Econ 203C: Systems Models

Problem Set 6

Michael Powell
Department of Economics, UCLA
May 28th, 2006

Question 1:

In this exercise we will conduct a Monte Carlo experiment that is designed to examine the *small sample* properties of the three tests discussed in Lecture Note 11, namely the Wald, Lagrange multiplier (LM), and the likelihood ratio (LR) tests.

Consider the linear regression model given by

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

Let $X_i^0 = \begin{bmatrix} X_{2i} \\ X_{3i} \\ X_{4i} \end{bmatrix}$ and it is given that

$$X_i^0 \sim N(\mu_X, \Omega_X),$$

where

$$\begin{aligned} \mu_X &= \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}, \text{ and} \\ \Omega_X &= \begin{bmatrix} 4.456 & -0.274 & 0.227 \\ -0.274 & 5.323 & 0.017 \\ 0.227 & 0.017 & 5.247 \end{bmatrix}. \end{aligned}$$

The value for the parameter vector β is given by:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \\ -0.5 \\ 0.25 \end{bmatrix}.$$

In this exercise we draw 500 samples. A draw of an observation in a sample is constructed as follows

$$X_{d_1 i} = \mu_X + P\xi_i,$$

where P is such that $PP' = \Omega_X$, and ξ_i is a draw from a standard normal distribution. Then,

$$Y_{di} = X_{di}'\beta + u_i,$$

where u_i is a draw from a student t distribution with 5 degrees of freedom and $X_{di} = \begin{bmatrix} 1 \\ X_{d_1 i}' \end{bmatrix}$.

Draw 500 samples each of 100 observations.

For each sample do the following:

(1) Compute the optimal GMM estimator for β, say $\hat{\beta}_n$, based on the moment conditions given by

$$\varphi(Y_i, X_i, \beta) = (Y_i - X_i'\beta) Z_i,$$

where

$$Z_i' = \begin{bmatrix} 1 & X_{2i} & X_{3i} & X_{4i} & X_{2i}^2 & X_{3i}^2 & X_{4i}^2 \end{bmatrix}'.$$

Solution: Recall from problem set 5 that the GMM estimator for β will be given by

$$\hat{\beta}_n = \left(X'ZV_n^{-1}Z'X\right)^{-1} X'ZV_n^{-1}Z'Y.$$

It can be shown that the optimal weight matrix is given by

$$V_n = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2$$

where $\hat{\varepsilon} = M_X Y$. This gives the following estimator

$$\begin{aligned}\hat{\beta}_n &= \left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)'\right]^{-1} \cdot \\ &\quad \left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Y_i\right)\end{aligned}$$

(2) Compute a consistent estimator for the asymptotic covariance of the optimal GMM estimate.
Solution: Recall from problem set 5 that the asymptotic covariance of the optimal GMM estimator is given by

$$\Lambda_0 = \left[\Sigma_{XZ} E_0\left[Z_i Z_i' \varepsilon_i^2\right]^{-1} \Sigma_{XZ}'\right]^{-1}$$

Since $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2 \xrightarrow{p} E_0\left[Z_i Z_i' \varepsilon_i^2\right]$ and $\frac{1}{n}\sum_{i=1}^{n} X_i Z_i' \xrightarrow{p} \Sigma_{XZ}$, we have that, by Slutsky's theorem,

$$\hat{\Lambda}_n \equiv \left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)'\right]^{-1}$$

Is a consistent estimator for Λ_0.

(3) Construct the Wald, LM, LR test statistics for the hypothesis

$$H_0 : r(\beta) = \beta_2\beta_3 + \beta_4 = 0.$$

Solution: The Wald test statistic is given by

$$W_n = nr\left(\hat{\beta}_n^{UR}\right)'\left(R\left(\hat{\beta}_n^{UR}\right)'\hat{\Lambda}_n R\left(\hat{\beta}_n^{UR}\right)\right)^{-1} r\left(\hat{\beta}_n^{UR}\right)$$

