CIRJE-F-499

On Likelihood Ratio Tests of Structural Coefficients: Anderson-Rubin (1949) revisited

T. W. Anderson
Stanford Univeristy

Naoto Kunitomo
University of Tokyo

May 2007

On Likelihood Ratio Tests of Structural Coefficients : Anderson-Rubin (1949) revisited *

T. W. Anderson †

and

Naoto Kunitomo ‡

May 28, 2007

Abstract

We develop the likelihood ratio criterion (LRC) for testing the coefficients of a structural equation in a system of simultaneous equations in econometrics. We relate the likelihood ratio criterion to the AR statistic proposed by Anderson and Rubin (1949, 1950), which has been widely known and used in econometrics over the past several decades. The method originally 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. The method of LRC can be extended to the linear functional relationships (or the errors-in-variables) model, the reduced rank regression and the cointegration models.

Key Words

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

---

*This is the version TWA07-5-28.

†Department of Statistics and Department of Economics, Stanford University

‡Graduate School of Economics, University of Tokyo

1. Introduction

In "Estimation of the parameters of a single equation in a complete system of stochastic equations" Anderson and Rubin (1949) gave a confidence region for the coefficients of the endogenous variables in that single equation. Such a confidence region leads to a test of the null hypothesis, say $\mathbf{H}_0$, that the vector of coefficients is a specified vector, say $\boldsymbol{\beta}_0$; the test consists of rejecting the null hypothesis if $\boldsymbol{\beta}_0$ is not included in the confidence region, that is, if

$$\frac{\boldsymbol{\beta}_0^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} > \frac{K_2}{T-K}F_{K_2,T-K}(\epsilon) \ , \tag{1.1}$$

where $\mathbf{P}_{2.1}$ is the regression of the "included" endogenous variables on the K_2 "excluded" exogenous variables, $\mathbf{A}_{22.1}$ is the sample covariance matrix of the "excluded" exogenous variables, $\mathbf{H}_{11}$ is the sample error covariance matrix of $T-K$ degrees of freedom, and $F_{K_2,T-K}(\epsilon)$ is the $1-\epsilon$ significance point of the F-distribution with K_2 and $T-K$ degrees of freedom. This test is a likelihood ratio test of $\mathbf{H}_0$ when the disturbances are normally distributed and the exogenous variables are nonstochastic.

The hypothesis $\mathbf{H}_0$ that the vector of coefficients of the endogenous variables is $\boldsymbol{\beta}_0$ is relevant only if the equation is identified. This fact suggests that the hypothesis $\mathbf{H}_0$ should be tested against the set of alternatives in which the equation is identified, say $\mathbf{H}_1$. The equation in question is identified if the relevant submatrix of the coefficients in the reduced form is of rank G_1-1 where G_1 is the number of coefficients in $\boldsymbol{\beta}_0$. The likelihood ratio test of identification is to reject the hypothesis, say $\mathbf{H}_1$, if

$$\frac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}} = \min_{\mathbf{b}} \frac{\mathbf{b}^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\mathbf{b}}{\mathbf{b}^{'}\mathbf{H}_{11}\mathbf{b}} \tag{1.2}$$

is greater than a constant. Here $\hat{\boldsymbol{\beta}}$ is the Limited Information Maximum Likelihood (LIML) estimator of coefficients of the endogenous variables in the selected structural equation. [Anderson-Rubin (1949)].

The likelihood ratio test that the coefficients vector is $\boldsymbol{\beta}_0$ given that the equation is identified is to reject $\mathbf{H}_0$ if

$$\frac{1+\dfrac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}}}{1+\dfrac{\boldsymbol{\beta}_0^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0}} \tag{1.3}$$

is less than a constant. The ratio $\boldsymbol{\beta}_0^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\boldsymbol{\beta}_0/\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0$ measures the effect of the excluded exogenous variables relative to the error variance of that linear combination. The ratio $\hat{\boldsymbol{\beta}}^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\hat{\boldsymbol{\beta}}/\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}$ is the relative variance of the linear combination on which the excluded exogenous variables have least effect.

The criterion for testing $\mathbf{H}_0$ vs. $\mathbf{H}_2$ has an asymptotic distribution of χ^2 with K_2 degrees of freedom, while the criterion for testing $\mathbf{H}_1$ vs. $\mathbf{H}_2$ has an asymptotic distribution of χ^2 with $K_2-(G_1-1)$ degrees of freedom under the standard regularity conditions. The ratio (1.3) has an asymptotic χ^2- distribution with $K_2-[K_2-(G_1-1)] = G_1-1$ degrees of freedom.

For a recent review of the study of testing of $\mathbf{H}_0$ against $\mathbf{H}_2$, see Andrews and Stock (2005). The shortcoming of the original method of Anderson and Rubin shows up particularly when the number of excluded exogenous variables (instruments) is large. Moreira (2003) developed a *conditional* likelihood test when the error covariance matrix is known. It was derived by a different approach and has a form slightly different from (1.3), to which we will mention at the end of Section 3.

In Section 2 we define the statistical model and a new (and simple) derivation of the likelihood ratio criterion (LRC) is given in Section 3. Then we give some results of the asymptotic distribution of LRC in Section 4 under a set of general conditions including some cases of the weak instruments and many instruments situations. The extensions of our approach to several problems (i.e. the errors-in-variables model, the reduced rank regression and the cointegration models) are discussed in Section 5 and concluding remarks are given in Section 6. The mathematical proofs of theorems are in Section 7.

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 is

(2.1)
$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{V}\ ,$$

where $\boldsymbol{\Pi}$ 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 $N(\mathbf{0},\boldsymbol{\Omega})$. The coefficients $\boldsymbol{\Pi}$ are estimated by the sample regression matrix

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

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

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

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 $\mathbf{u} = \mathbf{V}_1\boldsymbol{\beta}$ 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 (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 (2.4) and (2.5) is

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

The second part of (2.6),

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

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

Consider the null hypothesis

$$\mathbf{H}_0 : \ \mathbf{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0} \ ,$$

where $\boldsymbol{\beta}_0$ is a (non-zero) specified vector. The alternative hypothesis, say $\mathbf{H}_2$, consists of arbitrary $\mathbf{\Pi}$ and $\mathbf{\Omega}$.

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

$$\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{Z}\begin{bmatrix} -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \boldsymbol{I}_{K_2} \end{bmatrix} ,$$

where

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

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

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

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. A new derivation of the likelihood ratio criterion

The likelihood function is

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

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 $\mathbf{\Pi}_1^*$ occurrs at $\mathbf{\Pi}_1^* = \mathbf{P}_1^*$ and is

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

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

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

By *Lemma 1* in Section 7, the maximum of $L(\mathbf{\Pi}_2)$ with respect to $\mathbf{\Pi}_{22}$ is

$$(3.4) \qquad L(\mathbf{\Pi}_{21}) = (2\pi)^{-\frac{1}{2}TG} \left|(\mathbf{P}_{2.1} - \mathbf{\Pi}_{21})' \mathbf{A}_{22.1} (\mathbf{P}_{2.1} - \mathbf{\Pi}_{21}) + \mathbf{H}_{11}\right|^{-\frac{1}{2}T} |\mathbf{H}_{22.1}|^{-\frac{1}{2}T} e^{-\frac{1}{2}TG} ,$$

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

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

This is the likelihood maximized with respect to $\mathbf{\Pi}$ and $\mathbf{\Omega}$ without any rank restrictions on coefficient.

Let the $G_1 \times G_1$ matrix be

$$(3.6) \qquad \mathbf{G}_{11} = \mathbf{P}_{2.1}' \mathbf{A}_{22.1} \mathbf{P}_{2.1} .$$

Define ν_1 as the smallest root of

$$(3.7) \qquad |\mathbf{G}_{11} - \lambda \mathbf{H}_{11}| = 0 ;$$

that is

$$(3.8) \qquad \nu_1 = \min_{\mathbf{b}} \frac{\mathbf{b}' \mathbf{G}_{11} \mathbf{b}}{\mathbf{b}' \mathbf{H}_{11} \mathbf{b}} = \frac{\hat{\boldsymbol{\beta}}' \mathbf{G}_{11} \hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}' \mathbf{H}_{11} \hat{\boldsymbol{\beta}}} ,$$

where $\hat{\boldsymbol{\beta}}$ is the LIML estimator of $\boldsymbol{\beta}$. Then the likelihood function maximized under $\mathbf{H}_1$ is [1]

$$(3.9) \qquad \mathrm{L}_{H_1} = (2\pi)^{-\frac{1}{2}TG} |\mathbf{H}|^{-\frac{1}{2}T} (1 + \nu_1)^{-\frac{1}{2}T} e^{-\frac{1}{2}TG} .$$

Hence the likelihood ratio criterion for testing $\mathbf{H}_1$ against the alternative hypothesis that $\mathbf{\Pi}$ is unrestricted is

$$(3.10) \qquad \frac{\mathrm{L}_{H_1}}{\mathrm{L}_{H_2}} = (1 + \nu_1)^{-\frac{1}{2}T} = \left[1 + \min_{\mathbf{b}} \frac{\mathbf{b}' \mathbf{G}_{11} \mathbf{b}}{\mathbf{b}' \mathbf{H}_{11} \mathbf{b}}\right]^{-\frac{1}{2}T} .$$

[1] The result can be directly obtained by (3.15) below by substituting the parameter vector $\boldsymbol{\beta}$ for $\boldsymbol{\beta}_0$ and then maximizing the likelihood function with respect to $\boldsymbol{\beta}$.

(See Anderson and Rubin (1949), Theorem 2.)

