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

Suppose that we have the density $f(y;\theta)$ 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 $\hat{\theta}$ be the MLE of $\theta$.

Wald test is based on the very intuitive idea that we are willing to accept the null hypothesis when $\hat{\theta}$ is close to $\theta_0$. The distance between $\hat{\theta}$ and $\theta_0$ 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

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

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 $\left(Y_i, X_i\right)$ are i.i.d. and we can obtain conditional distribution of endogenous variables given the exogenous variables as $f\left(y_i \mid x_i;\theta\right)$ with $\theta \in \Theta \subset R^p$. The conditional density is known up to unknown parameter vector $\theta$. By i.i.d. assumption, we can write down the log-likelihood function of $n$ observations of $\left(Y_i, X_i\right)$ as

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

We assume all the regularity conditions for existence, consistency and asymptotic normality of MLE and denote MLE as $\hat{\theta}_n$. 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

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

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

pf) From the asymptotic characteristics of MLE, we know that

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

The first order Taylor series expansion of $g\left(\hat{\theta}_n\right)$ around the true value of $\theta_0$, we have

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

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

Hence, combining (1) and (2) gives

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

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

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

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

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

The statistic in (5) is useless since it depends on the unknown parameter $\theta_0$. However, we can consistently approximate the terms in inverse bracket by evaluating at MLE, $\hat{\theta}_n$. Therefore,

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

- 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, $g\left(\hat{\theta}_n\right) \stackrel{p}{\to} g(\theta_0) \neq 0$. Therefore, $g'\left(\hat{\theta}_n\right)\left(\frac{\partial g\left(\hat{\theta}_n\right)}{\partial \theta'}\mathcal{I}^{-1}\left(\hat{\theta}_n\right)\frac{\partial g'\left(\hat{\theta}_n\right)}{\partial \theta}\right)^{-1} g\left(\hat{\theta}_n\right)$ is converging to a constant instead of zero. By multiplying a constant by $n$, $W_n^W \to \infty$ 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 -

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

where $\mathcal{I}_n = E_X E_\theta\left(-\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'}\right) = E_X E_\theta\left(-\sum_{i=1}^{n}\frac{\partial^2 \log f(y_i|x_i;\theta)}{\partial \theta \partial \theta'}\right)$ and $\mathcal{I}_n^{-1}\left(\hat{\theta}_n\right)$ is the inverse of $\mathcal{I}_n$ evaluated at $\theta = \hat{\theta}_n$. 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.,

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

where $\theta_1$ is a $(q \times 1)$ subvector of $\theta$ with $q < p$. Then, the Wald statistic is given by

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

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

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

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

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

FOC's are given by

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

$$g\left(\tilde{\theta}_n\right) = 0 \qquad (7)$$

where $\tilde{\theta}_n$ is the solution of constrained maximization problem called constrained MLE and $\lambda$ is the vector of Lagrange multiplier.

The LM test is based on the idea that properly scaled $\tilde{\lambda}$ has an asymptotically normal distribution.

Proposition

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

Proof: First order Taylor expansions of $g\left(\hat{\theta}_n\right)$ and $g\left(\tilde{\theta}_n\right)$ around $\theta_0$ gives, ignoring $o_p(1)$ terms,

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

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

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

$$\sqrt{n}g\left(\hat{\theta}_n\right) = \frac{\partial g(\theta_0)}{\partial \theta'}\sqrt{n}\left(\hat{\theta}_n - \tilde{\theta}_n\right) \qquad (10)$$

On the other hand, taking first order Taylor series expansions of $\frac{\partial L(\hat\theta_n)}{\partial\theta}$ and $\frac{\partial L(\tilde\theta_n)}{\partial\theta}$ around $\theta_0$ gives, ignoring $o_p(1)$ terms,

$$\frac{\partial L(\hat\theta_n)}{\partial\theta} = \frac{\partial L(\theta_0)}{\partial\theta} + \frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}(\hat\theta_n-\theta_0) \cdots$$

$$\frac{1}{\sqrt n}\frac{\partial L(\hat\theta_n)}{\partial\theta} = \frac{1}{\sqrt n}\frac{\partial L(\theta_0)}{\partial\theta} + \frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}\sqrt n(\hat\theta_n-\theta_0) \cdots$$

