Econ 620

Three Classical Tests; Wald, LM(Score), and LR tests

Suppose that we have the density $\ell(y;\theta)$ of a model with the null hypothesis of the form $H_0; \theta = \theta_0$. Let $L(\theta)$ be the log-likelihood function of the model and $\widehat{\theta}$ be the MLE of θ.

Wald test is based on the very intuitive idea that we are willing to accept the null hypothesis when $\widehat{\theta}$ is close to θ_0. The distance between $\widehat{\theta}$ and θ_0 is the basis of constructing the test statistic. On the other hand, consider the following constrained maximization problem,

$$\max_{\theta\in\Theta} L(\theta) \qquad \text{s.t. } \theta = \theta_0$$

If the constraint is not binding (the null hypothesi is true), the Lagrangian multiplier associated with the constraint is zero. We can construct a test measuring how far the Lagrangian multiplier is from zero. - LM test. Finally, another way to check the validity of null hypothesis is to check the distance between two values of maximum likelihood function like

$$L\left(\widehat{\theta}\right) - L(\theta_0) = \log \frac{\ell\left(y;\widehat{\theta}\right)}{\ell(y;\theta_0)}$$

If the null hypothesis is true, the above statistic should not be far away from zero, again.

Asymptotic Distributions of the Three Tests

Assume that the observed variables can be partitioned into the endogenous variables X and exogenous variables Y. To simplify the presentation, we assume that the observations (Y_i, X_i) are i.i.d. and we can obtain conditional distribution of endogenous variables given the exogenous variables as $f(y_i \mid x_i;\theta)$ with $\theta \in \Theta \subseteq R^p$. The conditional density is known up to unknown parameter vector θ. By i.i.d. assumption, we can write down the log-likelihood function of n observations of (Y_i, X_i) as

$$L(\theta) = \sum_{i=1}^{n} \log f(y_i \mid x_i;\theta)$$

We assume all the regularity conditions for existence, consistency and asymptotic normality of MLE and denote MLE as $\widehat{\theta}_n$. The hypotheses of interest are given as

$$H_0; g(\theta_0) = 0 \qquad\qquad H_A; g(\theta_0) \neq 0$$

where $g(\cdot); R^p \to R^r$ and the rank of $\frac{\partial g}{\partial \theta}$ is r.

Wald test

Proposition 1

$$\xi_n^W = ng'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\widehat{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widehat{\theta}_n\right)\frac{\partial g'\left(\widehat{\theta}_n\right)}{\partial \theta}\right)^{-1} g\left(\widehat{\theta}_n\right) \sim \chi^2(r) \qquad \textit{under } H_0.$$

where $\mathcal{I} = E_X E_\theta\left(-\frac{\partial^2 \log f(Y|X;\theta)}{\partial\theta\partial\theta'}\right)$ *and* $\mathcal{I}^{-1}\left(\widehat{\theta}_n\right)$ *is the inverse of* $\mathcal{I}$ *evaluated at* $\theta = \widehat{\theta}_n$.

$\Rightarrow$ From the asymptotic characteristics of MLE, we know that

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}(\theta_0)\right) \tag{1}$$

The first order Taylor series expansion of $g\left(\widehat{\theta}_n\right)$ around the true value of θ_0, we have

$$g\left(\widehat{\theta}_n\right) = g\left(\theta_0\right) + \frac{\partial g\left(\theta_0\right)}{\partial \theta'}\left(\widehat{\theta}_n - \theta_0\right) + o_p\left(1\right)$$

$$\sqrt{n}\left(g\left(\widehat{\theta}_n\right) - g\left(\theta_0\right)\right) = \frac{\partial g\left(\theta_0\right)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) + o_p\left(1\right) \tag{2}$$

Hence, combining (1) and (2) gives

$$\sqrt{n}\left(g\left(\widehat{\theta}_n\right) - g\left(\theta_0\right)\right) \xrightarrow{d} N\left(0, \frac{\partial g\left(\theta_0\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial \theta}\right) \tag{3}$$

Under the null hypothesis, we have $g\left(\theta_0\right) = 0$. Therefore,

$$\sqrt{n}g\left(\widehat{\theta}_n\right) \xrightarrow{d} N\left(0, \frac{\partial g\left(\theta_0\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial \theta}\right) \tag{4}$$