Now consider maximizing the likelihood function under $\mathbf{H}_0 : \mathrm{rank}(\mathbf{\Pi}_{21}) = G_1 - 1$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. The matrix $\mathbf{\Pi}_{21}$ can be parameterized as

$$\mathbf{\Pi}_{21} = \boldsymbol{\mu}\mathbf{\Gamma}' \ , \tag{3.11}$$

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

$$\mathbf{\Gamma}'\boldsymbol{\beta}_0 = \mathbf{0} \ . \tag{3.12}$$

Then by a direct minimization [2]

$$\begin{aligned} &\min_{\boldsymbol{\mu}} \left|(\mathbf{P}_{2.1} - \boldsymbol{\mu}\mathbf{\Gamma}')'\mathbf{A}_{22.1}(\mathbf{P}_{2.1} - \boldsymbol{\mu}\mathbf{\Gamma}') + \mathbf{H}_{11}\right| \\ = &\left|(\mathbf{P}_{2.1} - \hat{\boldsymbol{\mu}}\mathbf{\Gamma}')'\mathbf{A}_{22.1}(\mathbf{P}_{2.1} - \hat{\boldsymbol{\mu}}\mathbf{\Gamma}') + \mathbf{H}_{11}\right| \ , \end{aligned} \tag{3.13}$$

where

$$\hat{\boldsymbol{\mu}} = \mathbf{P}_{2.1}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1} \ . \tag{3.14}$$

The determinant is then

$$\begin{aligned} &\left|[\mathbf{P}_{2.1} - \mathbf{P}_{2.1}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}']'\mathbf{A}_{22.1}[\mathbf{P}_{2.1} - \mathbf{P}_{2.1}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'] + \mathbf{H}_{11}\right| \\ = &\left|[\mathbf{I}_{G_1} - \mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}]\mathbf{G}_{11}[\mathbf{I}_{G_1} - \mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'] + \mathbf{H}_{11}\right| \\ = &|\mathbf{H}_{11}|\left|\left[\mathbf{I}_{G_1} - \mathbf{H}_{11}^{-1/2}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'\mathbf{H}_{11}^{-1/2}\right](\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2})\right. \\ &\left.\times\left[\mathbf{I}_{G_1} - \mathbf{H}_{11}^{-1/2}\mathbf{\Gamma}\left(\mathbf{\Gamma}'\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}'\mathbf{H}_{11}^{-1/2}\right] + \mathbf{I}_{G_1}\right| \\ = &|\mathbf{H}_{11}|\left|\left[\mathbf{I}_{G_1} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\right]\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2}\left[\mathbf{I}_{G_1} - \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\right] + \mathbf{I}_{G_1}\right| \ , \end{aligned}$$

where $\mathbf{Q} = \mathbf{H}_{11}^{-1/2}\mathbf{\Gamma}$. 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 $\mathbf{x} = \mathbf{H}_{11}^{1/2}\boldsymbol{\beta}_0$. Then (3.13) is

$$\begin{aligned} &|\mathbf{H}_{11}|\left|\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{H}_{11}^{-1/2}\mathbf{G}_{11}\mathbf{H}_{11}^{-1/2}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}' + \mathbf{I}_{G_1}\right| \\ = &|\mathbf{H}_{11}|\left[1 + \frac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}\right] \end{aligned} \tag{3.15}$$

[2] We can use the relation that $|(\mathbf{P}_{2.1} - \boldsymbol{\mu}\mathbf{\Gamma}')'\mathbf{A}_{22.1}(\mathbf{P}_{2.1} - \boldsymbol{\mu}\mathbf{\Gamma}') + \mathbf{H}_{11}| = |\mathbf{H}_{11}||(\boldsymbol{\mu}\mathbf{\Gamma}' - \mathbf{P}_{2.1})\mathbf{H}_{11}^{-1}(\mathbf{\Gamma}\boldsymbol{\mu} - \mathbf{P}_{2.1}')\mathbf{A}_{22.1} + \mathbf{I}_{G_1-1}|$ and minimize the quadratic form of $\boldsymbol{\mu}$.

by Corollary A.3.1 of Anderson (2003). The likelihood maximized over $\mathbf{H}_0$ is

(3.16)
$$\mathrm{L}_{H_0} = (2\pi)^{-\frac{1}{2}TG}\,|\mathbf{H}|\left[1+\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0}\right]^{-\frac{1}{2}T} .$$

Hence the likelihood ratio criterion for the null hypothesis $\mathbf{H}_0 : \mathbf{\Pi}_{21}$ has rank $G_1 - 1$ and $\mathbf{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$ vs. $\mathbf{H}_1 : \mathbf{\Pi}_{21}$ has rank $G_1 - 1$ is

(3.17)
$$\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \left[\frac{1+\nu_1}{1+\dfrac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0}}\right]^{\frac{1}{2}T} = \left[\frac{1+\dfrac{\hat{\boldsymbol{\beta}}^{'}\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}}}{1+\dfrac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0}}\right]^{\frac{1}{2}T} .$$

The null hypothesis that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ is rejected if the LRC is less than a suitable constant; that is, if

(3.18)
$$\frac{1+\dfrac{\hat{\boldsymbol{\beta}}^{'}\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}}}{1+\dfrac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0}} < c(K_2, T-K) .$$

The likelihood ratio test (3.18) can be written

(3.19)
$$\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} > \frac{1+\nu_1}{c(K_2, T-K)} - 1$$

In Anderson and Rubin (1949) the test is

(3.20)
$$\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} > \frac{K_2}{T-K}F_{K_2,T-K}(\epsilon) ,$$

where $F_{K_2,T-K}(\epsilon)$ denotes the $1-\epsilon$ significance point of the F-distribution with K_2 and $T-K$ degrees of freedom.

Comments :

1. The LRC does not depend on a normalization of the vector of coefficients. The ratio $\boldsymbol{\beta}_0^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\boldsymbol{\beta}_0/\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0$ is unchanged by replacing $\boldsymbol{\beta}_0$ by $\boldsymbol{\beta}_0$ times an arbitrary constant. Similarly, $\hat{\boldsymbol{\beta}}^{'}\mathbf{P}_{2.1}^{'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}\hat{\boldsymbol{\beta}}/\hat{\boldsymbol{\beta}}^{'}\mathbf{H}_{11}\hat{\boldsymbol{\beta}}$ 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 LRC (3.10) compares the hypothesized $\boldsymbol{\beta}_0$ with the LIML estimator $\hat{\boldsymbol{\beta}}$.

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

4. The LRC is invariant with respect to linear transformations $\mathbf{Y}_1 \rightarrow \mathbf{Y}_1\mathbf{C}$, $\boldsymbol{\beta}_0 \rightarrow \mathbf{C}^{-1}\boldsymbol{\beta}_0$ and $\mathbf{Z}_2 \rightarrow \mathbf{Z}_2\mathbf{D}$ 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 (3.7).

5. The logarithm of the likelihood ratio criterion is

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

which is approximately

$$\frac{1}{2}\mathrm{T}\left[\frac{\hat{\boldsymbol{\beta}}'\mathbf{G}_{11}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\mathbf{H}_{11}\hat{\boldsymbol{\beta}}} - \frac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0}\right] . \tag{3.22}$$

Moreira (2003) has arrived at (3.22) by another route. He considered criteria which are functions of the sufficient statistics that are invariant with respect to certain linear transformations when $\boldsymbol{\Omega}_{11}$ is known and expressed the statistic as

$$LR_0 = \bar{\mathbf{S}}'\bar{\mathbf{S}} - \lambda^{min} ,$$

where λ^{min} is the smallest eigenvalue of $(\bar{\mathbf{S}},\bar{\mathbf{T}})'(\bar{\mathbf{S}},\bar{\mathbf{T}})$,

$$(\bar{\mathbf{S}},\bar{\mathbf{T}}) = \mathbf{Z}_{2.1}'\mathbf{Z}_{2.1}\left[\mathbf{S}(\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0)^{-1/2}, \mathbf{T}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}_{11}\boldsymbol{\Gamma})^{-1/2}\right]$$

and

$$\mathbf{S} = \mathbf{Z}_{2.1}'\mathbf{Y}_1\boldsymbol{\beta}_0 \, , \mathbf{T} = \mathbf{Z}_{2.1}'\mathbf{Y}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Gamma} \, .$$

(We have used our notations here.) He has proposed to use the simulated distribution of LR_0 when $\boldsymbol{\Omega}_{11}$ is known for testing $\mathbf{H}_0$.

4. Asymptotic Distributions

We shall investigate the limiting distributions of the likelihood ratio (LR) statistic under conditions much more general than the conditions under which the test was developed. Let the $\sigma-$field $\mathcal{F}_{t-1}$ be generated by $\mathbf{z}_1, \mathbf{v}_1, \cdots, \mathbf{z}_{t-1}, \mathbf{v}_{t-1}, \mathbf{z}_t$ $(t=1,\cdots,T)$ and $\mathcal{F}_0$ is the initial $\sigma-$field generated by $\mathbf{z}_1$. We partition $1\times(G_1+G_2)$ vectors $\mathbf{v}_t' = (\mathbf{v}_{1t}', \mathbf{v}_{2t}')$ $(t=1,\cdots,T)$, and we assume that $\mathcal{E}(\mathbf{v}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ *a.s.*,

$\mathcal{E}(\mathbf{v}_t\mathbf{v}_t^{'}|\mathcal{F}_{t-1}) = \mathbf{\Omega}_t$ *a.s.*, and $\mathbf{\Omega}_t$ can be 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 $\mathcal{E}(u_t|\mathcal{F}_{t-1}) = \mathbf{0}$ *a.s.* and $\mathcal{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}_{11}^{(t)}\boldsymbol{\beta}$ *a.s.*, where $\mathbf{\Omega}_t$ is a $(G_1+G_2)\times(G_1+G_2)$ matrix

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

We first investigate the limiting distribution of LR statistic under the standard situation when T is large. Suppose

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

$$\textbf{(II)} \qquad \frac{1}{T}\max_{1\le t\le T}\|\mathbf{z}_t\|^2 \xrightarrow{p} 0 \ \ (\text{as } T\to\infty)\ ,$$

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

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

$$\textbf{(V)} \qquad \sup_{t\ge1}\mathcal{E}[\mathbf{v}_{1t}^{'}\mathbf{v}_{1t}\mathrm{I}(\mathbf{v}_{1t}^{'}\mathbf{v}_{1t}>c)|\mathcal{F}_{t-1}] \xrightarrow{p} 0 \ \ (\text{as } c\to\infty)\ ,$$

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

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

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 the stationary AR processes, for instance.