Where

$$R(\beta) = \begin{bmatrix} 0 \\ \beta_3 \\ \beta_2 \\ 1 \end{bmatrix}$$

And therefore,

$$W_n = \frac{n\left(\hat{\beta}_2^{UR}\hat{\beta}_3^{UR} + \hat{\beta}_4^{UR}\right)^2}{\begin{bmatrix} 0 & \hat{\beta}_3^{UR} & \hat{\beta}_2^{UR} & 1 \end{bmatrix}\hat{\Lambda}_n \begin{bmatrix} 0 \\ \hat{\beta}_3^{UR} \\ \hat{\beta}_2^{UR} \\ 1 \end{bmatrix}}$$

Where $\hat{\Lambda}_n$ is given above.

The Lagrange Multiplier (LM) test statistic is given by

$$LM_n = nm_n\left(\hat{\beta}_n^R\right)' W^{-1}\left(\hat{\beta}_n^R\right) A\left(\hat{\beta}_n^R\right)' \left[A\left(\hat{\beta}_n^R\right) W^{-1}\left(\hat{\beta}_n^R\right) A\left(\hat{\beta}_n^R\right)'\right]^{-1} A\left(\hat{\beta}_n^R\right) W^{-1}\left(\hat{\beta}_n^R\right) m_n\left(\hat{\beta}_n^R\right)$$

Where $\hat{\beta}_n^R$ is the restricted GMM estimator for β_0,

$$\begin{aligned}
m_n\left(\hat{\beta}_n^R\right) &\equiv \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i'\hat{\beta}_n^R\right) Z_i \\
A\left(\hat{\beta}_n^R\right) &= \frac{\partial}{\partial\beta} m_n\left(\hat{\beta}_n^R\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} X_i Z_i' \\
W^{-1}\left(\hat{\beta}_n^R\right) &= \left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \hat{\varepsilon}_i^2\right)^{-1}
\end{aligned}$$

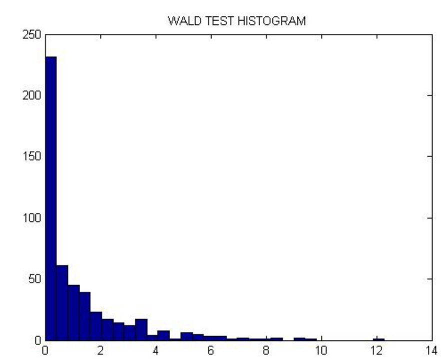
Finally, the Likelihood Ratio (LR) test statistic is given by

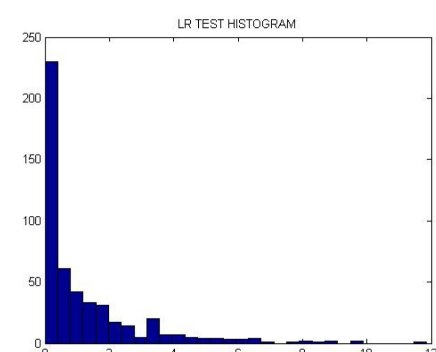
$$LR_n = n\left(m_n\left(\hat{\beta}_n^R\right)' W\left(\hat{\beta}_n^R\right) m_n\left(\hat{\beta}_n^R\right) - m_n\left(\hat{\beta}_n^{UR}\right)' W\left(\hat{\beta}_n^{UR}\right) m_n\left(\hat{\beta}_n^{UR}\right)\right)$$

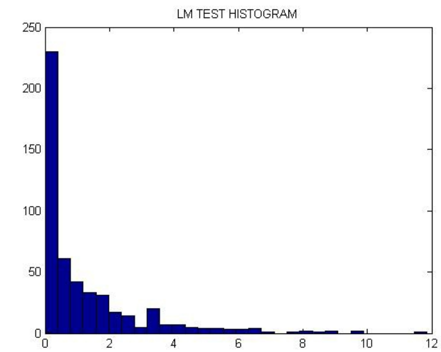
Where everything is defined as above.

(4) Store (for later use) the three test statistics obtained in (3).
Solution: Sounds like a plan.