By forming the quadratic form of the normal random variables, we can conclude that

$$ng'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\theta_0\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial \theta}\right)^{-1} g\left(\widehat{\theta}_n\right) \sim \chi^2\left(r\right) \qquad \text{under } H_0. \tag{5}$$

The statistic in (5) is useless since it depends on the unknown parameter θ_0. However, we can consistently approximate the terms in inverse bracket by evaluating at MLE, $\widehat{\theta}_n$. Therefore,

$$\xi_n^W = ng'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\widehat{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widehat{\theta}_n\right)\frac{\partial g'\left(\widehat{\theta}_n\right)}{\partial \theta}\right)^{-1} g\left(\widehat{\theta}_n\right) \sim \chi^2\left(r\right) \qquad \text{under } H_0.$$

- An asymptotic test which rejects the null hypothesis with probability one when the alternative hypothesis is true is called a *consistent test*. Namely, a consistent test has asymptotic power of 1.
- The Wald test we discussed above is a consistent test. A heuristic argument is that if the alternative hypothesis is true instead of the null hypothesis, $g\left(\widehat{\theta}_n\right) \xrightarrow{p} g\left(\theta_0\right) \neq 0$. Therefore, $g'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\widehat{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widehat{\theta}_n\right)\frac{\partial g'\left(\widehat{\theta}_n\right)}{\partial \theta}\right)^{-1} g\left(\widehat{\theta}_n\right)$ is converging to a constant instead of zero. By multiplying a constant by n, $\xi_n^W \to \infty$ as $n \to \infty$, which implies that we always reject the null hypothesis when the alternative is true.
- Another form of the Wald test statistic is given by - *caution*: this is quite confusing -

$$\xi_n^W = g'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\widehat{\theta}_n\right)}{\partial \theta'}\mathcal{I}_n^{-1}\left(\widehat{\theta}_n\right)\frac{\partial g'\left(\widehat{\theta}_n\right)}{\partial \theta}\right)^{-1} g\left(\widehat{\theta}_n\right) \sim \chi^2\left(r\right) \qquad \text{under } H_0.$$

  where $\mathcal{I}_n = E_X E_\theta\left(-\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'}\right) = E_X E_\theta\left(-\sum_{i=1}^n \frac{\partial^2 \log(y_i|x_i;\theta)}{\partial \theta \partial \theta'}\right)$ and $\mathcal{I}_n^{-1}\left(\widehat{\theta}_n\right)$ is the inverse of $\mathcal{I}_n$ evaluated at $\theta = \widehat{\theta}_n$. Note that $\mathcal{I}_n = n\mathcal{I}$.
- A quite common form of the null hypothesis is the zero restriction on a subset of parameters, i.e.,

$$H_0; \theta_1 = 0 \qquad\qquad H_A; \theta_1 \neq 0$$

  where θ_1 is a $(q \times 1)$ subvector of θ with $q < p$. Then, the Wald statistic is given by

$$\xi_n^W = n\widehat{\theta}_1'\left(\mathcal{I}^{11}\left(\widehat{\theta}_n\right)\right)^{-1}\widehat{\theta}_1 \sim \chi^2\left(q\right) \qquad \text{under } H_0.$$

  where $\mathcal{I}^{11}\left(\theta\right)$ is the upper left block of the inverse information matrix,

$$\mathcal{I}\left(\theta\right) = \begin{bmatrix} \mathcal{I}_{11}\left(\theta\right) & \mathcal{I}_{12}\left(\theta\right) \\ \mathcal{I}_{21}\left(\theta\right) & \mathcal{I}_{22}\left(\theta\right) \end{bmatrix}$$

  then, $\mathcal{I}^{11}\left(\theta\right) = \left(\mathcal{I}_{11}\left(\theta\right) - \mathcal{I}_{12}\left(\theta\right)\mathcal{I}_{22}^{-1}\left(\theta\right)\right)^{-1}$ by the formula for partitioned inverse. $\mathcal{I}^{11}\left(\widehat{\theta}_n\right)$ is $\mathcal{I}^{11}\left(\theta\right)$ evaluated at MLE.