In order to investigate the limiting null-distribution and the local power of LRC, we consider a sequence of local alternatives as

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

where $\boldsymbol{\xi}_i$ $(i=1,2)$ are $K_i \times 1$ $(i=1,2)$ vectors, each element of the $(K_1+K_2)\times(G_1+G_2)$ matrix $\mathbf{\Pi}$ are functions of T (say $\mathbf{\Pi}_T$) and it is partitioned as $\mathbf{\Pi}_T = (\mathbf{\Pi}_{ij}^{(T)})$. Hence $\lim_{T\to\infty}\mathbf{\Pi}_{21}^{(T)} = \mathbf{\Pi}_{21}$ and $\mathbf{\Pi}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$ as the limit $(T\to\infty)$ in (4.2). (See (2.6) and (2.7) in Section 2.) Then Theorem 1 is an extension of Theorem 4 of Anderson and Kunitomo (1994). The proof is given in Section 7.

Theorem 1 : 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 the non-central χ^2 with G_1-1 degrees of freedom and the non-centrality parameter $\kappa_1 = \theta_1\sigma^{-2}$, where $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}_{11}\boldsymbol{\beta}$, $\mathbf{M}_{22.1} = \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12}$,

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

and a $(K_1+K_2)\times(K_1+K_2)$ matrix

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

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

Under $\mathbf{H}_0$ $(\boldsymbol{\xi}=\mathbf{0})$, the limiting distribution of LR_1 is χ^2 with G_1-1 degrees of freedom under the general conditions on disturbances. Then by using the χ^2 distribution in Theorem 1 when T is large in (3.19), we can take the rejection region as

$$(4.5)\qquad \frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}\boldsymbol{\beta}_0} > [1+\nu_1]\,e^{\frac{1}{T}\chi^2_{G_1-1}(\epsilon)} - 1$$

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

Next, we consider the case of so-called *weak* instruments in econometrics. Let $\mathbf{\Pi}_T = \mathbf{C}/T^{\delta}$ for a constant matrix $\mathbf{C}$ and $\delta > 0$. 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 $\mathbf{\Pi}$, accordingly. Then Condition (I) implies

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

We rewrite

(4.6)
$$(\mathbf{Y}_1^{(T)}, \mathbf{Y}_2^{(T)}) = \mathbf{Z}\mathbf{\Pi}_T + \mathbf{V} \ ,$$

$$\mathbf{G}_{11}^{(T)} = \mathbf{P}_{2.1}^{(T)'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}^{(T)}$$

and

$$\mathbf{H}_{11}^{(T)} = \mathbf{Y}_1^{(T)'}\mathbf{Y}_1^{(T)} - \mathbf{P}_{2.1}^{(T)'}\mathbf{A}_{22.1}\mathbf{P}_{2.1}^{(T)} \ .$$

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

(4.7)
$$\nu_1^{(T)} = \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{G}_{11}^{(T)}\mathbf{b}}{\mathbf{b}'\mathbf{H}_{11}^{(T)}\mathbf{b}} = \frac{\hat{\boldsymbol{\beta}}^{(T)'}\mathbf{G}_{11}^{(T)}\hat{\boldsymbol{\beta}}^{(T)}}{\hat{\boldsymbol{\beta}}^{(T)'}\mathbf{H}_{11}^{(T)}\hat{\boldsymbol{\beta}}^{(T)}} \ ,$$

where $\hat{\boldsymbol{\beta}}^{(T)} = (1, -\hat{\boldsymbol{\beta}}_2^{(T)'})'$ is the LIML estimator of $\boldsymbol{\beta}$.
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 could be measured by the parameter δ. Theorem 2 states the limiting distribution of the LR statistic when $0 < \delta < 1/2$, of which the proof is similar to Theorem 1 and it is omitted.

Theorem 2 : Assume $\mathbf{\Pi}_T = \mathbf{C}/T^{\delta}$ for a (constant) $K \times G_1$ matrix $\mathbf{C}$ with $0 < \delta < 1/2$ and Conditions (I) – (V). Under the local alternative sequences with $\eta = 1/2$

(4.8)
$$\begin{bmatrix} \mathbf{\Pi}_{11}^{(T)} & \mathbf{\Pi}_{12}^{(T)} \\ \mathbf{\Pi}_{21}^{(T)} & \mathbf{\Pi}_{22}^{(T)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} + \frac{1}{\mathrm{T}^{\eta}} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \ ,$$

as $T \to \infty$ the limiting distribution of LR_1 is the non-central χ^2 with $G_1 - 1$ degrees of freedom and the non-centrality parameter $\kappa_2 = \theta_2\sigma^{-2}$, where $\sigma^2 = \boldsymbol{\beta}_0'\mathbf{\Omega}_{11}\boldsymbol{\beta}_0$,

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

and a $K_2 \times (G_1 - 1)$ matrix

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

has rank $G_1 - 1$.

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

When $\delta \geq 1/2$ and the instruments are extremely weak, however, the limiting distribution of LR_1 under $\mathbf{H}_0$ is not a χ^2 distribution. First we consider the case when $\delta = 1/2$. Define

$$\mathbf{X}_T = \frac{1}{\sqrt{T}} \mathbf{A}_{22.1}^{-1/2} \mathbf{Z}_{2.1}^{'} \mathbf{V}_1 \ . \tag{4.10}$$

Then for any constant vector $\mathbf{a}$, $\mathbf{X}_T\mathbf{a}$ converges to $\mathbf{X}\mathbf{a}$ weakly as $T \to \infty$ under Conditions (I)-(V) and $\mathbf{X}\mathbf{a}$ follows $N_K\left[\mathbf{0}, (\mathbf{a}^{'}\mathbf{\Omega}_{11}\mathbf{a})\mathbf{I}_K\right]$. Since $\mathbf{H}_{11}^{(T)}/T \xrightarrow{p} \mathbf{\Omega}_{11}$, we find that for any $\mathbf{C}_{21}^{'}$

$$T\,\frac{\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}^{(T)}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{H}_{11}^{(T)}\boldsymbol{\beta}_0} \xrightarrow{w} \frac{\boldsymbol{\beta}_0^{'}(\mathbf{C}_{21}^{'}\mathbf{M}_{22.1}^{1/2} + \mathbf{X}^{'})(\mathbf{M}_{22.1}^{1/2}\mathbf{C}_{21} + \mathbf{X})\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{\Omega}_{11}\boldsymbol{\beta}_0}$$

and

$$\min_{\mathbf{b}} \frac{\mathbf{b}^{'}\mathbf{G}_{11}^{(T)}\mathbf{b}}{\mathbf{b}^{'}\mathbf{H}_{11}^{(T)}\mathbf{b}} \xrightarrow{w} \nu_1^* \ ,$$

where ν_1^* is the smallest root of

$$\left|\mathbf{G}_{11}^{(*)} - \lambda^*\mathbf{\Omega}_{11}\right| = 0 \ ; \tag{4.11}$$

that is

$$\nu_1^* = \frac{\hat{\boldsymbol{\beta}}^{*'}(\mathbf{C}_{21}^{'}\mathbf{M}_{22.1}^{1/2} + \mathbf{X}^{'})(\mathbf{M}_{22.1}^{1/2}\mathbf{C}_{21} + \mathbf{X})\hat{\boldsymbol{\beta}}^{*}}{\hat{\boldsymbol{\beta}}^{*'}\mathbf{\Omega}_{11}\hat{\boldsymbol{\beta}}^{*}}$$

and $\hat{\boldsymbol{\beta}}^{*}$ is the characteristic vector of (4.11) with ν_1^*. Then as $T \to \infty$ under the condition $\mathbf{C}_{21}\boldsymbol{\beta}_0 = \mathbf{0}$,

$$\begin{aligned} LR_1 &\xrightarrow{w} \frac{\boldsymbol{\beta}_0^{'}(\mathbf{C}_{21}^{'}\mathbf{M}_{22.1}^{1/2} + \mathbf{X}^{'})(\mathbf{M}_{22.1}^{1/2}\mathbf{C}_{21} + \mathbf{X})\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{\Omega}_{11}\boldsymbol{\beta}_0} - \nu_1^* \\ &= \frac{\boldsymbol{\beta}_0^{'}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{\Omega}_{11}\boldsymbol{\beta}_0} - \nu_1^* \ (= LR_1^*, \text{ say}), \end{aligned} \tag{4.12}$$

where the first term of the limiting random variable follows $\chi^2(K_2)$ and the second term (i.e. ν_1^*) follows the minimum of a non-central Wishart matrix. Hence we have a representation of the $G_1 \times G_1$ random matrix

$$\mathbf{G}_{11}^{(*)} = (\mathbf{C}_{21}^{'}\mathbf{M}_{22.1}^{1/2} + \mathbf{X}^{'})(\mathbf{M}_{22.1}^{1/2}\mathbf{C}_{21} + \mathbf{X})$$

and it is actually a central Wishart if and only if $\mathbf{C}_{21} = \mathbf{O}$.
When $\delta > 1/2$, we have

$$LR_1 \xrightarrow{w} \frac{\boldsymbol{\beta}_0^{'}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\mathbf{\Omega}_{11}\boldsymbol{\beta}_0} - \nu_1^{**} \ = LR_1^* \ , \tag{4.13}$$

where ν_1^{**} is the smallest root of

$$\left|\mathbf{X}^{'}\mathbf{X} - \lambda^{**}\boldsymbol{\Omega}_{11}\right| = 0 \ . \tag{4.14}$$

In our formulation it is possible to analyze the asymptotic behavior of LR_1 under the local alternatives when $\eta \geq 1/2$ with some complications. There is no technical difficulty, but we need some further notations.

Then we can investigate the asymptotic behavior of LR_1^* when K or K_2 is large. An interesting observation is the fact that the LIML estimator is still consistent when K_2 is large and $\delta = 1/2$. On the other hand, when $\delta > 1/2$ the structural relation is asymptotically under-identification. This leads to the asymptotic behavior of LR_1^* when K_2 is large, which is different from the standard situation. We summarize the results on the asymptotic distributions of LR_1^* under $\mathbf{H}_0$ when K_2 is large. The proof is given in Section 7.