(5) Once you are done with (1) through (4) for each of the 500 samples drawn, plot a histogram for each of the 500 Wald, LM, and LR statistics obtained for the 500 samples.
Solution: Attached is the MATLAB code that generates the following histograms







(6) Discuss briefly the results obtained in (5). In particular compute the actual (empirical) type I error that one would get for each of the above three statistics.

Solution: From MATLAB, we have that the probability of type one error is 0.1156 for the Wald test, 0.1239 for the Lagrange Multiplier test, and 0.0726 for the Likelihood Ratio test. That is, for this particular hypothesis, for this particular estimator, and for this particular sample size, the Lagrange Multiplier test rejects the null (which is indeed true) more often than does the Wald test which does so more often than does the Likelihood Ratio test. We know that all three of these tests are asymptotically equivalent, but this example shows quite well that asymptotic properties do not hold in the finite sample case. In this case, we reject the null under the Lagrange Multiplier test almost twice as often as we do under the Likelihood Ratio test!

Question 2:
Consider the model given by

$$Y_i = g(X_i;\theta_0) + u_i,$$
$$E[u_i|X_i] = 0,$$

for $i = 1, \ldots, n$ where the parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^K$, X_i is a $K \times 1$ vector of regressors, and $g(\cdot)$ is a known non-linear function.

(1) Show that the population parameter vector is obtained as a solution to

$$\min_{\theta\in\Theta} E\left[(Y_i - g(X_i;\theta))^2\right].$$

Solution: First, assume that $g(X_i;\theta)$ is differentiable. Then, making the necessary assumptions to be able to change the order of differentiation and integration (i.e. assuming that $(Y_i - g(X_i;\theta))^2$ is locally uniformly continuous in θ), we have the following first order conditions (clearly the first order necessary conditions are also sufficient conditions since $(Y_i - g(X_i;\theta))^2$ is a convex function in θ.)

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta} E\left[(Y_i - g(X_i;\theta))^2\right] = E\left[\frac{\partial}{\partial\theta}(Y_i - g(X_i;\theta))^2\right] \\ &= -E\left[2(Y_i - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\right] \end{aligned}$$

Or

$$\begin{aligned} 0 &= 2E\left[(Y_i - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\right] \\ &= 2E\left[E\left[(Y_i - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\middle| X_i\right]\right] \\ &= 2E\left[(E[Y_i|X_i] - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\right] \\ &= 2E\left[(g(X_i;\theta_0) - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\right] \end{aligned}$$

Where the last equality holds because $E[u_i|X_i] = 0$ implies that $E[Y_i|X_i] = g(X_i;\theta_0)$. This equality holds when $\theta = \theta_0$. (i.e. θ_0 is a solution to the above problem.)

(2) Provide the conditions that make the parameter vector θ_0 unique.
Solution: Let $g(X_i;\theta) \neq g(X_i;\theta_0)$ for all $\theta \neq \theta_0$ and let $\frac{\partial}{\partial\theta} g(X_i;\theta) \neq 0$ for all $\theta \neq \theta_0$. Then

$$2E\left[(g(X_i;\theta_0) - g(X_i;\theta))\frac{\partial}{\partial\theta} g(X_i;\theta)\right] = 0$$

If and only if $\theta = \theta_0$ and therefore θ_0 is unique.

Note that this is an unnecessarily strong condition for the uniqueness of θ_0. We usually assume the identification condition, which is sufficient: $g(X_i;\theta) \neq g(X_i;\theta_0)$ for all $\theta \neq \theta_0$. Here, however, in order to be consistent with the manner in which θ_0 was derived in part (1), I also made the first derivative condition assumption, since otherwise the equation may have multiple solutions. (i.e. any θ satisfying $\frac{\partial}{\partial\theta} g(X_i;\theta) = 0$ would work.)

