\ \Delta \mathbf{x}_{t-p} \end{bmatrix} + \mathbf{\Pi}_2' \mathbf{x}_{t-1} + \mathbf{v}_t \\
&= \mathbf{\Pi}_1' \mathbf{z}_{1t} + \mathbf{\Pi}_2' \mathbf{z}_{2t} + \mathbf{v}_t \ ,
\end{aligned} \tag{6.1}
$$

where $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, $\mathbf{\Pi}_1' = (\mathbf{\Pi}'(1), \cdots, \mathbf{\Pi}'(p))$ and $\mathbf{\Pi}_2'$ are $G_1 \times G_1 p$ and $G_1 \times G_1$ matrices of coefficients, $\mathbf{z}_{1t}' = (\Delta \mathbf{x}_{t-1}', \cdots, \Delta \mathbf{x}_{t-p}')$, $\mathbf{z}_{2t}' = \mathbf{x}_{t-1}'$ and $\mathcal{E}(\mathbf{v}_t \mathbf{v}_t') = \mathbf{\Omega}$. The model (6.1) is of the form of the reduced form (2.1) with $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$ and $\mathbf{\Pi}' = (\mathbf{\Pi}_1', \mathbf{\Pi}_2')$ with $G = G_1$. The $t-$th row of $\mathbf{Y}$ is $\Delta \mathbf{x}_t'$; the $t-$th row of $\mathbf{Z}_2$ is $\mathbf{x}_{t-1}'$, $t = 1, \cdots, T$.

Suppose that $\mathbf{\Pi}_2$ is of rank $G_1 - r$ and hence can be written as $\mathbf{\Pi}_2 = \boldsymbol{\mu} \mathbf{\Gamma}'$, where $\boldsymbol{\mu}$ is $G_1 \times (G_1 - r)$ and $\mathbf{\Gamma}$ is $G_1 \times (G_1 - r)$. Note that $\boldsymbol{\mu}$ can be multiplied on the right by an arbitrary nonsingular matrix and $\mathbf{\Gamma}'$ on the left by the inverse of that arbitrary matrix. (The matrix $\mathbf{\Gamma} \boldsymbol{\mu}'$ is $\boldsymbol{\alpha} \boldsymbol{\beta}'$ in Johansen's notation.)

The likelihood ratio criterion for testing $\mathbf{\Gamma} = \mathbf{\Gamma}_0$ against alternatives rank $\mathbf{\Gamma} = G_1 - r$ is the same as the likelihood ratio criterion for testing $\mathbf{\Pi}_2 \mathbf{B}_0 = \mathbf{O}$ when $\mathbf{B}_0$ is a $G_1 \times r$ matrix satisfying $\mathbf{\Gamma}_0' \mathbf{B}_0 = \mathbf{O}$.

$$
LR_3 = T \log \left[\frac{\left| \mathbf{\Gamma}_0' (\mathbf{G}_{11} + \mathbf{H}_{11})^{-1} \mathbf{\Gamma}_0 \right|}{\left| \mathbf{\Gamma}_0' \mathbf{H}_{11}^{-1} \mathbf{\Gamma}_0 \right|} \right] - T \sum_{i=1}^{G_1 - r} \log \xi_i \ , \tag{6.2}
$$

where $\xi_{G_1+1-i} = 1/(1+\nu_i)$ $(i = 1, \cdots, G_1)$ are the characteristic roots $(\xi_1 \leq \xi_2 \leq \cdots \leq \xi_{G_1})$ of

$$
\left| (\mathbf{G}_{11} + \mathbf{H}_{11})^{-1} - \xi \mathbf{H}_{11}^{-1} \right| = 0 \ . \tag{6.3}
$$

In the cointegration case instead of Conditions (I)-(III) in Section 4, we assume the condition [2] that all characteristic roots of

$$\text{(VII)} \qquad \left|(\lambda-1)\lambda^p \mathbf{I}_G - \lambda^p \mathbf{\Pi}_2^{'} - (\lambda-1)\sum_{i=1}^{p} \lambda^{p-i}\mathbf{\Pi}_1^{'}(i)\right| = 0$$

are in the range $(-1,1]$ or their absolute values are in the range $[0,1)$.