Theorem 3 : Suppose each row of $\mathbf{X}$ follows $N_{G_1}(\mathbf{0}, \boldsymbol{\Omega}_{11})$. $[i]$ Let $\mathbf{M}_{K_2}^* = \mathbf{C}_{21}^{'}\mathbf{M}_{22.1}\mathbf{C}_{21}$ and assume

$$\textbf{(VI)} \qquad \frac{1}{K_2}\mathbf{M}_{K_2}^* \rightarrow \mathbf{M}^*$$

as $K_2 \rightarrow \infty$ and the lower-right corner $((G_1-1)\times(G_1-1))$ submatrix $\mathbf{M}_{22}^*$ of $\mathbf{M}^*$ is a non-singular matrix (i.e. $\mathbf{M}_{22}^*$ is of rank G_1). When $\delta = 1/2$, as $K_2 \rightarrow \infty$

$$LR_1^* \xrightarrow{w} \mathbf{x}^{'}\mathbf{x} \ , \tag{4.15}$$

where $\mathbf{x}$ follows the $(G_1 - 1)$-dimensional normal distribution with the covariance matrix

$$\mathbf{Q}^* = \mathbf{I}_{G_1-1} + \mathbf{M}_{22}^{*-1/2}\left[\boldsymbol{\Omega}_{11} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}}{\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]_{22}\mathbf{M}_{22}^{*-1/2} \ ,$$

which is a positive definite matrix and $[\ \cdot\ ]$ stands for the $(G_1-1)\times(G_1-1)$ lower-right corner of the matrix. $[ii]$ When $\delta > 1/2$, as $K_2 \rightarrow \infty$

$$\frac{1}{\sqrt{K_2}}LR_1^* \xrightarrow{w} \tau_{G_1} \ , \tag{4.16}$$

where τ_{G_1} is the largest characteristic root of the symmetric matrix

$$\mathbf{W}^* = \frac{1}{\sqrt{K_2}}\left[\left(\frac{\boldsymbol{\beta}_0^{'}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right)\mathbf{I}_{G_1} - \boldsymbol{\Omega}_{11}^{-1/2}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\Omega}_{11}^{-1/2}\right] = (w_{ij}^*) \ , \tag{4.17}$$

and each elements w^*_{ij} follows the Gaussian distributions with zero means and $\mathcal{E}(w^{*2}_{ii}) = 4(1-a_i^2)$, $\mathcal{E}(w^{*2}_{ij}) = 1$ $(i \neq j)$, $\mathcal{E}(w^*_{ii}w^*_{ij}) = -2a_i a_j$ $(i \neq j)$, $\mathcal{E}(w^*_{ii}w^*_{jj}) = 2[1-a_i^2-a_j^2]$ $(i \neq j)$ and

$$\mathbf{a} = (a_i) = \frac{\mathbf{\Omega}_{11}^{1/2}\boldsymbol{\beta}_0}{\|\mathbf{\Omega}_{11}^{1/2}\boldsymbol{\beta}_0\|} .$$

When $G_1 = 2$, the larger root of the determinant equation is

$$\tau_2 = \frac{w^*_{11}+w^*_{22}+\sqrt{(w^*_{11}+w^*_{22})^2 - 4(w^*_{11}w^*_{22}-w^{*2}_{12})}}{2} . \tag{4.18}$$

The above situations when $T \to \infty$ and $K_2 \to \infty$ can be regarded as some cases of *many weak* instruments recently discussed in econometrics. The alternative (and it may be more natural) formulation of many weak instruments is to relate K_2 to T, and take each elements and the size of $\mathbf{\Pi}$ as functions of T. Let the size $K \times G$ of $\mathbf{\Pi}$ be dependent on T, and we denote a sequence of $K_T \times G$ ($K_T = K_1 + K_{2T}, T \geq 3$, G is a fixed integer) matrices $\mathbf{\Pi}_T$, which is partitioned into the $(K_1+K_{2T}) \times (G_1+G_2)$ 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

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

Also 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 non-negative definite constant matrix (the upper-left $G_1 \times G_1$ sub-matrix of $\mathbf{\Phi}$ is of rank $G_1 - 1$), and $\mathbf{z}_t^{(T)}$ are the $K_T \times 1$ vectors of instruments.

In the many-weak instruments cases, there can be alternative assumptions among the relative magnitudes of T, K_T and $\mathbf{\Pi}_T$. The condition (VII) is a very mild condition and it is not possible to obtain the χ^2- distribution [3] without (VII). The many-weak instruments cases are different from the standard situation for (2.1) and (2.4) with a fixed K (and K_2). We have the next result and we have omitted the proof because it is similar to those of Theorem 1 and Theorem 3.

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

$$\mathbf{\Pi}_T \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} + \frac{1}{\sqrt{T}} \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_{2T} \end{bmatrix} , \tag{4.19}$$

as $T \to \infty$ the statistic LR_1 has the limiting distribution of the non-central χ^2 with $G_1 - 1$ degrees of freedom and the non-centrality parameter $\kappa_4 = \theta_4 \sigma^{-2}$, provided that the probability limits of

$$\theta_4 = \left[\text{plim} \frac{1}{T} \boldsymbol{\xi}_{2T}^{'} \mathbf{A}_{22.1} \mathbf{\Pi}_{2T} \right] \left[\text{plim} \frac{1}{T} \mathbf{\Pi}_{2T}^{'} \mathbf{A}_{22.1} \mathbf{\Pi}_{2T} \right]^{-1} \left[\text{plim} \frac{1}{T} \mathbf{\Pi}_{2T}^{'} \mathbf{A}_{22.1} \boldsymbol{\xi}_{2T} \right] , \tag{4.20}$$

exist and θ_4 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

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

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

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

Thus we also find that the rejection region and confidence region based on χ^2 distribution with $G_1 - 1$ degrees of freedom are asymptotically valid for some cases of weak instruments including some many weak instruments situation. The assumptions of Theorem 2 on weak instruments (with $\boldsymbol{\xi}_2 = \mathbf{C}_{21} \boldsymbol{\beta}_0 = \mathbf{0}$) or Theorem 4 (with $\theta_4 = 0$) on many instruments are sufficient for χ^2 with $G_1 - 1$ degrees of freedom as the asymptotic null-distribution.

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

5. Some Extensions

The likelihood ratio criterion we have developed can be extended to several statistical models, which have been often treated separately in the literatures. We shall discuss three important problems which have many applications in statistical and econometric analyses.

5.1 Linear Functional Relationships

We shall pay an attention to the fundamental relationship between the simultaneous equation system and the linear functional (or the errors-in-variables) models in the statistical literature, which are mathematically equivalent. A linear functional relationships model can be defined as follows.

Let the observed G_1-component vector $\mathbf{X}_{\alpha j}$ $(\alpha = 1, \cdots, K_2; j = 1, \cdots, m)$ be modeled as

$$\mathbf{X}_{\alpha j} = \boldsymbol{\xi}_\alpha + \boldsymbol{V}_{\alpha j} \ , \tag{5.1}$$

where $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{K_2}$ are *incidental parameters*, $\mathbf{V}_{\alpha j}$ are unobserved random vectors distributed as $N(\mathbf{0}, \boldsymbol{\Omega})$, and m is the number of repeated measurements. The assumed linear relationship among $\boldsymbol{\xi}_\alpha$ is

$$\boldsymbol{\xi}_\alpha^{'} \boldsymbol{\beta} = \mathbf{0} \ , \ \alpha = 1, \cdots, K_2 \ . \tag{5.2}$$

Then (5.1) can be written as $\mathbf{X} = \mathbf{Z\Pi} + \mathbf{V}$, where $mK_2 = T$ and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11}^{'} \\ \vdots \\ \mathbf{X}_{1m}^{'} \\ \mathbf{X}_{21}^{'} \\ \vdots \\ \mathbf{X}_{K_2 m}^{'} \end{bmatrix} , \ \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} , \ \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11}^{'} \\ \vdots \\ \mathbf{V}_{1m}^{'} \\ \mathbf{V}_{21}^{'} \\ \vdots \\ \mathbf{V}_{K_2 m}^{'} \end{bmatrix} , \tag{5.3}$$

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\xi}_1^{'} \\ \boldsymbol{\xi}_2^{'} \\ \vdots \\ \boldsymbol{\xi}_{K_2}^{'} \end{bmatrix} .$$

The linear relationship (5.3) implies that the rank of $\boldsymbol{\Pi}$ is $G_1 - 1$. The estimator of $\boldsymbol{\xi}_\alpha$ is $\bar{\mathbf{x}}_\alpha = (1/m) \sum_{j=1}^{m} \mathbf{X}_{\alpha j}$; the estimator of $\boldsymbol{\Pi}^{'} = (\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{K_2})$ of unrestricted rank is $(\bar{\mathbf{x}}_1, \cdots, \bar{\mathbf{x}}_{K_2})$; further for statistical inference it may be natural to use two matrices

$$\mathbf{G}_{11} = m \sum_{\alpha=1}^{K_2} \bar{\mathbf{x}}_\alpha \bar{\mathbf{x}}_\alpha^{'} \, , \mathbf{H}_{11} = \sum_{\alpha=1}^{K_2} \sum_{j=1}^{m} (\mathbf{x}_{\alpha j} - \bar{\mathbf{x}}_\alpha)(\mathbf{x}_{\alpha j} - \bar{\mathbf{x}}_\alpha)^{'} \, . \tag{5.4}$$

The relation between the estimation problem of structural equations in econometrics and the linear functional relationships model has been investigated by Anderson (1984). (See Sections 12 and 13 of Anderson (2003) for the details.) However, the likelihood ratio criteria for testing coefficients have not been fully developed although there were some test statistics proposed. In this respect, the test statistic and confidence region in the form of (3.18), (3.19) and (4.5) are directly applicable.