(3) Provide the sample analog of the population moments from which one can obtain an estimator for θ_0. Denote these moment conditions by $\varphi_1(Y, X;\theta)$.
Solution: Since $E[u_i|X_i] = 0$, we have that $E[Y_i|X_i] = g(X_i;\theta_0)$ or

$$E[Y_i - g(X_i;\theta_0)|X_i] = 0$$

By the law of iterated expectations, we have

$$E\left[Y_i - g\left(X_i;\theta_0\right)\right] = 0$$

Define $\varphi_1\left(Y_i, X_i;\theta\right) \equiv Y_i - g\left(X_i;\theta\right)$. Let

$$m_n^1\left(\theta\right) \equiv \frac{1}{n}\sum_{i=1}^{n}\varphi_1\left(Y_i, X_i;\theta\right)$$

And

$$Q_n^1\left(\theta\right) = m_n^1\left(\theta\right)' V_n^{-1} m_n^1\left(\theta\right)$$

The GMM estimator is therefore obtained as a solution to

$$\min_{\theta\in\Theta} Q_n^1\left(\theta\right).$$

(4) Suggest additional K moment conditions to those in (3). Denote the additional moment conditions by $\varphi_2\left(Y, X;\theta\right)$. Show that

$$E\left[\varphi_2\left(Y, X;\theta_0\right)\right] = 0.$$

Solution: Since $E\left[\left.u_i\right| X_i\right] = 0$, we have that $E\left[\left.Y_i\right| X_i\right] = g\left(X_i;\theta_0\right)$ or

$$\begin{aligned} E\left[\left.Y_i - g\left(X_i;\theta_0\right)\right| X_i\right] &= 0 \\ E\left[\left.Y_i - g\left(X_i;\theta_0\right)\right| X_i\right] X_i &= 0 \\ E\left[\left.\left(Y_i - g\left(X_i;\theta_0\right)\right) X_i\right| X_i\right] &= 0 \end{aligned}$$

Define $\varphi_2\left(Y_i, X_i;\theta\right) \equiv \left(Y_i - g\left(X_i;\theta\right)\right) X_i$. Let

$$m_n^2\left(\theta\right) \equiv \frac{1}{n}\sum_{i=1}^{n}\varphi_2\left(Y_i, X_i;\theta\right)$$

And

$$Q_n^2\left(\theta\right) = m_n^2\left(\theta\right)' V_n^{-1} m_n^2\left(\theta\right).$$

The GMM estimator is therefore obtained as a solution to

$$\min_{\theta\in\Theta} Q_n^2\left(\theta\right).$$

(5) Suggest an optimal GMM estimator for θ_0 based on $\varphi_1\left(Y, X;\theta\right)$ and $\varphi_2\left(Y, X;\theta\right)$ from (3) and (4).

Solution: Let $Z_i = \begin{cases} \begin{bmatrix} 1 \\ X_i \end{bmatrix} & \text{if } X \text{ contains a column of ones} \\ X_i & \text{if } X \text{ does not contain a column of ones} \end{cases}$

Then we can combine the moment conditions $\varphi_1\left(Y, X;\theta\right)$ and $\varphi_2\left(Y, X;\theta\right)$ into the following

$$\varphi_3\left(Y_i, X_i;\theta\right) = \left(Y_i - g\left(X_i;\theta\right)\right) Z_i$$

Let

$$m_n^3\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n}\varphi_3\left(Y_i, X_i;\theta\right)$$

And

$$Q_n^3\left(\theta\right) = m_n^3\left(\theta\right)' V_n^{-1} m_n^3\left(\theta\right)$$

Define the GMM estimator by

$$\hat{\theta}_{n,GMM}^3 = \arg\min_{\theta\in\Theta} Q_n^3\left(\theta\right)$$

As shown in problem set 5, the asymptotic distribution of $\hat{\theta}^{3}_{n,GMM}$ is given by

$$\sqrt{n}\left(\hat{\theta}^{3}_{n,GMM}-\theta_0\right)\xrightarrow{d}N\left(0,B\left(\theta_0\right)\Lambda\left(\theta_0\right)B\left(\theta_0\right)'\right)$$