Theorem 5 : Assume that $\{\mathbf{v}_t\}$ is a sequence of i.i.d. random vectors with $\mathcal{E}(\mathbf{v}_t) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_t\mathbf{v}_t^{'}) = \mathbf{\Omega}$, and Condition (VII). Then under the rank condition $H_0^{'}$: rank$(\mathbf{\Pi}_2) = G_1 - r$ and $\mathbf{\Gamma} = \mathbf{\Gamma}_0$, as $T \to \infty$ LR_3 has the limiting distribution of χ^2 with $r(G_1 - r)$ degrees of freedom.

The resulting test procedure and confidence region are invariant to orthogonal transformations of $\mathbf{\Gamma}_0$ (i.e. cointegrating vectors) and they are direct extensions of Section 4 to the cointegration problem.

There are some applications of *Weak Instruments* and *Many Weak Instruments* in Section 4 to the cointegration problem.

7. Concluding remarks

This paper has shed a new light on the classical problem of the likelihood ratio tests of structural coefficients in a structural equation in the simultaneous equation system. The method developed by Anderson and Rubin (1949, 1950) can be modified to the situation when there are many (or weak in some sense) instruments which may have some relevance in recent econometrics. We have found that the asymptotic null distribution of LRC is often the χ^2-distribution with G_1-1 degrees of freedom under a set of fairly general conditions.

Then we have shown that the testing problems in the structural equation (simultaneous equations) model, the reduced rank regression and the cointegration models

[2] It is sufficient that $\Delta\mathbf{x}_t$ is stationary and $\mathbf{x}_t$ is an $I(1)-$process.

are essentially the same. Furthermore, the testing problems in the linear functional relationships or the errors-in-variables models are also mathematically the same to those in the reduced regression problem, which are related to the testing problems in factor models. (See Anderson (1984).) Since these statistical models have been used in many applications, it is worthwhile and useful to show that the problems can be indeed formulated as direct extensions of the classical method by Anderson and Rubin for a single structural equation model [3] in an unified fashion.

This paper is written in terms of testing a null hypothesis $\mathbf{B} = \mathbf{B}_0$ or $\mathbf{\Gamma} = \mathbf{\Gamma}_0$. Any of resulting test procedure can be inverted to obtain a confidence region of $\mathbf{B}$ or $\mathbf{\Gamma}$; that is, a confidence region for $\mathbf{B}$ consists of all $\mathbf{B}_0$ not rejected by the test.

8. Mathematical Details

In this section we give some technical details which were omitted in the previous sections. At the last part of this section, we shall refer to Anderson and Kunitomo (1994) as AK (1994) and use their method for Theorem 4. Also we shall use the notation of projection operators $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'$. (These matrices are idempotent.)

Lemma 3 : Let a $p \times p$ nonsingular matrix $\mathbf{D}$ be decomposed into $(p_1+p_2)\times(p_1+p_2)$ submatrices $\mathbf{D} = (\mathbf{D}_{ij})$ and $\mathbf{D}^{-1} = (\mathbf{D}^{ij})$. For any $q \times p_1$ matrix $\mathbf{B}$, $q \times p_2$ matrix $\mathbf{C}$ and any positive definite matrix $\mathbf{A}$,

$$(8.1) \qquad \min_{\mathbf{C}} \left| \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}\,(\mathbf{B}, \mathbf{C}) + \mathbf{D} \right| = \left|\mathbf{D}_{22} - \mathbf{D}_{21}\mathbf{D}_{11}^{-1}\mathbf{D}_{12}\right| \left|\mathbf{D}_{11} + \mathbf{B}'\mathbf{A}\mathbf{B}\right|$$

and the minimum occurs at $\mathbf{C} = -\mathbf{B}\mathbf{D}_{11}^{-1}\mathbf{D}_{12}$.