5.2 Reduced Rank Regression

In (2.1) and (2.5) we consider the null hypothesis

$$\mathbf{H}_0^{'} : \ \ \mathbf{\Pi}_{21}\mathbf{B}_0 = \mathbf{0} \, ,$$

where $\mathbf{B}_0$ is a specified $G_1 \times r$ $(1 \leq r < G_1)$ matrix of rank r. The alternative hypothesis consists of arbitrary $\mathbf{\Pi}$ and $\mathbf{\Omega}$. Consider also

$$\mathbf{H}_1^{'} : \ \ \mathrm{rank}(\mathbf{\Pi}_{21}) = G_1 - r \, .$$

Note that $\mathbf{H}_0$ includes $\mathbf{H}_1$. By using the same argument as in Section 3, the likelihood ratio test of the null hypothesis $\mathbf{H}_2^{'} : \mathbf{\Pi}_{21}$ has rank $G_1 - r$ and $\mathbf{\Pi}_{21}\mathbf{B}_0 = \mathbf{0}$ vs. $\mathbf{H}_1^{'} : \mathbf{\Pi}_{21}$ has rank $G_1 - r$ can be developed. In the derivations of (3.13)-(3.15), we notice that the matrix $\mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}$ is idempotent of rank $G_1 - r$ and $\mathbf{I}_{G_1} - \mathbf{Q}(\mathbf{Q}^{'}\mathbf{Q})^{-1}\mathbf{Q}^{'}$ is 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 $\mathbf{X} = \mathbf{H}_{11}^{1/2}\mathbf{B}_0$. Since (3.13) becomes

$$|\mathbf{H}_{11}| \left| \mathbf{I}_r + \mathbf{B}_0^{'}\mathbf{G}_{11}\mathbf{B}_0 \left[\mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0 \right]^{-1} \right| \, , \tag{5.5}$$

then (3.16) can be replaced by a monotone function of

$$\frac{\prod_{i=1}^{r}(1+\nu_i) \left| \mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0 \right|}{\left| \mathbf{B}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})\mathbf{B}_0 \right|} = \min_{\mathbf{B}} \frac{\left| \mathbf{B}^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})\mathbf{B} \right|}{|\mathbf{B}^{'}\mathbf{H}_{11}\mathbf{B}|} \frac{\left| \mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0 \right|}{\left| \mathbf{B}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})\mathbf{B} \right|} \, , \tag{5.6}$$

where $\mathbf{B}$ is a $G_1 \times r$ matrix and ν_i is the i-th smallest root $(i = 1, \cdots, G_1)$ of (3.7). The likelihood ratio test in (3.17) becomes the form of

$$\frac{\left| \mathbf{B}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})\mathbf{B}_0 \right|}{\left| \mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0 \right|} > \frac{\prod_{i=1}^{r}(1+\nu_i)}{c^*(K_2, T-K)} \, , \tag{5.7}$$

where $c^*(K_2, T-K)$ is a suitable constant.

The resulting test procedure and confidence region are invariant to the linear transformations of $\mathbf{\Xi}_0$ and they are direct extensions of Section 3 to the reduced rank regression problem. (See Anderson (1951), Anderson and Amemiya (1991) for the details, for instance.) The degrees of freedom of χ^2-distribution for the statistic LR_2

$$LR_2 = T \log \left[\frac{\frac{\left|\mathbf{B}_0^{'}(\mathbf{G}_{11}+\mathbf{H}_{11})\mathbf{B}_0\right|}{\left|\mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0\right|}}{\prod_{i=1}^{r}(1+\nu_i)} \right] \tag{5.8}$$

is $r(G_1 - r)$ in the reduced rank regression.
It is straightforward to extend our analysis of the limiting distribution of LRC in Section 4 to the present case.

5.3 Cointegration

It has been known that the conintegration problem in econometrics can be essentially reduced to the reduced rank regression in the previous subsection. The main interest in the former is to make statistical inference on cointegrating vectors $\mathbf{\Gamma} = \mathbf{\Gamma}_0$ for

$$\mathbf{\Gamma}_0^{'}\mathbf{B}_0 = \mathbf{O} \tag{5.9}$$

under the hypothesis $\mathbf{H}_0^{'}$ when $\mathbf{\Gamma}_0$ is a $G_1 \times (G_1 - r)$ matrix consisting of cointegrating vectors. (See Johansen (1995) and Anderson (2000), for instance.)

Let a $G \times 1$ time series vector $\mathbf{x}_t$ follows

$$\begin{aligned} \Delta \mathbf{x}_t &= \left[\mathbf{\Pi}_1^{'}(1), \cdots, \mathbf{\Pi}_1^{'}(p)\right] \begin{bmatrix} \Delta \mathbf{x}_{t-1} \\ \vdots \\ \Delta \mathbf{x}_{t-p} \end{bmatrix} + \mathbf{\Pi}_2^{'}\mathbf{x}_{t-1} + \mathbf{v}_t \\ &= \mathbf{\Pi}_1^{'}\mathbf{z}_{1t} + \mathbf{\Pi}_2^{'}\mathbf{z}_{2t} + \mathbf{v}_t \ , \end{aligned} \tag{5.10}$$

where $\mathbf{\Pi}_1^{'} = (\mathbf{\Pi}_1^{'}(1), \cdots, \mathbf{\Pi}_1^{'}(p))$ and $\mathbf{\Pi}_2^{'}$ are $G \times Gp$ and $G \times G$ matrices of coefficients. Then we take a $T \times G$ matrix $\mathbf{Y} = (\Delta \mathbf{x}_t^{'})$ and a $T \times (Gp + G)$ matrix $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, where $\Delta \mathbf{x}_{t-i} = \mathbf{x}_{t-i} - \mathbf{x}_{t-(i+1)}$ $(i = 1, \cdots, p)$, $\mathbf{z}_{1t}^{'} = (\Delta \mathbf{x}_{t-1}^{'}, \cdots, \Delta \mathbf{x}_{t-p}^{'})$ and $\mathbf{z}_{2t}^{'} = \mathbf{x}_{t-1}^{'}$. In the cointegration case $(G = G_1)$ instead of Conditions (I)-(III), we assume the condition [4] 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 $[0,1)$.

By using the same arguments of Section 3 and Section 5.2, the determinant of the maximized likelihood function in (3.4) (and thus (3.13) or (5.5)) under $\mathbf{H}_0^{'}$: $\text{rank}(\mathbf{\Pi}_2) = G_1 - r$ and $\mathbf{\Gamma} = \mathbf{\Gamma}_0$ $(G = G_1)$ is proportional to

$$\begin{aligned}
& |\mathbf{H}_{11}| \left|\left[\mathbf{H}_{11}^{-1} - \mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\right]\mathbf{G}_{11} + \mathbf{I}_{G_1}\right| \\
= \; & |\mathbf{H}_{11}|\,|\mathbf{G}_{11}| \left|(\mathbf{G}_{11}^{-1} + \mathbf{H}_{11}^{-1}) - \mathbf{H}_{11}^{-1}\mathbf{\Gamma}\left(\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\right| \\
= \; & |\mathbf{H}_{11}|\,|\mathbf{G}_{11}| \left|\mathbf{G}_{11}^{-1} + \mathbf{H}_{11}^{-1}\right| \times \frac{\left|\mathbf{\Gamma}^{'}\left[\mathbf{H}_{11}^{-1} - \mathbf{H}_{11}^{-1}(\mathbf{G}_{11}^{-1} + \mathbf{H}_{11}^{-1})^{-1}\mathbf{H}_{11}^{-1}\right]\mathbf{\Gamma}\right|}{\left|\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right|} \\
= \; & \frac{|\mathbf{G}_{11} + \mathbf{H}_{11}|}{|\mathbf{H}_{11}|} \times \frac{\left|\mathbf{\Gamma}^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}\right|}{\left|\mathbf{\Gamma}^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}\right|} .
\end{aligned}$$

Hence the likelihood ratio test in (3.17) can be replaced by

$$\text{(5.11)} \qquad \frac{\left|\mathbf{\Gamma}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}_0\right|}{\left|\mathbf{\Gamma}_0^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}_0\right|} > \frac{\prod_{i=1}^{r}(1+\nu_i)}{c^{**}(K_2, T-K)} ,$$

where $c^{**}(K_2, T-K)$ is a suitable constant.

In the cointegrating case, the LRC can be written in terms of

$$\text{(5.12)} \qquad LR_3 = T\log\left[\prod_{i=1}^{r}\xi_i\frac{\left|\mathbf{\Gamma}_0^{'}(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1}\mathbf{\Gamma}_0\right|}{\left|\mathbf{\Gamma}_0^{'}\mathbf{H}_{11}^{-1}\mathbf{\Gamma}_0\right|}\right] ,$$

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

$$\text{(5.13)} \qquad \left|(\mathbf{G}_{11} + \mathbf{H}_{11})^{-1} - \zeta\mathbf{H}_{11}^{-1}\right| = 0 .$$

Then we have the next result on the limiting distribution of LR_3, which is analogous to the reduced rank regression case. The outline of derivation is given in Section 7.

Theorem 5 : Assume that $\mathbf{v}_t$ are an i.i.d. sequence of random variables with

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

$\mathcal{E}(\mathbf{v}_t) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_t\mathbf{v}_t^{'}) = \mathbf{\Omega}$, and Condition (VII). Then under the rank condition $\mathbf{H}_0^{'}$: 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 the orthogonal transformations of $\mathbf{\Gamma}_0$ (i.e. cointegrating vectors) and they are direct extensions of Section 3 to the cointegration problem.

6. 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 (not always, but) 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 linear functional relationship (errors-in-variables) models, the reduced rank regression and the cointegration models are essentially the same. 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.

7. 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 5. Also we shall use the notation of projection operators $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}^{'}\mathbf{Z})^{-1}\mathbf{Z}^{'}$ and $\mathbf{P}_{Z_1} = \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}$.