LM test (Score test)

If we have a priori reason or evidence to believe that the parameter vector satisfies some restrictions in the form of $g(\theta) = 0$, incorporating the information into the maximization of the likelihood function through constrained optimization will improve the efficiency of estimator compared to MLE from unconstrained maximization. We solve the following problem;

$$\max L(\theta) \qquad\qquad s.t. g(\theta) = 0$$

FOC's are given by

$$\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta} + \frac{\partial g'\left(\widetilde{\theta}_n\right)}{\partial \theta}\widetilde{\lambda} = 0 \tag{6}$$

$$g\left(\widetilde{\theta}_n\right) = 0 \tag{7}$$

where $\widetilde{\theta}_n$ is the solution of constrained maximization problem called constrained MLE and λ is the vector of Lagrange multiplier. The LM test is based on the idea that properly scaled $\widetilde{\lambda}$ has an asymptotically normal distribution.

Proposition 2

$$\begin{aligned}
\xi_n^S &= \frac{1}{n}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widetilde{\theta}_n\right)\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta} \\
&= \frac{1}{n}\widetilde{\lambda}'\frac{\partial g\left(\widetilde{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widetilde{\theta}_n\right)\frac{\partial g'\left(\widetilde{\theta}_n\right)}{\partial \theta}\widetilde{\lambda} \sim \chi^2(r) \qquad \textit{under } H_0.
\end{aligned}$$

$\Rightarrow$ First order Taylor expansions of $g\left(\widehat{\theta}_n\right)$ and $g\left(\widetilde{\theta}_n\right)$ around θ_0 gives, ignoring $o_p(1)$ terms,

$$\sqrt{n}g\left(\widehat{\theta}_n\right) = \sqrt{n}g(\theta_0) + \frac{\partial g(\theta_0)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \tag{8}$$

$$\sqrt{n}g\left(\widetilde{\theta}_n\right) = \sqrt{n}g(\theta_0) + \frac{\partial g(\theta_0)}{\partial \theta'}\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) \tag{9}$$

Note that $g\left(\widetilde{\theta}_n\right) = 0$ from (7) and substracting (9) from (8), we have

$$\sqrt{n}g\left(\widehat{\theta}_n\right) = \frac{\partial g(\theta_0)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \tag{10}$$

On the other hand, taking first order Taylor series expansions of $\frac{\partial L(\widehat{\theta}_n)}{\partial \theta}$ and $\frac{\partial L(\widetilde{\theta}_n)}{\partial \theta}$ around θ_0 gives, ignoring $o_p(1)$ terms,

$$\begin{aligned}
\frac{\partial L\left(\widehat{\theta}_n\right)}{\partial \theta} &= \frac{\partial L(\theta_0)}{\partial \theta} + \frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'}\left(\widehat{\theta}_n - \theta_0\right) \Rightarrow \\
\frac{1}{\sqrt{n}}\frac{\partial L\left(\widehat{\theta}_n\right)}{\partial \theta} &= \frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial \theta} + \frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \Rightarrow \\
\frac{1}{\sqrt{n}}\frac{\partial L\left(\widehat{\theta}_n\right)}{\partial \theta} &= \frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)
\end{aligned} \tag{11}$$

note that $-\frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'} = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log(y_i|x_i;\theta)}{\partial \theta \partial \theta'} \xrightarrow{p} \mathcal{I}(\theta_0)$ by the law of large numbers. Similarly,

$$\frac{1}{\sqrt{n}}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta} = \frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial \theta} - \mathcal{I}(\theta_0)\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) \tag{12}$$

Considering the fact that $\frac{\partial L(\widehat{\theta}_n)}{\partial\theta} = 0$ by FOC of the unconstrained maximization problem, we take the difference between (11) and (12). Then,

$$\frac{1}{\sqrt{n}}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial\theta} = -\mathcal{I}\left(\theta_0\right)\sqrt{n}\left(\widetilde{\theta}_n - \widehat{\theta}_n\right) = \mathcal{I}\left(\theta_0\right)\sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \tag{13}$$

Hence,

$$\sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) = \mathcal{I}^{-1}\left(\theta_0\right)\frac{1}{\sqrt{n}}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial\theta} \tag{14}$$

From (10) and (14), we obtain

$$\sqrt{n}g\left(\widehat{\theta}_n\right) = \frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{1}{\sqrt{n}}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial\theta}$$