Where

$$\begin{aligned}\Lambda\left(\theta_0\right) &= E_0\left[\varphi_3\left(Y_i,X_i;\theta_0\right)\varphi\left(Y_i,X_i;\theta_0\right)'\right]\\ &= E_0\left[\left(Y_i-g\left(X_i;\theta_0\right)\right)Z_iZ_i'\left(Y_i-g\left(X_i;\theta_0\right)\right)'\right]\\ &= E_0\left[Z_iZ_i'\left(Y_i-g\left(X_i;\theta_0\right)\right)^2\right]\end{aligned}$$

$$\begin{aligned}A\left(\theta_0\right) &= E_0\left[\frac{\partial\varphi_3\left(Y_i,X_i;\theta_0\right)}{\partial\theta}\right]\\ &= E_0\left[-\frac{\partial}{\partial\theta}g\left(X_i;\theta_0\right)Z_i'\right]\end{aligned}$$

$$\begin{aligned}B\left(\theta_0\right) &= \left[-A\left(\theta_0\right)\left(V^{-1}\right)A\left(\theta_0\right)'\right]^{-1}A\left(\theta_0\right)\left(V^{-1}\right)'\\ &= \left[-E_0\left[-\frac{\partial}{\partial\theta}g\left(X_i;\theta_0\right)Z_i'\right]V^{-1}E_0\left[-\frac{\partial}{\partial\theta}g\left(X_i;\theta_0\right)Z_i'\right]'\right]^{-1}E_0\left[-\frac{\partial}{\partial\theta}g\left(X_i;\theta_0\right)Z_i'\right]\left(V^{-1}\right)'\end{aligned}$$

Next, let

$$V_n^*\equiv\frac{1}{n}\sum_{i=1}^{n}Z_iZ_i'\left(Y_i-g\left(X_i;\hat{\theta}^{3}_{n,GMM}\right)\right)^2$$

Then

$$V_n^*\xrightarrow{p}E_0\left[Z_iZ_i'\left(Y_i-g\left(X_i;\theta_0\right)\right)^2\right]\equiv V^*$$

Define

$$Q_n^{3*}\left(\theta\right)\equiv m_n^3\left(\theta\right)'\left(V_n^*\right)^{-1}m_n^3\left(\theta\right)$$

The optimal GMM estimator based on $\varphi_3\left(Y_i,X_i;\theta\right)$ is therefore given by

$$\hat{\theta}^{3*}_{n,GMM}=\arg\min_{\theta\in\Theta}Q_n^{3*}\left(\theta\right)$$

(6) Provide the asymptotic distribution for the estimator suggested in (5).
Solution: Half the work for this is already done. As shown in part (5), we have that

$$\sqrt{n}\left(\hat{\theta}^{3}_{n,GMM}-\theta_0\right)\xrightarrow{d}N\left(0,B\left(\theta_0\right)\Lambda\left(\theta_0\right)B\left(\theta_0\right)'\right)$$

Where

$$\Lambda\left(\theta_0\right)=E_0\left[Z_iZ_i'\left(Y_i-g\left(X_i;\theta_0\right)\right)^2\right]$$

$$B\left(\theta_0\right)=\left[-A\left(\theta_0\right)V^{-1}A\left(\theta_0\right)'\right]^{-1}A\left(\theta_0\right)\left(V^{-1}\right)'$$

And

$$A\left(\theta_0\right)=E_0\left[-\frac{\partial}{\partial\theta}g\left(X_i;\theta_0\right)Z_i'\right]$$

For the optimal estimator suggested in (5), we have that

$$\sqrt{n}\left(\hat{\theta}^{3*}_{n,GMM}-\theta_0\right)\xrightarrow{d}N\left(0,B\left(\theta_0\right)\Lambda\left(\theta_0\right)B\left(\theta_0\right)'\right)$$

is such that

$$\begin{aligned}
& B(\theta_0)\Lambda(\theta_0)B(\theta_0) \\
= & \left[-A(\theta_0)\Lambda(\theta_0)^{-1}A(\theta_0)'\right]^{-1} A(\theta_0)\Lambda(\theta_0)^{-1}\Lambda(\theta_0)\Lambda(\theta_0)^{-1}A(\theta_0)\left[-A(\theta_0)\Lambda(\theta_0)^{-1}A(\theta_0)\right]^{-1} \\
= & \left[A(\theta_0)\Lambda(\theta_0)^{-1}A(\theta_0)'\right]^{-1} \\
= & \left[E_0\left[-\frac{\partial}{\partial\theta}g(X_i;\theta_0)Z_i'\right]E_0\left[Z_iZ_i'(Y_i-g(X_i;\theta_0))^2\right]^{-1}E_0\left[-\frac{\partial}{\partial\theta}g(X_i;\theta_0)Z_i'\right]'\right]^{-1}
\end{aligned}$$

Which is analogous to the "sandwich form" of the generalized linear regression model.