Lemma 1 : 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 $\mathbf{A}$,

$$\min_{\mathbf{C}} \left| \begin{pmatrix} \mathbf{B}^{'} \\ \mathbf{C}^{'} \end{pmatrix} \mathbf{A}\,(\mathbf{B},\mathbf{C}) + \mathbf{D} \right| = \left|\mathbf{D}_{22} - \mathbf{D}_{21}\mathbf{D}_{11}^{-1}\mathbf{D}_{12}\right| \left|\mathbf{D}_{11} + \mathbf{B}^{'}\mathbf{A}\mathbf{B}\right| \tag{7.1}$$

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

Proof of Lemma 1: For $|\mathbf{D}| \neq 0$ and $\mathbf{A} > 0$,

$$\text{(7.2)} \quad \left|\mathbf{D} + \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}\,(\mathbf{B},\mathbf{C})\right| = \left|\begin{array}{cc} \mathbf{D} & -\begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix}\mathbf{A}^{1/2} \\ \mathbf{A}^{1/2}(\mathbf{B},\mathbf{C}) & \mathbf{I}_q \end{array}\right|$$

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

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

$$\text{(7.3)} \quad \left|\mathbf{D} + \begin{pmatrix} \mathbf{B}' \\ \mathbf{C}' \end{pmatrix} \mathbf{A}\,(\mathbf{B},\mathbf{C})\right| \geq |\mathbf{D}|\left|\mathbf{I}_q + \mathbf{A}^{1/2}\mathbf{B}\mathbf{D}^{22}\mathbf{B}'\mathbf{A}^{1/2}\right|$$

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

which is the right-hand side of (7.1).
Q.E.D

In order to prove Theorem 1, we first prove two lemmas. (Similar arguments can be used for the proof of Theorem 2 and Theorem 4.)

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

$$\text{(7.4)} \quad T^{\epsilon}\nu_1 \xrightarrow{p} 0 .$$

Proof of Lemma 2 : It is immediate to see that $(1/T)\mathbf{H}_{11} \xrightarrow{p} \mathbf{\Omega}_{11}$ and

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

of which each component of the right-hand side 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

$$LR_d = T\left[\frac{\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}_{11}\boldsymbol{\beta}_0} - \min_{\mathbf{b}}\frac{\mathbf{b}'\mathbf{G}_{11}\mathbf{b}}{\mathbf{b}'\mathbf{H}_{11}\mathbf{b}}\right] . \tag{7.5}$$

Lemma 3 : Under the assumptions of Theorem 1, as $T \to \infty$

$$LR_1 - LR_d \xrightarrow{p} 0 . \tag{7.6}$$

Proof of Lemma 3 : 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 2 as $T \to \infty$.
Q.E.D.

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

$$\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 + \mathbf{G}^{(0)}\left[\begin{array}{c} 0 \\ -\sqrt{T}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right)\end{array}\right] = o_p(1) . \tag{7.7}$$

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

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) = \left[(0,\mathbf{I}_{G_1-1})\mathbf{G}^{(0)}\left(\begin{array}{c}\mathbf{0}' \\ \mathbf{I}_{G_1-1}\end{array}\right)\right]^{-1}(0,\mathbf{I}_{G_1-1})\frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 + o_p(1) . \tag{7.8}$$

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

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

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

$$\boldsymbol{\beta}_0'\mathbf{G}_{11}\boldsymbol{\beta}_0 - T\nu_1\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 - \frac{1}{\sqrt{T}}\boldsymbol{\beta}_0'\mathbf{G}_{11}\left[\begin{array}{c} 0 \\ -\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right)\end{array}\right] = o_p(1) . \tag{7.9}$$

Then by using (7.8) and (7.9) we find that

$$(7.10)\boldsymbol{\beta}_0^{'}\mathbf{G}_{11}\boldsymbol{\beta}_0 - T\nu_1\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0$$

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

The limiting distribution of (7.10) 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

$$(7.11)\quad \frac{1}{\sqrt{T}}\mathbf{G}_{11}\boldsymbol{\beta}_0 = \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{Z}_{2.1}\boldsymbol{\Pi}_{21}\boldsymbol{\beta}_0 + \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}_1\boldsymbol{\beta}_0 + o_p(1)$$
$$= \frac{1}{\sqrt{T}}\boldsymbol{\Pi}_{21}^{'}\mathbf{Z}_{2.1}^{'}\mathbf{V}_1\boldsymbol{\beta}_0 + \boldsymbol{\Pi}_{21}^{'}\mathbf{M}_{22.1}\boldsymbol{\xi}_2 + o_p(1)\ .$$

By applying the CLT (Lindeberg-type Central Limit Theorem, see Anderson and Kunitomo (1992) for instance) to the first term of (7.11) and using (7.10), we have the result.
Q.E.D.

Proof of Theorem 3 :

[i] As $K_2 \to \infty$,

$$(7.12)\quad 0 = \left|\text{plim}\frac{1}{K_2}(\mathbf{C}_{21}^{'}\mathbf{M}_{22.1}^{1/2} + \mathbf{X}^{'})(\mathbf{M}_{22.1}^{1/2}\mathbf{C}_{21} + \mathbf{X}) - \left[\text{plim}\frac{1}{K_2}\nu_1^*\right]\boldsymbol{\Omega}_{11}\right|$$
$$= \left|\mathbf{M}^* + \boldsymbol{\Omega}_{11} - \left[\text{plim}\frac{1}{K_2}\nu_1^*\right]\boldsymbol{\Omega}_{11}\right|$$

and then $(1/K_2)\nu_1^* \xrightarrow{p} 1 = \nu^{(0)}$. Hence $\hat{\boldsymbol{\beta}}^* \xrightarrow{p} \boldsymbol{\beta}_0$. Define $\mathbf{G}^{(1)}$, $\nu^{(1)}$, and $\mathbf{b}_1$ by $\mathbf{G}^{(0)} = \mathbf{M}^* + \boldsymbol{\Omega}_{11}$, $\mathbf{G}^{(1)} = \sqrt{K_2}(\frac{1}{K_2}\mathbf{G}^{(*)} - \mathbf{G}^{(0)})$, $\nu^{(1)} = \sqrt{K_2}(\nu_1^* - 1)$ and $\mathbf{b}_1 = \sqrt{K_2}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0)$. Then

$$\left[\mathbf{G}^{(1)} - \nu^{(1)}\boldsymbol{\Omega}_{11}\right]\boldsymbol{\beta}_0 + \left[\mathbf{G}^{(0)} - \nu^{(0)}\boldsymbol{\Omega}_{11}\right]\begin{bmatrix}0\\ -\mathbf{e}_1\end{bmatrix} = o_p(1)\ ,$$

where $\mathbf{e}_1 = \sqrt{K}(\hat{\boldsymbol{\beta}}_2^* - \boldsymbol{\beta}_2)$. By multiplying $\boldsymbol{\beta}_0^{'}$ from the left and using the fact that $\boldsymbol{\beta}_0^{'}\left[\mathbf{G}^{(0)} - \nu^{(0)}\boldsymbol{\Omega}_{11}\right] = \mathbf{0}^{'}$, we find

$$(7.13)\quad \nu^{(1)} = \frac{\boldsymbol{\beta}_0^{'}\mathbf{G}^{(1)}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0} + o_p(1)\ .$$

Let also define $\mathbf{G}^{(2)}$, $\nu^{(2)}$, and $\mathbf{b}_2$ by $\mathbf{G}^{(2)} = K_2(\frac{1}{K_2}\mathbf{G}^{(*)} - \mathbf{G}^{(0)} - \frac{1}{\sqrt{K_2}}\mathbf{G}^{(1)})$, $\nu^{(1)} = K_2(\nu_1^* - 1 - \frac{1}{\sqrt{K_2}}\nu^{(1)})$ and $\mathbf{b}_2 = K_2(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0 - \frac{1}{\sqrt{K_2}}\mathbf{b}_1)$. Then

$$\left[\mathbf{G}^{(2)} - \nu^{(2)}\boldsymbol{\Omega}_{11}\right]\boldsymbol{\beta}_0 + \left[\mathbf{G}^{(1)} - \nu^{(1)}\boldsymbol{\Omega}_{11}\right]\begin{bmatrix} 0 \\ -\mathbf{e}_1 \end{bmatrix} + \left[\mathbf{G}^{(0)} - \nu^{(0)}\boldsymbol{\Omega}_{11}\right]\begin{bmatrix} 0 \\ -\mathbf{e}_2 \end{bmatrix} = o_p(1)\ ,$$

where $\mathbf{e}_2$ is defined accordingly.

By multiplying $\boldsymbol{\beta}_0'$ from the left and using the above expression for $\nu^{(1)}$, we find that $\boldsymbol{\beta}_0'\mathbf{G}^{(0)}\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 = 0$, $\boldsymbol{\beta}_0'\mathbf{G}^{(1)}\boldsymbol{\beta}_0 - \nu^{(1)}\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 = o_p(1)$,

$$\boldsymbol{\beta}_0'\mathbf{G}^{(2)}\boldsymbol{\beta}_0 - \nu^{(2)}\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0'\mathbf{G}^{(1)}\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\begin{bmatrix} \mathbf{0}' \\ \mathbf{I}_{G_1-1} \end{bmatrix}\mathbf{e}_1 = o_p(1)\ ,$$

and

$$\text{(7.14)}\quad \begin{aligned} \mathbf{M}_{22}^*\mathbf{e}_1 &= [\mathbf{0}, \mathbf{I}_{G_1-1}]\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\mathbf{G}^{(1)}\boldsymbol{\beta}_0 + o_p(1) \\ &= [\mathbf{0}, \mathbf{I}_{G_1-1}]\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right] \\ &\quad \times \left[\frac{1}{\sqrt{K_2}}\mathbf{C}_{21}'\mathbf{M}_{22.1}^{1/2}\mathbf{X}\boldsymbol{\beta}_0 + \sqrt{K_2}(\frac{1}{K_2}\mathbf{X}'\mathbf{X} - \boldsymbol{\Omega}_{11})\boldsymbol{\beta}_0\right] + o_p(1)\ . \end{aligned}$$

We need to evaluate the covariance of the asymptotic distribution and use the relation that the limiting distribution of

$$\boldsymbol{\beta}_0'\mathbf{G}^{(1)}\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\begin{bmatrix} \mathbf{0}' \\ \mathbf{I}_{G_1-1} \end{bmatrix}\mathbf{M}_{22}^{*-1}[\mathbf{0}, \mathbf{I}_{G_1-1}]\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\mathbf{G}^{(1)}\boldsymbol{\beta}_0$$

is the same as the limiting distribution of $[\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0] \times LR_1^*$ as $K_2 \to \infty$. Then by applying CLT to

$$\text{(7.15)}\quad \begin{aligned} &\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\mathbf{G}^{(1)}\boldsymbol{\beta}_0 \\ = &\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\left[\mathbf{C}_{21}'\mathbf{M}_{22.1}^{1/2}\mathbf{X} + \sqrt{K_2}(\frac{1}{K_2}\mathbf{X}'\mathbf{X} - \boldsymbol{\Omega}_{11})\right]\boldsymbol{\beta}_0 + o_p(1) \end{aligned}$$

and using the relation

$$\text{(7.16)}\quad \begin{aligned} &\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right]\left[\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\Omega}_{11} + \boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\right]\left[\mathbf{I}_{G_1} - \frac{\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right] \\ = &\ \boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\left(\boldsymbol{\Omega}_{11} - \frac{\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}}{\boldsymbol{\beta}_0'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_0}\right)\ , \end{aligned}$$

we have the result.