$$\frac{1}{\sqrt n}\frac{\partial L(\hat\theta_n)}{\partial\theta} = \frac{1}{\sqrt n}\frac{\partial L(\theta_0)}{\partial\theta} - \mathcal{I}(\theta_0)\sqrt n(\hat\theta_n-\theta_0) \qquad (11)$$

note that $-\frac1n\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'} = -\frac1n\sum_{i=1}^n \frac{\partial^2\log(y_i|x_i;\theta)}{\partial\theta\partial\theta'} \xrightarrow{p} \mathcal{I}(\theta_0)$ by the law of large numbers. Similarly,

$$\frac{1}{\sqrt n}\frac{\partial L(\tilde\theta_n)}{\partial\theta} = \frac{1}{\sqrt n}\frac{\partial L(\theta_0)}{\partial\theta} - \mathcal{I}(\theta_0)\sqrt n(\tilde\theta_n-\theta_0) \qquad (12)$$

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

$$\frac{1}{\sqrt n}\frac{\partial L(\tilde\theta_n)}{\partial\theta} = -\mathcal{I}(\theta_0)\sqrt n(\tilde\theta_n-\hat\theta_n) = \mathcal{I}(\theta_0)\sqrt n(\hat\theta_n-\tilde\theta_n) \qquad (13)$$

Hence,

$$\sqrt n(\hat\theta_n-\tilde\theta_n) = \mathcal{I}^{-1}(\theta_0)\frac{1}{\sqrt n}\frac{\partial L(\tilde\theta_n)}{\partial\theta} \qquad (14)$$

From (10) and (14), we obtain

$$\sqrt n g(\hat\theta_n) = \frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{1}{\sqrt n}\frac{\partial L(\tilde\theta_n)}{\partial\theta}$$

Using (6), we deduce

$$\sqrt n g(\hat\theta_n) = -\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{\partial g'(\tilde\theta_n)}{\partial\theta}\frac{\tilde\lambda}{\sqrt n}$$

$$\to -\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{\partial g'(\theta_0)}{\partial\theta}\frac{\tilde\lambda}{\sqrt n} \qquad (15)$$

since $\tilde\theta_n \xrightarrow{p} \theta_0$ hence $g(\tilde\theta_n) \xrightarrow{p} g(\theta_0)$. Therefore,

$$\frac{\tilde\lambda}{\sqrt n} = -\left(\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{\partial g'(\theta_0)}{\partial\theta}\right)^{-1}\sqrt n g(\hat\theta_n) \qquad (16)$$

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

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

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

$$\frac1n\tilde\lambda'\left(\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{\partial g'(\theta_0)}{\partial\theta}\right)\tilde\lambda \sim \chi^2(r) \quad \text{under } H_0. \qquad (18)$$

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

$$\frac1n\frac{\partial L(\tilde\theta_n)}{\partial\theta'}\mathcal{I}^{-1}(\theta_0)\frac{\partial L(\tilde\theta_n)}{\partial\theta} \sim \chi^2(r) \quad \text{under } H_0 \qquad (19)$$

Note that (18) and (19) are useless since they depend on the unknown parameter value $\theta_0$. We can evaluate the terms involved in $\theta_0$ at the *constrained MLE*, $\tilde\theta_n$ to get a usable statistic.

- Again, another form of LM test is $\xi_n^S = \frac{\partial L(\tilde\theta_n)}{\partial\theta'}\mathcal{I}_n^{-1}(\theta_0)\frac{\partial L(\tilde\theta_n)}{\partial\theta} = \tilde\lambda'\left(\frac{\partial g(\theta_0)}{\partial\theta'}\mathcal{I}_n^{-1}(\theta_0)\frac{\partial g'(\theta_0)}{\partial\theta}\right)\tilde\lambda$.
- We can approximate $\mathcal{I}(\theta_0)$ with either $-\frac1n\sum_{i=1}^n\frac{\partial^2\log(y_i|x_i;\tilde\theta_n)}{\partial\theta\partial\theta'}$ or $\frac1n\sum_{i=1}^n\frac{\partial\log(y_i|x_i;\tilde\theta_n)}{\partial\theta}\frac{\partial\log(y_i|x_i;\tilde\theta_n)}{\partial\theta'}$. If we choose the second approximation, the LM test statistic becomes