[3] Some results on the corresponding estimation problems have been investigated by Anderson, Kunitomo and Matsushita (2005, 2008).

Proof of Lemma 3: For $|\mathbf{D}| \neq 0$ and $\mathbf{A} > 0$,

$$
(8.2) \quad \left| \mathbf{D} + \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}\,(\mathbf{B}, \mathbf{C}) \right| = \left| \begin{matrix} \mathbf{D} & -\begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}^{1/2} \\ \mathbf{A}^{1/2}\,(\mathbf{B}, \mathbf{C}) & \mathbf{I}_q \end{matrix} \right|
$$

$$
= |\mathbf{D}| \left| \mathbf{I}_q + \mathbf{A}^{1/2}\,(\mathbf{B}, \mathbf{C})\,\mathbf{D}^{-1} \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}^{1/2} \right| \ .
$$

Also we have

$$
\mathbf{A}^{1/2}\,(\mathbf{B}, \mathbf{C})\,\mathbf{D}^{-1} \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}^{1/2}
$$

$$
= \mathbf{A}^{1/2} \left[\mathbf{C} + \mathbf{B}\mathbf{D}^{12}(\mathbf{D}^{22})^{-1} \right] \mathbf{D}^{22} \left[\mathbf{C} + \mathbf{B}\mathbf{D}^{12}(\mathbf{D}^{22})^{-1} \right]' \mathbf{A}^{1/2} \geq \mathbf{A}^{1/2}\mathbf{B}\mathbf{D}^{22}\mathbf{B}'\mathbf{A}^{1/2} \ .
$$

Then

$$
(8.3) \quad \left| \mathbf{D} + \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}\,(\mathbf{B}, \mathbf{C}) \right| \geq |\mathbf{D}| \left| \mathbf{I}_q + \mathbf{A}^{1/2}\mathbf{B}\mathbf{D}^{22}\mathbf{B}'\mathbf{A}^{1/2} \right|
$$

$$
= \frac{|\mathbf{D}|}{|\mathbf{D}_{11}|} \left| \mathbf{D}_{11} + \mathbf{B}'\mathbf{A}\mathbf{B} \right| \ ,
$$

which is the right-hand side of (8.1).

Q.E.D

In order to prove Theorem 2, we first prove two lemmas. (Similar arguments can be used for the proofs of Theorem 3 and Theorem 5.)

Lemma 4 : Under the assumptions of Theorem 2, for any $0 \leq \epsilon < 1$

$$
(8.4) \quad T^{\epsilon}\nu_1 \xrightarrow{p} 0 \ .
$$

Proof of Lemma 4 : It is immediate to see that $(1/T)\mathbf{H}_{11} \xrightarrow{p} \mathbf{\Omega}_{11}$ and

$$
\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0'\mathbf{V}_1'\mathbf{Z}_{2.1}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\mathbf{V}_1\boldsymbol{\beta}_0 + \frac{2}{\sqrt{T}}\boldsymbol{\beta}_0'\mathbf{V}_1'\mathbf{Z}_{2.1}\boldsymbol{\xi}_2 + \frac{1}{T}\boldsymbol{\xi}_2'\mathbf{Z}_{2.1}'\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\boldsymbol{\xi}_2 \ ;
$$

for every $\boldsymbol{\beta}_0$, then $\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0$ converges to a limiting random variable as $T \to \infty$. Then for $0 \leq \epsilon < 1$,

$$0 \leq T^{\epsilon} \min_{\mathbf{b}} \frac{\mathbf{b}^{'}\mathbf{G}_{11}\mathbf{b}}{\mathbf{b}^{'}\mathbf{H}_{11}\mathbf{b}} \leq \frac{1}{T^{1-\epsilon}} \frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\frac{1}{T}\mathbf{H}_{11}\boldsymbol{\beta}_0} \xrightarrow{p} 0 \ .$$

Q.E.D.