[*ii*] We apply the CLT to each elements of

$$\mathbf{W}^{**} = \sqrt{K_2}\left[\frac{1}{K_2}\mathbf{\Omega}_{11}^{-1/2}\mathbf{X}'\mathbf{X}\mathbf{\Omega}_{11}^{-1/2} - \mathbf{I}_{G_1+1}\right] \tag{7.17}$$

and $W_0 = \mathbf{a}'\mathbf{W}^{**}\mathbf{a}$ as $K_2 \to \infty$. Since the asymptotic distributions of $\mathbf{W}^{**}$ and W_0 are the Gaussian distributions when $K_2 \to \infty$, we only need to calculate their asymptotic covariance. By using direct evaluations $\mathcal{E}(w_{ii}^2) = 2$, $\mathcal{E}(w_{ij}^2) = 1$ $(i \neq j)$, $\mathcal{E}(w_{ii}w_{jj}) = 0$ $(i \neq j)$, $\mathcal{E}(w_0^2) = 2$ and $\mathcal{E}(w_{ij}w_0) = 2a_ia_j$ $(i \neq j)$. Then by evaluating the second moments of each elements of

$$\mathbf{W}^{*} = \mathbf{a}'\mathbf{W}^{**}\mathbf{a}\mathbf{I}_{G_1} - \mathbf{W}^{**} = (w_{ij}^*) \tag{7.18}$$

in (4.17) and noting the fact that LR_1^* is the maximum of $\mathbf{W}^*$, we have the result. **Q.E.D.**

Proof of Theorem 5: We shall consider the limiting distribution of LR_2 of (5.8), which is the same of LR_3 of (5.11), and we shall use the similar arguments as AR(1994). We utilize the fact that $K_2 = G = G_1$ in the cointegration case, and set $\mathbf{Y} = \mathbf{Y}_1$ and $\mathbf{V} = \mathbf{V}_1$. Let a $G_1 \times [(G_1 - r) + r]$ matrix $\mathbf{\Phi} = (\mathbf{\Gamma}_0, \mathbf{B}_0)$ and a $(K_1 + G_1) \times [K_1 + (G_1 - r) + r]$ matrix

$$\mathbf{\Psi}^* = \left[\begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}, \begin{pmatrix} \mathbf{O} \\ \mathbf{\Gamma}_0 \end{pmatrix}, \begin{pmatrix} \mathbf{O} \\ \mathbf{B}_0 \end{pmatrix}\right],$$

a $G_1 \times r$ matrix $\mathbf{B}_0 = (\mathbf{I}_r, -\mathbf{B}_2')'$ and a $K \times [K_1 + (G_1 - r) + r]$ matrix

$$\mathbf{\Gamma}^* = \left[\begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}, \begin{pmatrix} \mathbf{O} \\ \mathbf{\Gamma}_0 \end{pmatrix}\right].$$

For normalizations, we set $\mathbf{\Pi}_2\mathbf{\Gamma}_0 = (\mathbf{I}_{G_0-r}, \mathbf{O})'$ for convenience and

$$\mathbf{D}_T = \begin{bmatrix} \frac{1}{\sqrt{T}}\mathbf{I}_{K_1+G_1-r} & \mathbf{O} \\ \mathbf{O} & \frac{1}{T}\mathbf{I}_r \end{bmatrix}.$$

We use the fact that for $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, each row of $\mathbf{Z}_1$ and $\mathbf{Z}_2\mathbf{\Gamma}_0$ is a vector stationary process and each row of $\mathbf{Z}_2\mathbf{B}_0$ follows an $I(1)$ (the 1st order integrated) process under $\mathbf{H}_0'$. We prepare the following lemmas. (Their proofs are based on the similar arguments given in Appendix B of Johansen (1995) and so we have omitted the details.)

Lemma 4 : Under $\mathbf{H}_0^{'}$, we have the weak convergence

$$(7.19)\quad \begin{bmatrix} \frac{1}{T}\mathbf{\Gamma}^{*'}\boldsymbol{Z}^{'}\boldsymbol{Z}\mathbf{\Gamma}^{*} & \frac{1}{T\sqrt{T}}\mathbf{\Gamma}^{*'}\mathbf{Z}^{'}\mathbf{Z}\mathbf{B}_0 \\ \frac{1}{T\sqrt{T}}\mathbf{B}_0^{'}\mathbf{Z}_2^{'}\mathbf{Z}\mathbf{\Gamma}^{*} & \frac{1}{T^2}\mathbf{B}_0^{'}\mathbf{Z}_2^{'}\mathbf{Z}_2\mathbf{B}_0 \end{bmatrix} \xrightarrow{w} \mathbf{M}^{*} = \begin{bmatrix} \mathbf{\Gamma}^{*'}\mathbf{M}\mathbf{\Gamma}^{*} & \mathbf{\Gamma}^{*'}\mathbf{M}_{.2}\mathbf{B}_0 \\ \mathbf{B}_0^{'}\mathbf{M}_{2.}\mathbf{\Gamma}^{*} & \mathbf{B}_0^{'}\mathbf{M}_{22}\mathbf{B}_0 \end{bmatrix}$$

where $\mathbf{\Gamma}^{*'}\mathbf{M}\mathbf{\Gamma}^{*}$ is a $(K_1+G_1-r)\times(K_1+G_1-r)$ constant matrix, $\mathbf{B}_0^{'}\mathbf{M}_{2.}\mathbf{\Gamma}^{*}$ and $\mathbf{B}_0^{'}\mathbf{M}_{22}\mathbf{B}_0$ are random matrices.

Lemma 5 : Under $\mathbf{H}_0^{'}$, we have

$$(7.20)\quad \frac{1}{\sqrt{T}}\mathbf{\Gamma}^{*'}\mathbf{Z}\mathbf{U}\mathbf{c} \xrightarrow{w} N_{K_1+G_1-r}(\mathbf{0}, \mathbf{c}^{'}\mathbf{\Sigma}\mathbf{c}\mathbf{\Gamma}^{*'}\boldsymbol{M}\mathbf{\Gamma}^{*})$$

for any constant (non-zero) vector $\mathbf{c}$, where $\mathbf{U}=\mathbf{V}\mathbf{B}_0$ and $\mathbf{\Sigma}=\mathbf{B}_0^{'}\mathbf{\Omega}\mathbf{B}_0$.

We use the relations

$$\mathbf{\Pi}^{'}\mathbf{Z}^{'}\mathbf{Z}\mathbf{\Pi} = \mathbf{\Pi}^{'}\mathbf{\Psi}^{'-1}\mathbf{D}_T\left[\mathbf{D}_T^{-1}\mathbf{\Psi}^{'}\mathbf{Z}^{'}\mathbf{Z}\mathbf{\Psi}\mathbf{D}_T^{-1}\right]\mathbf{D}_T\mathbf{\Psi}^{-1}\mathbf{\Pi}\ ,$$

$$\begin{aligned} & \mathbf{\Pi}^{'}\mathbf{Z}^{'}\mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}\mathbf{\Pi} \\ = \ & \mathbf{\Pi}^{'}\mathbf{\Psi}^{'-1}\mathbf{D}_T\left[\mathbf{D}_T^{-1}\mathbf{\Psi}^{'}\mathbf{Z}^{'}(\frac{1}{\sqrt{T}})\mathbf{Z}_1(\frac{1}{T}\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}(\frac{1}{\sqrt{T}})\mathbf{Z}_1^{'}\mathbf{Z}\mathbf{\Psi}\mathbf{D}_T^{-1}\right]\mathbf{D}_T\mathbf{\Psi}^{-1}\mathbf{\Pi} \end{aligned}$$

and

$$\mathbf{\Pi}^{'}\mathbf{Z}^{'}(\mathbf{P}_Z-\mathbf{P}_{Z_1})\mathbf{Z}\mathbf{\Pi} = \mathbf{\Pi}_2^{'}\left[\mathbf{Z}_2^{'}\mathbf{Z}_2 - \mathbf{Z}_2^{'}\mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_2\right]\mathbf{\Pi}_2\ .$$

We consider the smaller characteristic roots $0\le\nu_1\le\cdots\le\nu_r$, which satisfy (3.7) and the corresponding characteristic vectors $\boldsymbol{\beta}_i$. We can use the relation that

$$(7.21)\quad \boldsymbol{\beta}_i^{'}\mathbf{Y}^{'}\mathbf{P}_Z\mathbf{Y}\boldsymbol{\beta}_i - \nu_i\boldsymbol{\beta}_i^{'}\mathbf{Y}^{'}\bar{\mathbf{P}}_Z\mathbf{Y}\boldsymbol{\beta}_i = 0\ \ (i=1,\cdots,r)$$

is equivalent to

$$\frac{1}{T^{\epsilon}}\boldsymbol{\beta}_i^{'}\mathbf{V}^{'}\mathbf{P}_Z\mathbf{V}\boldsymbol{\beta}_i - \left[T^{1-\epsilon}\nu_i\right]\frac{1}{T}\boldsymbol{\beta}_i^{'}\mathbf{V}^{'}\bar{\mathbf{P}}_Z\mathbf{V}\boldsymbol{\beta}_i = 0$$

for any $1>\epsilon>0$. Then we have

$$\mathrm{plim}_{T\to\infty}\left[T^{1-\epsilon}\nu_i\right]\boldsymbol{\beta}_i^{'}\mathbf{\Omega}\boldsymbol{\beta}_i = 0\ ,$$

which implies the next result.