Using (6), we deduce

$$\begin{aligned}\sqrt{n}g\left(\widehat{\theta}_n\right) &= -\frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\widetilde{\theta}_n\right)}{\partial\theta}\frac{\widetilde{\lambda}}{\sqrt{n}} \\ &\to -\frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial\theta}\frac{\widetilde{\lambda}}{\sqrt{n}} \end{aligned} \tag{15}$$

since $\widetilde{\theta}_n \xrightarrow{p} \theta_0$ hence $g\left(\widetilde{\theta}_n\right) \xrightarrow{p} g\left(\theta_0\right)$. Therefore,

$$\frac{\widetilde{\lambda}}{\sqrt{n}} = -\left(\frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial\theta}\right)^{-1}\sqrt{n}g\left(\widehat{\theta}_n\right) \tag{16}$$

From (4), under the null hypothesis, $\sqrt{n}g\left(\widehat{\theta}_n\right) \xrightarrow{d} N\left(0, \frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'(\theta_0)}{\partial\theta}\right)$. Consequently, we have

$$\frac{\widetilde{\lambda}}{\sqrt{n}} \xrightarrow{d} N\left(0, \left(\frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial\theta}\right)^{-1}\right) \tag{17}$$

Again, forming the quadratic form of the normal random variables, we obtain

$$\frac{1}{n}\widetilde{\lambda}'\left(\frac{\partial g\left(\theta_0\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial\theta}\right)\widetilde{\lambda} \sim \chi^2\left(r\right) \qquad \text{under } H_0. \tag{18}$$

Alternatively, using (6), another form of the test statistic is given by

$$\frac{1}{n}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial\theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial\theta} \sim \chi^2\left(r\right) \qquad \text{under } H_0 \tag{19}$$

Note that (18) and (19) are useless since they depend on the unknown parameter value θ_0. We can evaluate the terms involved in θ_0 at the *constrained MLE*, $\widetilde{\theta}_n$ to get a usable statistic.

- Again, another form of LM test is $\xi_n^S = \frac{\partial L(\widetilde{\theta}_n)}{\partial\theta'}\mathcal{I}_n^{-1}\left(\theta_0\right)\frac{\partial L(\widetilde{\theta}_n)}{\partial\theta} = \widetilde{\lambda}'\left(\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}_n^{-1}\left(\theta_0\right)\frac{\partial g'(\theta_0)}{\partial\theta}\right)\widetilde{\lambda}$.
- We can approximate $\mathcal{I}\left(\theta_0\right)$ with either $-\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial\theta\partial\theta'}$ or $\frac{1}{n}\sum_{i=1}^n \frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial\theta}\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial\theta'}$. If

we choose the second approximation, the LM test statistic becomes

$$\begin{aligned}
\xi_n^S &= \frac{1}{n}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta'}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\right)^{-1}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\right)^{-1}\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta} \\
&= \sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\left(\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\right)^{-1}\sum_{i=1}^{n}\frac{\partial \log\left(y_i \mid x_i;\widetilde{\theta}_n\right)}{\partial \theta}
\end{aligned}$$

this expression seems quite familiar to us - looks like a projection matrix -. The intuition is correct. The (uncentered) R_u^2 from the regression of 1 on $\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial \theta}$ is given by

$$R_u^2 = \frac{\mathbf{1}'X\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \frac{\mathbf{1}'X\left(X'X\right)^{-1}X'\mathbf{1}}{\mathbf{1}'\mathbf{1}}$$

where $\underset{(n\times p)}{X} = \left[\begin{array}{cccc} \frac{\partial \log\left(y_1|x_1;\widetilde{\theta}_n\right)}{\partial \theta'} & \frac{\partial \log\left(y_2|x_2;\widetilde{\theta}_n\right)}{\partial \theta'} & \ldots & \frac{\partial \log\left(y_n|x_n;\widetilde{\theta}_n\right)}{\partial \theta'} \end{array}\right]'$ and $\underset{(n\times 1)}{\mathbf{1}} = \left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \end{array}\right]'$. Then,

$$R_u^2 = \frac{\sum_{i=1}^{n}\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\left(\sum_{i=1}^{n}\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial \theta}\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial \theta'}\right)^{-1}\sum_{i=1}^{n}\frac{\partial \log\left(y_i|x_i;\widetilde{\theta}_n\right)}{\partial \theta}}{n}$$

Hence,

$$\xi_n^S = nR_u^2$$

This is quite an interesting result since the computation of LM statistic is nothing but an OLS regression. We regress 1 on the *scores* evaluated at constrained MLE and compute uncentered R^2 and then multiply it with the number of observations to get LM statistic. One thing to be cautious is that most software will automatically try to print out centered R^2, which is impossible in this case since the denominator of centered R^2 is simply zero.