Define

$$LR_d = T\left[\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} - \min_{\mathbf{b}} \frac{\mathbf{b}^{'}\mathbf{G}_{11}\mathbf{b}}{\mathbf{b}^{'}\mathbf{H}_{11}\mathbf{b}}\right] \ . \tag{8.5}$$

Lemma 5 : Under the assumptions of Theorem 2 (as $T \to \infty$)

$$LR_1 - LR_d \xrightarrow{p} 0 \ . \tag{8.6}$$

Proof of Lemma 5 : Taylor's expansion yields

$$|T\log(1+\nu_1) - T\nu_1| \leq \frac{1}{2}\left[T^{1/2}\nu_1\right]^2 \ ,$$

which converges to zero by Lemma 3 as $T \to \infty$.

Q.E.D.

Proof of Theorem 2 : By using Lemma 3, we find that as $T \to \infty$ $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Then $(1/T)\mathbf{G}_{11} \xrightarrow{p} \mathbf{\Pi}_{21}^{'}\mathbf{M}_{22.1}\mathbf{\Pi}_{21}$. By using the fact that $\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 = O_p(1)$ and substituting $\mathbf{\Pi}_{21}^{'}\mathbf{M}_{22.1}\mathbf{\Pi}_{21}$ into the set of equations $[\mathbf{G}_{11} - \nu_1\mathbf{H}_{11}]\,\hat{\boldsymbol{\beta}} = \mathbf{0}$, we have

$$\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 + \mathbf{\Pi}_{21}^{'}\mathbf{M}_{22.1}\mathbf{\Pi}_{21}\left[\begin{array}{c} 0 \\ -\sqrt{T}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \end{array}\right] = o_p(1) \ . \tag{8.7}$$

By multiplying $(0, \mathbf{I}_{G_1-1})$ from the left, we find

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) = \left[(0, \mathbf{I}_{G_1-1})\mathbf{\Pi}_{21}^{'}\mathbf{M}_{22.1}\mathbf{\Pi}_{21}\left(\begin{array}{c} \mathbf{0}^{'} \\ \mathbf{I}_{G_1-1} \end{array}\right)\right]^{-1}(0, \mathbf{I}_{G_1-1})\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 + o_p(1) \ . \tag{8.8}$$

Because $(1/T)\mathbf{H}_{11} = \boldsymbol{\Omega}_{11} + O_p(1/\sqrt{T})$, we rewrite the set of equations $[\mathbf{G}_{11} - \nu_1 \mathbf{H}_{11}]\,\hat{\boldsymbol{\beta}} = \mathbf{0}$ as

$$\mathbf{G}_{11}\boldsymbol{\beta}_0 - T\nu_1\left[\boldsymbol{\Omega}_{11} + O_p(\frac{1}{\sqrt{T}})\right]\boldsymbol{\beta}_0 - \left[\mathbf{G}_{11} - T\nu_1\left(\boldsymbol{\Omega}_{11} + O_p(\frac{1}{\sqrt{T}})\right)\right]\begin{bmatrix} 0 \\ -\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \end{bmatrix} = \mathbf{0}.$$

By multiplying $\boldsymbol{\beta}_0^{'}$ from the left, we find that

$$\text{(8.9)} \qquad \boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0 - T\nu_1\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 - \frac{1}{\sqrt{T}}\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\begin{bmatrix} 0 \\ -\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) \end{bmatrix} = o_p(1)\ .$$

Then by using (8.8) and (8.9) we find that

$$\text{(8.10)} \qquad \boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0 - T\nu_1\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0$$
$$= \frac{1}{\sqrt{T}}\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\begin{bmatrix} \mathbf{0}^{'} \\ \mathbf{I}_{G_1-1} \end{bmatrix}\left[(\mathbf{0}, \mathbf{I}_{G_1-1})\boldsymbol{\Pi}_{21}^{'}\mathbf{M}_{22.1}\boldsymbol{\Pi}_{21}\begin{pmatrix} \mathbf{0}^{'} \\ \mathbf{I}_{G_1-1} \end{pmatrix}\right]^{-1}[\mathbf{0}, \mathbf{I}_{G_1-1}]\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0$$
$$+ o_p(1)\ .$$