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

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 $\frac{\partial \log f(y_i \mid x_i;\tilde\theta_n)}{\partial\theta}$ is given by

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

where $\underset{(n\times p)}{X} = \begin{bmatrix} \frac{\partial \log f(y_1 \mid x_1;\tilde\theta_n)}{\partial\theta'} & \frac{\partial \log f(y_2 \mid x_2;\tilde\theta_n)}{\partial\theta'} & \cdots & \frac{\partial \log f(y_n \mid x_n;\tilde\theta_n)}{\partial\theta'} \end{bmatrix}'$ and $\underset{(n\times 1)}{\mathbf{1}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}'$. Then,

$$R_u^2 = \frac{\sum_{i=1}^n \frac{\partial \log f(y_i \mid x_i;\tilde\theta_n)}{\partial\theta'}\left(\sum_{i=1}^n \frac{\partial \log f(y_i \mid x_i;\tilde\theta_n)}{\partial\theta}\frac{\partial \log f(y_i \mid x_i;\tilde\theta_n)}{\partial\theta'}\right)^{-1}\sum_{i=1}^n \frac{\partial \log f(y_i \mid x_i;\tilde\theta_n)}{\partial\theta}}{n}$$

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),

$$\begin{aligned}
\xi_n^W &= ng'(\hat\theta_n)\left(\frac{\partial g(\hat\theta_n)}{\partial\theta'}I^{-1}(\hat\theta_n)\frac{\partial g'(\hat\theta_n)}{\partial\theta}\right)^{-1}g(\hat\theta_n) \\
&\approx ng'(\hat\theta_n)\left(\frac{\partial g(\theta_0)}{\partial\theta'}I^{-1}(\theta_0)\frac{\partial g'(\theta_0)}{\partial\theta}\right)^{-1}g(\hat\theta_n) = \xi_n^S
\end{aligned}$$

## Likelihood ratio(LR) test

Proposition

$$\xi_n^R = 2\left(L(\hat\theta_n) - L(\tilde\theta_n)\right) \xrightarrow{d} \chi^2(r) \qquad \text{under } H_0.$$

Proof: We consider the second order Taylor expansions of $L(\hat\theta_n)$ and $L(\tilde\theta_n)$ around $\theta_0$. Under the null hypothesis, ignoring stochastically dominated terms,

$$\begin{aligned}
L(\hat\theta_n) &= L(\theta_0) + \frac{\partial L(\theta_0)}{\partial\theta'}(\hat\theta_n - \theta_0) + \frac{1}{2}(\hat\theta_n - \theta_0)'\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}(\hat\theta_n - \theta_0) \\
&= L(\theta_0) + \frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial\theta'}\sqrt{n}(\hat\theta_n - \theta_0) + \frac{1}{2}\sqrt{n}(\hat\theta_n - \theta_0)'\frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\hat\theta_n - \theta_0) \\
L(\tilde\theta_n) &= L(\theta_0) + \frac{\partial L(\theta_0)}{\partial\theta'}(\tilde\theta_n - \theta_0) + \frac{1}{2}(\tilde\theta_n - \theta_0)'\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}(\tilde\theta_n - \theta_0) \\
&= L(\theta_0) + \frac{1}{\sqrt{n}}\frac{\partial L(\theta_0)}{\partial\theta'}\sqrt{n}(\tilde\theta_n - \theta_0) + \frac{1}{2}\sqrt{n}(\tilde\theta_n - \theta_0)'\frac{1}{n}\frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'}\sqrt{n}(\tilde\theta_n - \theta_0)
\end{aligned}$$

Taking differences and multiplying by 2, we obtain

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

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

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

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

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

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

- 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 $\chi^2(r)$.