- LM test is also an asymptotically consistent test.
- From (16) and (18),

$$\begin{aligned}
\xi_n^W &= ng'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\widehat{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\widehat{\theta}_n\right)\frac{\partial g'\left(\widehat{\theta}_n\right)}{\partial \theta}\right)^{-1}g\left(\widehat{\theta}_n\right) \\
&\to ng'\left(\widehat{\theta}_n\right)\left(\frac{\partial g\left(\theta_0\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial g'\left(\theta_0\right)}{\partial \theta}\right)^{-1}g\left(\widehat{\theta}_n\right) = \xi_n^S
\end{aligned}$$

Likelihood ratio(LR) test

Proposition 3

$$\xi_n^R = 2\left(L\left(\widehat{\theta}_n\right) - L\left(\widetilde{\theta}_n\right)\right) \sim \chi^2\left(r\right) \qquad \textit{under } H_0.$$

$\Rightarrow$ We consider the second order Taylor expansions of $L\left(\widehat{\theta}_n\right)$ and $L\left(\widetilde{\theta}_n\right)$ around θ_0. Under the null hypothesis, ignoring stochastically dominated terms,

$$
\begin{aligned}
L\left(\widehat{\theta}_n\right) &= L\left(\theta_0\right) + \frac{\partial L\left(\theta_0\right)}{\partial \theta'}\left(\widehat{\theta}_n - \theta_0\right) + \frac{1}{2}\left(\widehat{\theta}_n - \theta_0\right)' \frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\left(\widehat{\theta}_n - \theta_0\right) \\
&= L\left(\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial L\left(\theta_0\right)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) + \frac{1}{2}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)' \frac{1}{n}\frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \\
L\left(\widetilde{\theta}_n\right) &= L\left(\theta_0\right) + \frac{\partial L\left(\theta_0\right)}{\partial \theta'}\left(\widetilde{\theta}_n - \theta_0\right) + \frac{1}{2}\left(\widetilde{\theta}_n - \theta_0\right)' \frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\left(\widetilde{\theta}_n - \theta_0\right) \\
&= L\left(\theta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial L\left(\theta_0\right)}{\partial \theta'}\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) + \frac{1}{2}\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right)' \frac{1}{n}\frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right)
\end{aligned}
$$

Taking differences and multiplying by 2, we obtain

$$
\begin{aligned}
2\left(L\left(\widehat{\theta}_n\right) - L\left(\widetilde{\theta}_n\right)\right) &= \frac{2}{\sqrt{n}}\frac{\partial L\left(\theta_0\right)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) + \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)' \frac{1}{n}\frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad - \sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right)' \frac{1}{n}\frac{\partial^2 L\left(\theta_0\right)}{\partial \theta \partial \theta'}\sqrt{n}\left(\widetilde{\theta}_n - \theta_0\right) \\
&\quad \to 2n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) - n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad + n\left(\widetilde{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widetilde{\theta}_n - \theta_0\right)
\end{aligned}
$$

since $\frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial \theta'} = \mathcal{I}\left(\theta_0\right)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)$ from (11) and $-\frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} \mathcal{I}\left(\theta_0\right)$. Continuing the derivation,

$$
\begin{aligned}
2\left(L\left(\widehat{\theta}_n\right) - L\left(\widetilde{\theta}_n\right)\right) &= 2n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) - n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad + n\left(\widetilde{\theta}_n - \widehat{\theta}_n + \widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widetilde{\theta}_n - \widehat{\theta}_n + \widehat{\theta}_n - \theta_0\right) \\
&= 2n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) - n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad + n\left(\widetilde{\theta}_n - \widehat{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widetilde{\theta}_n - \widehat{\theta}_n\right) + n\left(\widetilde{\theta}_n - \widehat{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad + n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widetilde{\theta}_n - \widehat{\theta}_n\right) + n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) \\
&= 2n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) + n\left(\widehat{\theta}_n - \widetilde{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \\
&\quad - n\left(\widehat{\theta}_n - \widetilde{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right) - n\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \\
&= n\left(\widehat{\theta}_n - \widetilde{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \qquad (20)
\end{aligned}
$$

note that $\left(\widehat{\theta}_n - \theta_0\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) = \left(\widehat{\theta}_n - \widetilde{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\left(\widehat{\theta}_n - \theta_0\right)$.