The limiting distribution of (8.10) is the limiting distribution of $\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 \times LR_d$ as $T \to \infty$. The local alternatives of Theorem 1 imply

$$\mathbf{Y}_1\boldsymbol{\beta}_0 = \mathbf{Z}_1\left(\boldsymbol{\gamma}_1 + \frac{1}{\sqrt{T}}\boldsymbol{\xi}_1\right) + \mathbf{V}_1\boldsymbol{\beta}_0 + \frac{1}{\sqrt{T}}\mathbf{Z}_2\boldsymbol{\xi}_2$$

and then

$$\text{(8.11)} \quad \frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 = \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\boldsymbol{\Pi}_{21}\boldsymbol{\beta}_0 + \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}_1\boldsymbol{\beta}_0 + o_p(1)$$
$$= \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}_1\boldsymbol{\beta}_0 + \boldsymbol{\Pi}_{21}^{'}\mathbf{M}_{22.1}\boldsymbol{\xi}_2 + o_p(1)\ .$$

By applying the Lindeberg-type Central Limit Theorem (see Anderson and Kunitomo (1992) for instance) to the first term on the right of (8.11) and using (8.10), we have the result.

Q.E.D.

References

[1] Anderson, T.W. (1951), "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics*, Vol. 22, 327-351.

[2] Anderson, T.W. (1976), "Estimation of Linear Functional Relationships : Approximate Distributions and Connections to Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society*, B, Vol. 38, 1-36.

[3] Anderson, T.W. (1984), "Estimating linear statistical relationships," *Annals of Statistics*, Vol. 12, 1-45.

[4] Anderson, T.W. (2000), "The asymptotic distribution of canonical correlations and variates in cointegrated models," *Proceedings of the National Academy of Sciences*, Vol. 97, 7068-7073.

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

[6] Anderson, T.W., and Y. Amemiya (1991), "Testing dimensionality in multivariate analysis of variance," *Statistics and Probability Letters*, Vol. 12, 445-463.

[7] Anderson, T.W., and N. Kunitomo (1992), "Asymptotic Distributions of Regression and Autoregression Coefficients with Martingale Difference Disturbances," *Journal of Multivariate Analysis*, Vol. 40, 221-243.

[8] Anderson, T.W., and N. Kunitomo (1994), "Asymptotic robustness of tests of overidentification and predeterminedness," *Journal of Econometrics*, Vol. 62, 383-414.

[9] Anderson, T.W., and N. Kunitomo (2007), "On Likelihood Ratio Tests of Structural Coefficients : Anderson-Rubin (1949) revisited," Discussion Paper CIRJE-F-499, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007).

[10] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005), "A New Light from Old Wisdoms : Alternative Estimation Method of Simultaneous Equations and Microeconometric Models," Discussion Paper CIRJE-F-321, Graduate School of Economics, University of Tokyo.

[11] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2008), "On Finite Sample Properties of Alternative Estimators of Coefficients in a Structural Equation with Many Instruments," Discussion Paper CIRJE-F-577, Graduate School of Economics, University of Tokyo, forthcoming in *Journal of Econometrics*.

[12] 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.

[13] 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.

[14] Andrews, Donald W.H., Marcelo J. Moreira, and James H. Stock (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, Vol.74, 715-752.

[15] Johansen, S. (1995), *Likelihood-based Inference in Cointegrating Autoregressive Models*, Oxford UP.

[16] Matsushita, Y. (2007), "Approximate Distribution of the Likelihood Ratio Statistic in a Structural Equation with Many Instruments," Discussion Paper CIRJE-F-466, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007).

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

[18] Reinsel, G.C., and R.P. Valu (1998), *Multivariate Reduced-rank Regression*, Springer, New York.

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