Lemma 6 : Under $\mathbf{H}_0^{'}$ and $\mathbf{\Omega}_{11}$ is nonsingular, for any $0\le\delta<1$

$$(7.22)\quad T^{\delta}\nu_i \xrightarrow{p} 0\ (i=1,\cdots,r)\ .$$

We set the corresponding characteristic vectors as a $G_1 \times r$ matrix $\hat{\mathbf{B}}_{ML}$ and apply the similar arguments for $\hat{\hat{\beta}}_{LI}$ in AK (1994). By setting a $G_1 \times r$ matrix $\hat{\mathbf{B}}$ such that

$$\mathbf{Y}^{'}\left[\mathbf{P}_Z - \mathbf{P}_{Z_1}\right]\mathbf{Y}\hat{\mathbf{B}} = \mathbf{0} \, , \tag{7.23}$$

then the limiting distribution of $\hat{\mathbf{B}}$ is the same as the limiting distribution of $\hat{\mathbf{B}}_{ML}$.

Now we decompose

$$\begin{aligned}
& \mathbf{Y}^{'}\mathbf{P}_Z\mathbf{Y} \\
= \; & \mathbf{V}^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\left[\mathbf{D}_T^{-1}\boldsymbol{\Psi}^{'}\mathbf{Z}^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\right]^{-1}\mathbf{D}_T^{-1}\boldsymbol{\Psi}\mathbf{Z}\mathbf{V} + \left[\boldsymbol{V}^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right] \\
& + \left[\boldsymbol{V}^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right]^{'} + \boldsymbol{\Pi}^{'}\boldsymbol{\Psi}^{'-1}\mathbf{D}_T\left[\mathbf{D}_T^{-1}\boldsymbol{\Psi}^{'}\mathbf{Z}^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\right]\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}
\end{aligned}$$

and

$$\begin{aligned}
& \mathbf{Y}^{'}\mathbf{P}_{Z_1}\mathbf{Y} \\
= \; & \mathbf{V}^{'}\mathbf{Z}_1(\frac{1}{\sqrt{T}})(\frac{1}{T}\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}(\frac{1}{\sqrt{T}})\mathbf{Z}_1^{'}\mathbf{V} \\
& + \left[\boldsymbol{V}^{'}(\frac{1}{\sqrt{T}})\mathbf{Z}_1(\frac{1}{T}\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}(\frac{1}{\sqrt{T}})\mathbf{Z}_1^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right] \\
& + \left[\boldsymbol{V}^{'}(\frac{1}{\sqrt{T}})\mathbf{Z}_1(\frac{1}{T}\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}(\frac{1}{\sqrt{T}})\mathbf{Z}_1^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right]^{'} \\
& + \boldsymbol{\Pi}^{'}\boldsymbol{\Psi}^{'-1}\mathbf{D}_T\left[\mathbf{D}_T^{-1}\boldsymbol{\Psi}^{'}\mathbf{Z}^{'}(\frac{1}{\sqrt{T}})\mathbf{Z}_1(\frac{1}{T}\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}(\frac{1}{\sqrt{T}})\mathbf{Z}_1^{'}\mathbf{Z}\boldsymbol{\Psi}\mathbf{D}_T^{-1}\right]\mathbf{D}_T\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi} \, .
\end{aligned}$$

We utilize

$$\boldsymbol{\Psi}^{'}\mathbf{Y}^{'}\left[\mathbf{P}_Z - \mathbf{P}_{Z_1}\right]\mathbf{Y}\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}\hat{\mathbf{B}} = \mathbf{O} \, . \tag{7.24}$$

By dividing (7.22) by $1/T$ and using the relation $\boldsymbol{\Psi}^{-1}\mathbf{B}_0 = (\mathbf{O}, \mathbf{I}_r)^{'}$ and $\mathbf{M}_*$ of the $G_1 \times G_1$ left-lower corner sub-matrix of $\mathbf{M}^*$ in Lemma 4, we find that

$$\begin{bmatrix} \mathbf{I}_{G_1-r} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{M}^* \begin{bmatrix} \mathbf{I}_{G_1-r} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \boldsymbol{\Psi}^{-1}\text{plim}_{T\to\infty}\left[\hat{\mathbf{B}} - \mathbf{B}_0\right] = \mathbf{O}$$

and

$$\text{plim}_{T\to\infty}\hat{\mathbf{B}} = \mathbf{B}_0 \, .$$

By dividing (7.24) by $1/\sqrt{T}$, we have

$$\begin{aligned}
& \begin{bmatrix} \mathbf{I}_{G_1-r} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{M}^* \begin{bmatrix} \mathbf{I}_{G_1-r} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \boldsymbol{\Phi}^{-1}\sqrt{T}\left[\hat{\mathbf{B}} - \mathbf{B}_0\right] \\
+ \; & \begin{bmatrix} \mathbf{I}_{G_1-r} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{D}_T^{-1}\boldsymbol{\Phi}^{'}\mathbf{Z}\left[\mathbf{P}_Z - \mathbf{P}_{Z_1}\right]\mathbf{V}\boldsymbol{\Phi}\boldsymbol{\Phi}^{-1}\mathbf{B}_0 = o_p(1) \, .
\end{aligned} \tag{7.25}$$

By using the fact that the limiting distribution of

$$\begin{aligned} & \mathbf{B}_0^{'}\mathbf{G}_{11}\mathbf{B}_0 - \hat{\mathbf{B}}^{'}\mathbf{G}_{11}\hat{\mathbf{B}} \\ = \ & \mathbf{B}_0^{'}\mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}(\mathbf{B}_0 - \hat{\mathbf{B}}) + (\mathbf{B}_0 - \hat{\mathbf{B}})^{'}\mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}\mathbf{B}_0 \\ & -(\mathbf{B}_0 - \hat{\mathbf{B}})^{'}\mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}(\mathbf{B}_0 - \hat{\mathbf{B}}_{LI}) \ , \end{aligned}$$

is the same as the limiting distribution of

$$(\mathbf{B}_0 - \hat{\mathbf{B}})^{'}\mathbf{Y}^{'}(\mathbf{P}_Z - \mathbf{P}_{Z_1})\mathbf{Y}(\mathbf{B}_0 - \hat{\mathbf{B}}) \ .$$

Also by (7.23), we find that

$$\left[\hat{\mathbf{B}} - \mathbf{B}_0\right]^{'}\mathbf{Y}^{'}\left[\mathbf{P}_Z - \mathbf{P}_{Z_1}\right]\mathbf{Y}\left[\hat{\mathbf{B}} - \mathbf{B}_0\right]$$

is asymptotically equivalent to

$$\left[\frac{1}{\sqrt{T}}\mathbf{U}^{*'}\mathbf{Z}_2\mathbf{\Gamma}\right]\left[\frac{1}{T}\mathbf{\Gamma}^{'}\left(\mathbf{Z}_2^{'}\mathbf{Z}_2 - \mathbf{Z}_2^{'}\mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\mathbf{Z}_2\right)\mathbf{\Gamma}\right]^{-1}\left[\frac{1}{\sqrt{T}}\mathbf{\Gamma}^{'}\mathbf{Z}_2^{'}\mathbf{U}^{*}\right] \ ,$$

where we use the notations $\mathbf{U} = \mathbf{V}\mathbf{B}$ and $\mathbf{U}^{*} = \left[\mathbf{I}_T - \mathbf{Z}_1(\mathbf{Z}_1^{'}\mathbf{Z}_1)^{-1}\mathbf{Z}_1^{'}\right]\mathbf{U}$.
Then

$$\mathrm{tr}\left(\left[\hat{\mathbf{B}} - \mathbf{B}_0\right]^{'}\mathbf{Y}^{'}\left[\mathbf{P}_Z - \mathbf{P}_{Z_1}\right]\mathbf{Y}\left[\hat{\mathbf{B}} - \mathbf{B}_0\right]\mathbf{\Sigma}^{-1}\right)$$

converges to the χ^2-distribution as $T \to \infty$ under $\mathbf{H}_0^{'}$.

Finally, we notice that as $T \to \infty$

$$\frac{1}{T}\hat{\mathbf{B}}^{'}\mathbf{H}_{11}\hat{\mathbf{B}} \xrightarrow{p} \mathbf{\Sigma} \tag{7.26}$$

and

$$\frac{1}{T}\mathbf{B}_0^{'}\mathbf{H}_{11}\mathbf{B}_0 \xrightarrow{p} \mathbf{\Sigma} \ , \tag{7.27}$$

where $\mathbf{\Sigma} = \mathbf{B}_0^{'}\mathbf{\Omega}_{11}\mathbf{B}_0$. Then we use the fact that LR_2 and LR_3 are equivalent, and $T\sum_{i=1}^{r}\log(1+\nu_i) - T\sum_{i=1}^{r}\nu_i = o_p(1)$ by using Lemma 6. Since

$$LR_3 - \mathrm{tr}\left[(\mathbf{B}_0^{'}\mathbf{G}_{11}\mathbf{B}_0 - \hat{\mathbf{B}}^{'}\mathbf{G}_{11}\hat{\mathbf{B}})\mathbf{\Sigma}^{-1}\right] = o_p(1) \ ,$$

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. (1984), "Estimating linear statistical relationships," *Annals of Statistics*, Vol. 12, 1-45.

[3] 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.

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

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

[6] 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.

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

[8] 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 (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2005).

[9] 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.

[10] 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.

[11] Andrews, D. and J. Stock (2005), "Inference with Weak Instruments," Unpublished Manuscript.

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

[13] 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).

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