## 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

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

The log-likelihood function is given by

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

Then, the unconstrained MLE is given by

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

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,

$$\xi_n^W = n\left(R\hat{\beta}_n - r\right)'\left[\begin{bmatrix} R & 0 \end{bmatrix}' I^{-1}\left(\hat{\theta}_n\right)\begin{bmatrix} R \\ 0 \end{bmatrix}\right]^{-1}\left(R\hat{\beta}_n - r\right)$$

$$= \left(R\hat{\beta}_n - r\right)'\left[\begin{bmatrix} R & 0 \end{bmatrix}' I_n^{-1}\left(\hat{\theta}_n\right)\begin{bmatrix} R \\ 0 \end{bmatrix}\right]^{-1}\left(R\hat{\beta}_n - r\right)$$

$$= \left(R\hat{\beta}_n - r\right)'\left[R'\hat{\sigma}_n^2\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}_n - r\right)$$

$$= \frac{1}{\hat{\sigma}_n^2}\left(R\hat{\beta}_n - r\right)'\left[R'\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}_n - r\right) \sim \chi^2(r) \text{ under } H_0.$$

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

$$\tilde{\sigma}_n^2 - \hat{\sigma}_n^2 = \frac{1}{n}\left(y - X\tilde{\beta}_n\right)'\left(y - X\tilde{\beta}_n\right) - \frac{1}{n}\left(y - X\hat{\beta}_n\right)'\left(y - X\hat{\beta}_n\right)$$

$$= \frac{1}{n}\left(X\tilde{\beta}_n - X\hat{\beta}_n\right)'\left(X\tilde{\beta}_n - X\hat{\beta}_n\right)$$

$$= \frac{1}{n}\left(\tilde{\beta}_n - \hat{\beta}_n\right)'X'X\left(\tilde{\beta}_n - \hat{\beta}_n\right) = \frac{1}{n}\left(R\hat{\beta}_n - r\right)'\left[R'\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}_n - r\right)$$

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

$$\xi_n^W = \frac{n\left[\tilde{\sigma}_n^2 - \hat{\sigma}_n^2\right]}{\hat{\sigma}_n^2} = \frac{\left(R\hat{\beta}_n - r\right)'\left[R'\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}_n - r\right)}{\frac{1}{n}\left(y - X\hat{\beta}_n\right)'\left(y - X\hat{\beta}_n\right)}$$

$$= \frac{\left[\left(R\hat{\beta}_n - r\right)'\left[R'\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}_n - r\right)\right]/r}{\left[\left(y - X\hat{\beta}_n\right)'\left(y - X\hat{\beta}_n\right)\right]/\left(n - K\right)} \times \frac{nr}{n - K} = \frac{nr}{n - K}F$$

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

$$\tilde{\lambda}_n = -\frac{2}{\tilde{\sigma}_n^2}\left(R\left(X'X\right)^{-1}R'\right)^{-1}\left(r - R\hat{\beta}_n\right)$$

Under $H_0$, the distribution of the Lagrange multiplier is

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

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

$$\xi_n^S = \frac{\tilde{\sigma}_n^2}{4}\tilde{\lambda}_n'\left(R\left(X'X\right)^{-1}R'\right)\tilde{\lambda}_n$$

$$= \frac{1}{\tilde{\sigma}_n^2}\left(R\hat{\beta}_n - r\right)'\left(R\left(X'X\right)^{-1}R'\right)^{-1}\left(R\hat{\beta}_n - r\right)$$

$$= n\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}{\tilde{\sigma}_n^2} = \frac{n}{1 - 1 + \frac{\tilde{\sigma}_n^2}{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}} = \frac{n}{1 + \frac{\hat{\sigma}_n^2}{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}} = \frac{n}{1 + \frac{(n+K)}{rF}}$$

To obtain LR test statistic, note that

$$L\left(\hat{\theta}_n\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_n^2 - \frac{1}{2\hat{\sigma}_n^2}\left(y - X\hat{\beta}_n\right)'\left(y - X\hat{\beta}_n\right)$$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_n^2 - \frac{n}{2\hat{\sigma}_n^2} \times \frac{1}{n}\left(y - X\hat{\beta}_n\right)'\left(y - X\hat{\beta}_n\right)$$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_n^2 - \frac{n}{2\hat{\sigma}_n^2} \times \hat{\sigma}_n^2$$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_n^2 - \frac{n}{2}$$

On the other hand,

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

Hence,

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

An interesting result can be obtained using the following inequalities,

$$
\frac{x}{1+x} \le \log(1+x) \le x \quad x > -1
$$

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

$$
Y_n^S \le Y_n^R \le Y_n^W
$$