(7) Provide the Wald, LM, and LR test statistics for the null hypothesis

$$H_0 : \prod_{k=1}^{K}\theta_{k0} = 1.$$

Solution: For all that follows, I will use the estimator defined in part (5). Recall from pages 8-10 in lecture notes 11 that for a general hypothesis of the form $H_0 : r(\theta) = 0$, the Wald, LM, and LR test statistics in the GMM framework are given by

$$W_n = nr\left(\hat{\theta}_n^{UR}\right)'\left(R\left(\hat{\theta}_n^{UR}\right)'\hat{\Lambda}_nR\left(\hat{\theta}_n^{UR}\right)\right)^{-1}r\left(\hat{\theta}_n^{UR}\right)$$

$$LM_n = nm_n\left(\hat{\theta}_n^{R}\right)'V_n^{-1}\frac{\partial}{\partial\theta'}m_n\left(\hat{\theta}_n^{R}\right)\left[A\left(\hat{\theta}_n^{R}\right)W^{-1}\left(\hat{\theta}_n^{R}\right)A\left(\hat{\theta}_n^{R}\right)'\right]^{-1}\frac{\partial}{\partial\theta}m_n\left(\hat{\theta}_n^{R}\right)V_n^{-1}m_n\left(\hat{\theta}_n^{R}\right)$$

$$LR_n = n\left(m_n\left(\hat{\theta}_n^{R}\right)'W\left(\hat{\theta}_n^{R}\right)m_n\left(\hat{\theta}_n^{R}\right)-m_n\left(\hat{\theta}_n^{UR}\right)'W\left(\hat{\theta}_n^{UR}\right)m_n\left(\hat{\theta}_n^{UR}\right)\right)$$

Where

$$\hat{\theta}_n^{UR} \equiv \arg\min_{\theta\in\Theta} Q_n^{3*}(\theta)$$

$$\hat{\theta}_n^{R} \equiv \arg\min_{\theta\in\Theta:r(\theta)=0} Q_n^{3*}(\theta)$$

$$r(\theta) \equiv \prod_{k=1}^{K}\theta_k - 1$$

$$R(\theta) \equiv \frac{\partial}{\partial\theta}r(\theta) = \begin{bmatrix} \prod_{k\neq1}\theta_k \\ \vdots \\ \prod_{k\neq K}\theta_k \end{bmatrix}$$

$$\hat{\Lambda}_n \equiv \left\{\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}g(X_i;\theta_0)Z_i'\right]\left[\frac{1}{n}\sum_{i=1}^{n}Z_iZ_i'(Y_i-g(X_i;\theta_0))^2\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}g(X_i;\theta_0)Z_i'\right]'\right\}^{-1}$$

The Wald statistic can be shown explicitly to be

$$W_n = n\frac{\left(\prod_{k=1}^{K}\hat{\theta}_{n,k}^{UR}-1\right)^2}{\begin{bmatrix} \prod_{k\neq1}\hat{\theta}_{n,k} \\ \vdots \\ \prod_{k\neq K}\hat{\theta}_{n,k} \end{bmatrix}'\hat{\Lambda}_n\begin{bmatrix} \prod_{k\neq1}\hat{\theta}_{n,k} \\ \vdots \\ \prod_{k\neq K}\hat{\theta}_{n,k} \end{bmatrix}}$$

Without more specific information regarding the functional form of g, it is not possible to characterize explicitly the Lagrange Multiplier and Likelihood Ratio test statistics.