Now, from (13) and (20), we have

$$
\begin{aligned}
2\left(L\left(\widehat{\theta}_n\right) - L\left(\widetilde{\theta}_n\right)\right) &= \sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right)' \mathcal{I}\left(\theta_0\right)\sqrt{n}\left(\widehat{\theta}_n - \widetilde{\theta}_n\right) \\
&= \frac{1}{\sqrt{n}}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\mathcal{I}\left(\theta_0\right)\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta}\frac{1}{\sqrt{n}} \\
&= \frac{1}{n}\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\theta_0\right)\frac{\partial L\left(\widetilde{\theta}_n\right)}{\partial \theta} = \xi_n^S \sim \chi^2\left(r\right) \qquad \text{under } H_0.
\end{aligned}
$$

- Calculating LR test statistic requires two maximizations of likelihood function one with and the other without constraint.
- LR test is also an asymptotically consistent test.
- As shown above, Wald, LM and LR test are asymptotically equivalent with $\chi^2\left(r\right)$.

Examples of tests in the linear regression model

Suppose the regression model such as

$$\begin{aligned} y_i &= \beta' x_i + \varepsilon_i \\ \varepsilon_i &\sim i.i.n. \left(0, \sigma^2\right) \end{aligned}$$

The hypotheses are given by

$$H_0; \underset{(r\times p)}{R} \underset{(p\times 1)}{\beta} = \gamma \qquad\qquad H_0; R\beta \neq \gamma$$

The log-likelihood function is given by

$$L(\beta) \approx -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

Then, the unconstrained MLE is given by

$$\begin{aligned} \widehat{\beta}_n &= (X'X)^{-1} X'y \\ \widehat{\sigma}_n^2 &= \frac{1}{n}\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right) \end{aligned}$$

Information matrix is given by

$$\mathcal{I}_n(\theta_0) = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

The Wald statistic is, from Proposition 1,

$$\begin{aligned} \xi_n^W &= n\left(R\widehat{\beta}_n - \gamma\right)'\left[\begin{bmatrix} R & 0 \end{bmatrix}' \mathcal{I}^{-1}\left(\widehat{\theta}_n\right)\begin{bmatrix} R \\ 0 \end{bmatrix}\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right) \\ &= \left(R\widehat{\beta}_n - \gamma\right)'\left[\begin{bmatrix} R & 0 \end{bmatrix}' \mathcal{I}_n^{-1}\left(\widehat{\theta}_n\right)\begin{bmatrix} R \\ 0 \end{bmatrix}\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right) \\ &= \left(R\widehat{\beta}_n - \gamma\right)'\left[R'\widehat{\sigma}_n^2 (X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right) \\ &= \frac{1}{\widehat{\sigma}_n^2}\left(R\widehat{\beta}_n - \gamma\right)'\left[R'(X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right) \sim \chi^2(r) \text{ under } H_0. \end{aligned}$$

Denote the constrained MLE as $\widetilde{\beta}_n$ and $\widetilde{\sigma}_n^2$, respectively. Then,

$$\begin{aligned} \widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2 &= \frac{1}{n}\left(y - X\widetilde{\beta}_n\right)'\left(y - X\widetilde{\beta}_n\right) - \frac{1}{n}\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right) \\ &= \frac{1}{n}\left(X\widetilde{\beta}_n - X\widehat{\beta}_n\right)'\left(X\widetilde{\beta}_n - X\widehat{\beta}_n\right) \\ &= \frac{1}{n}\left(\widetilde{\beta}_n - \widehat{\beta}_n\right)' X'X\left(\widetilde{\beta}_n - \widehat{\beta}_n\right) = \frac{1}{n}\left(R\widehat{\beta}_n - \gamma\right)'\left[R'(X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right) \end{aligned}$$

since $\widetilde{\beta}_n = \widehat{\beta}_n + (X'X)^{-1} R'\left[R'(X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right)$. Therefore,

$$\begin{aligned} \xi_n^W &= \frac{n\left[\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2\right]}{\widehat{\sigma}_n^2} = \frac{\left(R\widehat{\beta}_n - \gamma\right)'\left[R'(X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right)}{\frac{1}{n}\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right)} \\ &= \frac{\left[\left(R\widehat{\beta}_n - \gamma\right)'\left[R'(X'X)^{-1} R\right]^{-1}\left(R\widehat{\beta}_n - \gamma\right)\right]/r}{\left[\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right)\right]/(n-K)} \times \frac{nr}{n-K} = \frac{nr}{n-K}F \end{aligned}$$

On the other hand, the Lagrange multiplier of the constrained maximization problem is

$$\widetilde{\lambda}_n = -\frac{2}{\widetilde{\sigma}_n^2}\left(R\left(X'X\right)^{-1}R'\right)^{-1}\left(\gamma - R\widehat{\beta}_n\right)$$

Under H_0, the distribution of the Lagrange multiplier is

$$\widetilde{\lambda}_n \sim N\left(0, \frac{4}{\widetilde{\sigma}_n^2}\left(R\left(X'X\right)^{-1}R'\right)^{-1}\right)$$

since $\left(\gamma - R\widehat{\beta}_n\right) \sim N\left(0, \widetilde{\sigma}_n^2 R\left(X'X\right)^{-1}R'\right)$. Then, the LM test statistic is

$$\begin{aligned}
\xi_n^S &= \frac{\widetilde{\sigma}_n^2}{4}\widetilde{\lambda}_n'\left(R\left(X'X\right)^{-1}R'\right)\widetilde{\lambda}_n \\
&= \frac{1}{\widetilde{\sigma}_n^2}\left(R\widehat{\beta}_n - \gamma\right)'\left(R\left(X'X\right)^{-1}R'\right)^{-1}\left(R\widehat{\beta}_n - \gamma\right) \\
&= n\frac{\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2}{\widetilde{\sigma}_n^2} = \frac{n}{1 - 1 + \frac{\widetilde{\sigma}_n^2}{\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2}} = \frac{n}{1 + \frac{\widehat{\sigma}_n^2}{\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2}} = \frac{n}{1 + \frac{(n+K)}{rF}}
\end{aligned}$$

To obtain LR test statistic, note that

$$\begin{aligned}
L\left(\widehat{\theta}_n\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}_n^2 - \frac{1}{2\widehat{\sigma}_n^2}\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}_n^2 - \frac{n}{2\widehat{\sigma}_n^2} \times \frac{1}{n}\left(y - X\widehat{\beta}_n\right)'\left(y - X\widehat{\beta}_n\right) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}_n^2 - \frac{n}{2\widehat{\sigma}_n^2} \times \widehat{\sigma}_n^2 \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widehat{\sigma}_n^2 - \frac{n}{2}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
L\left(\widetilde{\theta}_n\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widetilde{\sigma}_n^2 - \frac{1}{2\widetilde{\sigma}_n^2}\left(y - X\widetilde{\beta}_n\right)'\left(y - X\widetilde{\beta}_n\right) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widetilde{\sigma}_n^2 - \frac{n}{2\widetilde{\sigma}_n^2} \times \frac{1}{n}\left(y - X\widetilde{\beta}_n\right)'\left(y - X\widetilde{\beta}_n\right) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widetilde{\sigma}_n^2 - \frac{n}{2\widetilde{\sigma}_n^2} \times \widetilde{\sigma}_n^2 \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\widetilde{\sigma}_n^2 - \frac{n}{2}
\end{aligned}$$

Hence,

$$\begin{aligned}
\xi_n^R &= 2\left(L\left(\widehat{\theta}_n\right) - L\left(\widetilde{\theta}_n\right)\right) = 2\left(-\frac{n}{2}\log\widehat{\sigma}_n^2 + \frac{n}{2}\log\widetilde{\sigma}_n^2\right) \\
&= n\log\frac{\widetilde{\sigma}_n^2}{\widehat{\sigma}_n^2} = n\log\left(1 - 1 + \frac{\widetilde{\sigma}_n^2}{\widehat{\sigma}_n^2}\right) = n\log\left(1 + \frac{\widetilde{\sigma}_n^2 - \widehat{\sigma}_n^2}{\widehat{\sigma}_n^2}\right) \\
&= n\log\left(1 + \frac{rF}{n-K}\right)
\end{aligned}$$

An interesting result can be obtained using the following inequalities,

$$\frac{x}{1+x} \leq \log\left(1+x\right) \leq x \;\forall x > -1$$

Let $x = \frac{rF}{n-K}$ and applying the above inequalities,we obtain

$$\xi_n^S \leq \xi_n^R \leq \xi_n^W$$